Н. К. Верещагин
В. А. Успенский
А. Шень

КОЛМОГОРОВСКАЯ СЛОЖНОСТЬ И АЛГОРИТМИЧЕСКАЯ СЛУЧАЙНОСТЬ

Москва
Издательство МЦНМО
2013

УДК 510.5
ББК 22.12
В77

Верещагин Н. К. и др.

В77 Колмогоровская сложность и алгоритмическая случайность / Н. К. Верещагин, В. А. Успенский, А. Шень. — М.: МЦНМО, 2013. — 576 с.

ISBN 978-5-4439-0212-8

Классическая (шенноновская) теория информации измеряет количество информации, заключённой в случайных величинах. В середине 1960-х годов А. Н. Колмогоров (и другие авторы) предложили измерять количество информации в конечных объектах с помощью теории алгоритмов, определив сложность объекта как минимальную длину программы, порождающей этот объект. Это определение послужило основой для алгоритмической теории информации, а также для алгоритмической теории вероятностей: объект считается случайным, если его сложность близка к максимальной.

Предлагаемая книга содержит подробное изложение основных понятий алгоритмической теории информации и теории вероятностей, а также наиболее важных работ, выполненных в рамках «колмогоровского семинара по сложности определений и сложности вычислений», основанного А. Н. Колмогоровым в начале 1980-х годов.

Книга рассчитана на студентов и аспирантов математических факультетов и факультетов теоретической информатики.

ББК 22.12

В оформлении форзацев использованы фрагменты рукописи статьи А. Н. Колмогорова [65].

Николай Константинович Верещагин
Владимир Андреевич Успенский
Александр Шень

КОЛМОГОРОВСКАЯ СЛОЖНОСТЬ
И АЛГОРИТМИЧЕСКАЯ СЛУЧАЙНОСТЬ

Издательство Московского центра
непрерывного математического образования
119002, Москва, Большой Власьевский пер., 11. Тел. (499) 241-74-83

Подписано в печать 02.11.2012 г. Формат 70×100 1/16.
Бумага офсетная. Печать офсетная. Печ. л. 36. Тираж 1000. Заказ .

Отпечатано в ППП «Типография „Наука“».
121099, Москва, Шубинский пер., д. 6.

ISBN 978-5-4439-0212-8

Памяти Андрея Мучника

Предисловие

Понятие *колмогоровской сложности* (или, как ещё говорят, *алгоритмической энтропии*) появилось в 1960-е годы на стыке теории алгоритмов, теории информации и теории вероятностей.

Идея А.Н. Колмогорова, опубликованная им в знаменитой статье 1965 года [63], состояла в том, чтобы измерять количество информации, заключённой в индивидуальных конечных объектах (а не в случайных величинах, как в шенноновской теории информации). Оказалось, что это возможно (хотя лишь с точностью до ограниченного слагаемого).

На несколько лет раньше сходные идеи высказывал Р. Соломонов (см. [160] и другие его статьи, ссылки на которые даны в [85])[1]. Он исходил из несколько другой мотивировки, пытаясь ввести понятие «априорной вероятности» — с какими вероятностями мы ожидаем появления тех или иных объектов, если ничего не знаем о порождающем их процессе? Оказалось, что вопросы эти тесно связаны. В самом первом приближении можно сказать так: если объект простой, то он имеет большую априорную вероятность, а если сложный — то малую. (К сожалению, работы Соломонова приобрели известность в основном лишь после того, как Колмогоров упомянул их в своей статье.)

В 1965 году американский математик Г. Чейтин (тогда только что окончивший школу) представил к публикации две своих статьи [19] и [20], напечатанные в 1966 и 1969 годах. Во второй из них он дал то же определение алгоритмической сложности, что и Колмогоров.

Основные свойства колмогоровской сложности были исследованы в 1970-е годы. К. П. Шнорр и Л. А. Левин (ученик Колмогорова) установили связь понятия сложности с алгоритмическим определением случайности (предложенным учеником Колмогорова шведским математиком П. Мартин-Лёфом). При этом они ввели некоторый новый вариант понятия сложности (так называемую монотонную сложность). Идеи Соломонова об априорной вероятности привели к (введённому Левиным и позднее Чейтиным) понятию префиксной сложности, которое оказалось не только философски важным, но и технически удобным. Были найдены применения понятия сложности как в теории алгоритмов, так и в теории вероятностей.

Интерес к колмогоровской сложности (заслуженный — это, безусловно, одно из базисных понятий теории алгоритмов, имеющее фундаментальное значение) возрос в последние десятилетия, когда появилась монография Ли и Витаньи [85] (первое издание вышло в 1993 году). Без преувеличения можно сказать, что практически

[1]«Я пришел к аналогичной концепции, еще не зная работ Соломонова», писал Колмогоров в названной выше статье [63]

всё известное к тому моменту о колмогоровской сложности вошло в эту книгу (в основной текст или в упражнения). Книга включает в себя подробный рассказ об истории области, ссылки на первые публикации и так далее. Затем вышли книги К. Калюде [16], А. Ниса [127], а также Р. Доуни и Д. Хиршфельдта [37]; в них вошли многочисленные результаты, полученные в последние годы (в частности, связывающие понятия сложности и случайности с классической теорией вычислимости).

Наша книга представляет собой учебник, ни в коей мере не претендующий на полноту (и, в частности, не включающий упомянутые результаты последних лет). Мы пытались отобрать наиболее важные (философски и технически) факты о колмогоровской сложности и доступно о них рассказать. За редкими исключениями, мы почти ничего не говорим об истории вопроса. Как и в любом учебнике, многие утверждения формулируются без указания авторства (и это ни в коей мере не означает, что мы пытаемся приписать их себе).

В следующем разделе («О чём эта книга?») мы пытаемся кратко очертить круг вопросов, относящихся к колмогоровской сложности. Грубо говоря, он предназначен для тех, кто хочет понять, что это за область и стоит ли знакомиться с нею подробно.

В конце книги в качестве приложения воспроизведен текст небольшой брошюры одного из авторов (В.У.), написанной по материалам его лекции для студентов младших курсов и старшеклассников (23 июля 2005 года в летней школе «Современная математика» в Дубне; первое издание брошюры вышло в 2006 г. в издательстве МЦНМО). На этой лекции было рассказано о различных подходах к математическому определению понятия случайности, и мы приводим её текст (с небольшими изменениями), чтобы читатель, не имеющий возможности или охоты подробно изучать соответствующие разделы основного текста, мог составить некоторое представление об этой теме.

К сожалению, с названиями, обозначениями и терминологией в этой области большая путаница — отчасти и по вине авторов этой книги. Например, в Приложении 2 использована терминология, отличная от терминологии, принятой в основном тексте. Там, в частности, появляются термины *хаотический*, *типический* и *непредсказуемый*[1]. Но суть, конечно, не в терминах, а в стоящих за ними понятиях. Каждый из названных терминов отражает некоторое специфическое свойство случайного объекта; хаотичность опирается на представление о колмогоровской сложности, типичность — на представление о мере множества, непредсказуемость — на представление о выигрывающем плане игры. Понятия хаотичности, типичности и непредсказуемости допускают строгие определения, каждое из которых может претендовать на то, что оно-то и есть истинное определение случайности. Совпадение по объёму понятий хаотичности и типичности для двоичных последовательностей, установленное теоремой Левина – Шнорра, является глубоким фактом, не очевидным априори и демонстрирующим адекватность соответствующих определе-

[1] Термины «хаотический» и «типический» дороги автору приложения тем, что были одобрены Колмогоровым и впервые публично прозвучали в 1986 г. на открытии Первого Всемирного конгресса Общества математической статистики и теории вероятностей им. Бернулли в докладе «Алгоритмы и случайность» А. Н. Колмогорова и В. А. У. Они же были использованы в статьях [67, 182, 167, 120].

ний. Само основное понятие называют *алгоритмической сложностью,* колмогоровской сложностью, а также алгоритмической (или колмогоровской) *энтропией*. Мы будем считать все эти слова синонимами.

Термин «энтропия» в рассматриваемую тематику был введён Колмогоровым и не был для него случайным. Основатель теории информации Клод Шеннон построил её на вероятностном фундаменте. Колмогоров имел амбициозный план построить параллельную теорию, заменив вероятностный фундамент алгоритмическим. Для него было естественным сохранить основной термин шенноновской теории, термин «энтропия», но наполнить его новым содержанием. В своей статье [64] Колмогоров писал:

> Обычно пользуются вероятностным определением энтропии, которое относится, однако, не к индивидуальным объектам, а к «случайным», т. е. по существу к распределениям вероятностей в каких-либо классах объектов. $\langle\ldots\rangle$ Далеко не все применения теории информации укладываются разумным образом в такую интерпретацию её основных понятий. $\langle\ldots\rangle$ Насколько я знаю, первой печатной публикацией, содержащей замысел перестройки теории информации, удовлетворяющей высказанному пожеланию, является статья Соломонова [160], опубликованная в 1964 г. Я пришел к аналогичной концепции, еще не зная работ Соломонова в 1963 – 1964 гг., и опубликовал первую заметку на эту тему [63] в начале 1965 г. $\langle\ldots\rangle$ Замысел определения очень прост: энтропия $H(x\backslash y)$ есть минимальная длина записанной в виде последовательности нулей и единиц «программы» P, которая позволяет построить объект x, имея в своем распоряжении объект y:
>
> $$H(x\backslash y) = \min_{A(P,y)=x} l(P).$$
>
> Точное проведение этого замысла опирается на общую теорию «вычислимых» (частично рекурсивных) функций, т. е. на общую теорию алгоритмов во всей их общности.
>
> $\langle\ldots\rangle$ Предшествующее краткое изложение должно оправдать два общих тезиса:
>
> 1) основные понятия теории информации должны и могут быть обоснованы без помощи обращения к теории вероятностей и так, что понятия «энтропия» и «количество информации» оказываются применимы к индивидуальным объектам;
>
> 2) введённые таким образом понятия теории информации могут лечь в основу новой концепции *случайного*, соответствующей естественной мысли о том, что случайность есть отсутствие закономерности.

Ещё раньше, 23 апреля 1965 года, Колмогоров так формулировал свой замысел, выступая в Институте философии Академии наук СССР с докладом «Понятие „информация“ и основы теории вероятностей»:

> Возникают последовательно такие проблемы:
>
> 1. Нельзя ли теорию информации (и понятие «количество информации» освободить от вероятностей?
>
> 2. Нельзя ли воспользоваться идеей, что случайность — это не поддающийся сокращению (описанию более коротким, чем уже привычный, способом) закон?

(цитируется по стенограмме, опубликованной в сборнике [68], с. 126).

Ещё бо́льшая путаница, чем с терминами, происходит с обозначениями. В оригинальной статье Колмогорова [63] сложность слова x обозначалась $K(x)$, а в его же статье 1969 года [64] использовалось обозначение $H(x)$ и для шенноновской энтропии, и для алгоритмической сложности. В других работах (например, у Ли и Витаньи) $K(x)$ обозначает префиксную сложность, а для исходного понятия используется обозначение $C(x)$. Чтобы избежать путаницы, мы вообще не будем использовать обозначение $K(x)$, а писать $KS(x)$ для исходного варианта сложности и $KP(x)$ для префиксной сложности. Есть и другие варианты понятия сложности; монотонную сложность (в её наиболее естественной форме) мы будем обозначать $KM(x)$, а сложность разрешения — $KR(x)$. (При этом некоторая путаница остаётся — у Ли и Витаньи монотонная сложность называется $Km(x)$, а обозначение $KM(x)$ используется для логарифма априорной вероятности, которые мы обозначаем $KA(x)$. Надеемся, это не собьёт читателя с толку.)

Есть и другие, менее существенные, различия в названиях и обозначениях; для удобства в конце книги приведён список используемых обозначений и предметный указатель.

* * *

В начале 1980-х годов на механико-математическом факультете МГУ по инициативе А.Н. Колмогорова и при участии А.Л. Семёнова начал работу семинар «Сложность определений и сложность вычислений» (его наследником является «Колмогоровский семинар», работающий и поныне). Авторы книги глубоко признательны всем своим коллегам по этому семинару, без которых эта книга не могла бы возникнуть, в том числе А. Звонкину, Е. Асарину и В. Вовку (ученикам Колмогорова), С. Ф. Сопрунову, В. Вьюгину, А. Ромащенко, М. Вялому, С. Тарасову, А. Чернову, М. Вьюгину, С. Посицельскому, К. Макарычеву, Ю. Макарычеву, М. Ушакову, М. Устинову, С. Сальникову, А. Румянцеву, Д. Мусатову, В. Подольскому, И. Межирову, Ю. Притыкину, М. Раскину, А. Ходыреву, П. Карповичу, А. Минасяну, Е. Калининой, Г. Челнокову, И. Разенштейну, М. Андрееву, А. Савину, М. Дектярёву, А. Савчику, А. Кумку, В. Арзуманяну, А. Махлину.

Лягушку и её проекции (для обложки) нарисовала Марина Фейгельман; обложку оформила Ольга Лехтонен. Как всегда, мы благодарны (в частности, за помощь при подготовке оригинал-макета) Виктору Шувалову.

Во время написания книги авторы пользовались финансовой поддержкой фонда Сороса, шведского фонда STINT, Российского фонда фундаментальных исследований по грантам РФФИ 01-01-00493-а, 01-01-01028-а, 03-01-00475-а, 06-01-00122-а,

09-01-00709-а, 12-01-00864-а, CNRS, а также французского Agence National de Recherche по гранту ANR-08-EMER-008 NAFIT.

Книга не была бы написана, если бы не поддержка (научная и человеческая) зарубежных коллег, в том числе Брюно Бауэнса, Лорана Бьенвеню, Гарри Бюрмана, Сергея Воробьёва, Пола Витаньи, Питера Гача, Брюно Дюрана, Криса Калюде, Михала Коуцки, Леонида Левина, Вольфганга Меркле, Джо Миллера, Андре Ниса, Яна Раймана, Майкла Сипсера, Стива Симпсона, Денниса Хиршфельдта, Матьё Хойрупа и многих других.

Авторы благодарны судьбе за предоставленную возможность научного и человеческого общения с Андреем Мучником (1958 – 2007), выдающимся математиком, глубоким мыслителем и замечательным человеком, многие годы участвовавшим в работе колмогоровского семинара. Мы посвящаем наш скромный труд его памяти.

Н. К. Верещагин, В. А. Успенский, А. Шень *2 ноября 2012 г.*

О чём эта книга?

Что такое колмогоровская сложность?

«На пальцах» это проще всего объяснить так. Существуют программы, которые сжимают файлы (для экономии места в архиве, без потери информации). Возможно, вы встречались с ними: наиболее распространённые называются `zip`, `gzip`, `bzip2`, `compress`, `rar`, `arj` (есть и другие).

Применив такую программу к некоторому файлу (с текстом, данными, программой), мы получаем его сжатую версию (которая, как правило, короче исходного файла). По ней можно восстановить исходный файл (с помощью парной программы-«декомпрессора»; часто сжатие и восстановление объединены в одну программу).

Так вот, в первом приближении колмогоровскую сложность файла можно описать как длину его сжатой версии. Файл, имеющий регулярную структуру и хорошо сжимаемый, имеет малую колмогоровскую сложность (в сравнении с его длиной). Напротив, плохо сжимаемый файл имеет сложность, близкую к длине.

Это описание весьма приблизительно и нуждается в уточнениях — как технических, так и принципиальных. Начнём с технического: вместо файлов (последовательностей байтов) мы будем рассматривать двоичные слова (конечные последовательности битов, то есть нулей и единиц). Длиной такого слова называется число знаков (так что слово 1001, скажем, имеет длину 4, а пустое слово имеет длину 0).

Более существенны другие отличия:

- Мы рассматриваем только программы восстановления файла из архива и не рассматриваем программы сжатия файлов в архив. Точнее, назовём *декомпрессором* произвольный алгоритм (программу), который получает на вход двоичные слова и выдаёт на выход также двоичные слова. Если декомпрессор D на входе x даёт y, мы пишем $D(x) = y$ и говорим, что x является *описанием* y при данном D (относительно данного D). Декомпрессоры мы также будем называть *способами описания*.

- Мы не требуем, чтобы декомпрессор был определён на всех словах. При некоторых x вычисление $D(x)$ может зацикливаться (не останавливаться) и не давать результата. Мы также не накладываем никаких ограничений на время работы D: на некоторых входах программа D может дать ответ лишь после очень долгой работы.

Используя терминологию теории вычислимых функций, можно сказать, что способ описания есть вычислимая функция из Ξ в Ξ, где Ξ — множество всех двоичных слов. Напомним, что с каждым алгоритмом D, входами и выходами которого являются двоичные слова, мы связываем функцию d, которая определена на слове x тогда и только тогда, когда D даёт результат на слове x; при этом значение $d(x)$ равно результату применения D ко входу x. Функции d, получаемые таким образом из всевозможных алгоритмов D, называются *вычислимыми функциями из* Ξ *в* Ξ. Обычно используют одну и ту же букву для обозначения алгоритма и вычисляемой им функции, и вместо $d(x)$ пишут $D(x)$. Мы также будем следовать этой традиции в тех случаях, когда это не вызывает путаницы.

Пусть фиксирован некоторый способ описания (декомпрессор) D. Для данного слова x рассмотрим все его описания, то есть все слова y, для которых $D(y)$ определено и равно x. Длину самого короткого из них и называют *колмогоровской сложностью* слова x при данном способе описания D:

$$KS_D(x) = \min\{l(y) \mid D(y) = x\}.$$

Здесь и далее $l(y)$ обозначает длину слова y. Индекс D подчёркивает, что определение зависит от выбора способа D. Минимум пустого множества, как обычно, считается равным $+\infty$, так что $KS_D(x)$ бесконечно для тех x, которые не входят в область значений функции D (не могут быть результатами декомпрессии, не имеют описаний).

Это определение кажется бессодержательным, поскольку для разных D получаются разные определения, в том числе абсурдные. Например, если D нигде не определён, то KS_D всегда бесконечно. Если $D(y) = \Lambda$ (пустое слово) при всех y, то сложность пустого слова равна нулю (поскольку $D(\Lambda) = \Lambda$ и $l(\Lambda) = 0$), а сложность всех остальных слов бесконечна.

Более осмысленный пример: если программа-декомпрессор копирует вход на выход и $D(x) = x$ при всех x, то каждое слово имеет единственное описание (самого себя) и $KS_D(x) = l(x)$.

Естественно, для любого слова x можно подобрать способ описания D, при котором x имеет малую сложность. Достаточно положить $D(\Lambda) = x$, и тогда $KS_D(x)$ будет равно нулю. Можно также подобрать способ описания, благоприятствующий целому классу слов: например, для слов из одних нулей хорош такой способ описания:

$$D(\mathrm{bin}(n)) = 000\ldots000 \quad (n \text{ нулей}),$$

где $\mathrm{bin}(n)$ — двоичная запись числа n. Легко проверить, что длина слова $\mathrm{bin}(n)$ примерно равна $\log_2 n$ (не превосходит $\log_2 n + 1$). Мы получаем, что для построенного способа описания сложность слова из n нулей близка к $\log_2 n$ (и много меньше длины слова, то есть n). Зато все другие слова (содержащие хотя бы одну единицу) имеют бесконечную сложность.

На первый взгляд кажется, что определение сложности настолько сильно зависит от выбора способа описания, что никакая общая теория невозможна.

Оптимальные способы описания

Способ описания тем лучше, чем короче описания. Поэтому естественно дать такое определение: способ D_1 *не хуже* способа D_2, если

$$KS_{D_1}(x) \leqslant KS_{D_2}(x) + c$$

при некотором c и при всех x.

В этом определении требует пояснения константа c. Мы соглашаемся пренебрегать увеличением сложности не более чем на константу. Конечно, можно сказать, что это лишает определение практического смысла, так как константа c может быть сколь угодно велика — однако без такого «огрубления» обойтись не удаётся.

Пример. Пусть даны два произвольных способа описания D_1 и D_2. Покажем, что существует способ описания D, который не хуже их обоих.

В самом деле, положим

$$D(0y) = D_1(y),$$
$$D(1y) = D_2(y).$$

Другими словами, первый бит описания мы воспринимаем как номер способа описания, а остальную часть — как собственно описание. Ясно, что если y является описанием x при D_1 (или D_2), то $0y$ (соответственно $1y$) является описанием x при D, и это описание всего лишь на один бит длиннее. Поэтому

$$KS_D(x) \leqslant KS_{D_1}(x) + 1,$$
$$KS_D(x) \leqslant KS_{D_2}(x) + 1$$

при всех x, так что способ D не хуже обоих способов D_1 и D_2.

Этот приём используется на практике. Например, формат `zip`-архива предусматривает заголовок, в котором указывается номер используемого метода сжатия, а затем идёт сам сжатый файл. При этом использование N различных способов описания приводит к тому, что начальные $\log_2 N$ битов (или около того) приходится зарезервировать для указания используемого способа.

Несколько обобщив эту же идею, можно доказать такую теорему:

Теорема 1 (Соломонова – Колмогорова). *Существует способ описания D, который не хуже любого другого: для всякого способа описания D' найдётся такая константа c, что*

$$KS_D(x) \leqslant KS_{D'}(x) + c$$

для любого слова x.

Способ описания, обладающий указанным в теореме свойством, называют *оптимальным.*

◀ Напомним, что по нашему определению способами описания являются вычислимые функции. Программы, их вычисляющие, можно считать двоичными словами. Будем при этом предполагать, что по программе можно определить, где она

кончается (такой способ записи по-английски называется self-delimiting; подходящего русского слова, пожалуй, нет, и мы будем называть такие программы «самоограниченными»). Формально говоря, самоограниченность означает, что никакая программа не может быть началом другой программы. Если выбранный способ записи программ не таков, то его можно модифицировать. Например, можно продублировать каждый знак в записи программы (заменив 0 на 00 и 1 на 11), а в конце программы добавить группу 01.

Теперь определим новый способ описания D, положив

$$D(py) = p(y),$$

где p — произвольная программа (при выбранном способе записи), а y — любое двоичное слово. Другими словами, D читает слово слева направо, выделяя из него начало-программу. (Если таковой не оказывается, D делает что угодно, например, зацикливается.) Далее D применяет найденную программу (p) к остатку входа (y) и выдаёт полученный результат. Самоограниченность гарантирует, что вход алгоритма D однозначно разлагается на программу и остаток: если $py = p'y'$, то одна из программ p, p' является началом другой, что противоречит самоограниченности.

Покажем, что D не хуже любого другого способа описания P. Пусть p — программа, вычисляющая функцию P, причём записанная в выбранной нами форме. Если слово y является кратчайшим описанием слова x относительно P, то py будет описанием x относительно D (не обязательно кратчайшим). Поэтому при переходе от P к D длина описания увеличится не более чем на $l(p)$, и

$$KS_D(x) \leqslant KS_P(x) + l(p).$$

Видно, что константа (то есть $l(p)$) зависит лишь от выбора способа описания P, но не от x. ►

По существу здесь используется тот же приём, что и в предыдущем примере, только вместо двух способов описания соединяются все мыслимые способы — каждый со своим префиксом. Именно так устроены так называемые «саморазархивирующиеся» архивы (self-extracting archives). Такой архив представляет собой исполняемый файл, в котором, однако, собственно программа занимает лишь небольшой начальный кусок. Эта программа загружается в память, после чего читает и разархивирует остаток архива.

Заметим, что построенный нами оптимальный способ описания на некоторых входах работает очень долго (поскольку бывают долго работающие программы), а на некоторых входах и вовсе не определён.

Колмогоровская сложность

Фиксируем некоторый оптимальный способ описания D и будем называть *колмогоровской сложностью* слова x значение $KS_D(x)$. В обозначении $KS_D(x)$ будем опускать индекс D и писать просто $KS(x)$.

Замена оптимального способа на другой оптимальный способ приводит к изменению сложности не более чем на константу: для любых оптимальных способов D_1 и D_2 найдётся такая константа $c(D_1, D_2)$, что

$$|KS_{D_1}(x) - KS_{D_2}(x)| \leqslant c(D_1, D_2)$$

при всех x. Это неравенство записывают как

$$KS_{D_1}(x) = KS_{D_2}(x) + O(1),$$

понимая под $O(1)$ произвольную ограниченную функцию от x.

Имеет ли смысл говорить о колмогоровской сложности конкретного слова x согласно этому определению, не указывая, какой оптимальный способ мы используем? Нет — легко понять, что подбором подходящего оптимального способа можно сделать сложность данного слова любой. Точно так же нельзя говорить, что слово x проще слова y (имеет меньшую сложность): выбирая тот или иной оптимальный способ, можно сделать любое из двух слов более простым.

Каков же тогда смысл колмогоровской сложности, если ни про какое конкретное слово ничего нельзя сказать?

В защиту этого понятия можно сказать следующее. Свойства оптимального способа, построенного при доказательстве теоремы Колмогорова – Соломонова, зависят от использованного способа записи программ (то есть от выбора языка программирования). Два таких способа приводят к сложностям, отличающимся не более чем на константу, которая есть длина интерпретатора одного языка программирования, написанного на другом языке. Можно надеяться, что при естественном выборе языков эта константа будет измеряться тысячами или даже сотнями. Тем самым, если мы говорим о сложностях порядка сотен тысяч (скажем, для текста романа) или миллионов (скажем, для ДНК), то уже не так важно, какой именно язык программирования мы выбрали.

Но тем не менее надо всегда помнить, что все утверждения о колмогоровской сложности носят асимптотический характер. Чтобы утверждение имело точный математический смысл, в нём должна идти речь о сложности не изолированного слова, а последовательности слов. Для теории сложности вычислений такого рода ситуация типична: скажем, оценки на сложность решения какой-либо задачи обычно также имеют асимптотический характер.

Сложность и информация

Колмогоровскую сложность слова x можно назвать также *количеством информации* в слове x. В самом деле, слово из одних нулей, которое может быть описано коротко, содержит мало информации, а какое-то сложное слово, которое не поддаётся сжатию, содержит много информации (пусть даже и бесполезной — мы не пытаемся отделить осмысленную информацию от бессмысленной, так что любая абракадабра для нас содержит много информации, если её нельзя задать коротко).

Если слово x имеет сложность k, мы говорим, что x содержит k битов информации. Естественно ожидать, что число битов информации в слове не превосходит

его длины, то есть что $KS(x) \leqslant l(x)$. Так и оказывается (но только надо добавить константу, о чём мы уже говорили).

Теорема 2. *Существует такая константа c, что*

$$KS(x) \leqslant l(x) + c$$

для любого слова x.

◀ Мы уже говорили, что если $P(y) = y$ при всех y, то $KS_P(x) = l(x)$. В силу оптимальности найдётся такое c, что

$$KS(x) \leqslant KS_P(x) + c = l(x) + c$$

при всех x. ▶

Утверждение этой теоремы обычно записывают так: $KS(x) \leqslant l(x) + O(1)$. Из него вытекает, в частности, что колмогоровская сложность любого слова конечна (то есть что любое слово имеет описание).

Вот ещё один пример свойства, которого естественно ожидать от понятия «количество информации»: при алгоритмических преобразованиях количество информации не увеличивается (точнее, увеличивается не более чем на константу, зависящую от алгоритма преобразования).

Теорема 3. *Для любого алгоритма A существует такая константа c, что*

$$KS(A(x)) \leqslant KS(x) + c$$

для всех x, при которых $A(x)$ определено.

◀ Пусть D — оптимальный декомпрессор, используемый при определении колмогоровской сложности. Рассмотрим другой декомпрессор D':

$$D'(p) = A(D(p))$$

(D' применяет сначала D, а затем A). Если p является описанием некоторого x относительно D, причём $A(x)$ определено, то p является описанием $A(x)$ относительно D'. Взяв в качестве p кратчайшее описание x относительно D, находим, что

$$KS_{D'}(A(x)) \leqslant l(p) = KS_D(x) = KS(x),$$

а в силу оптимальности

$$KS(A(x)) \leqslant KS_{D'}(A(x)) + c \leqslant KS(x) + c$$

при некотором c и при всех x, что и требовалось доказать. ▶

Эта теорема гарантирует, что количество информации «не зависит от кодировки» — если мы, скажем, заменим в каком-то слове все нули на единицы и наоборот (или разбавим это слово нулями, добавив после каждой цифры по нулю), то полученное слово будет иметь ту же сложность, что и исходное (с точностью до

константы), поскольку преобразования в ту и другую сторону выполняются некоторым алгоритмом.

Ещё один пример. Пусть x и y — два слова. Соединим их в одно, приписав y к x справа. Сколько битов информации будет иметь полученное слово? Естественно ожидать, что количество информации в нём не превосходит суммы количеств информации в x и в y. И действительно это так, — правда, с некоторой поправкой.

Теорема 4. *Существует такая константа c, что при любых x и y выполнено неравенство*

$$KS(xy) \leqslant KS(x) + 2\log KS(x) + KS(y) + c.$$

(Здесь и далее log обозначает двоичный логарифм, если иное основание не указано явно.)

◀ Попробуем для начала доказать теорему без добавочного члена $2\log KS(x)$, её ослабляющего. Это естественно делать так. Пусть D — оптимальный способ описания, используемый при определении колмогоровской сложности. Рассмотрим новый способ описания D'. Именно, если $D(p) = x$ и $D(q) = y$, будем считать pq описанием слова xy, то есть положим $D'(pq) = xy$. Тогда сложность слова xy относительно D' не превосходит длины слова pq, то есть $l(p) + l(q)$. Если в качестве p и q взять кратчайшие описания, то получится $KS_{D'}(xy) \leqslant KS_D(x) + KS_D(y)$, и остаётся воспользоваться оптимальностью D и перейти от D' к D (при этом возникает константа c).

Что неверно в этом рассуждении? Дело в том, что определение D' некорректно: мы полагаем $D'(pq) = D(p)D(q)$, но D' не имеет средств разделить p и q. Вполне может оказаться, что есть два разбиения слова на части, дающие разные результаты: $p_1q_1 = p_2q_2$, но $D(p_1)D(q_1) \ne D(p_2)D(q_2)$.

Есть два способа исправить эту ошибку. Первый, который мы и применим, состоит в том, чтобы перед словом p написать длину слова p в двоичной записи. При этом в этой двоичной записи мы удвоим каждый бит, а после неё напишем 01, чтобы алгоритм мог понять, где кончается двоичная запись и начинается само p. Более точно, обозначим через $\mathrm{bin}(k)$ двоичную запись числа k, а через $\bar{x}$ результат удвоения каждого бита в слове x. (Например, $\mathrm{bin}(5) = 101$, а $\overline{\mathrm{bin}(5)} = 110011$.) Теперь положим

$$D'(\overline{\mathrm{bin}(l(p))}01pq) = D(p)D(q).$$

Это определение корректно: алгоритм D' сначала читает $\overline{\mathrm{bin}(l(p))}$, пока цифры идут парами. Когда появляется группа 01, он определяет $l(p)$, затем отсчитывает $l(p)$ цифр и получает p. Остаток входа есть q, после чего уже можно вычислить $D(p)D(q)$.

Величина $KS_{D'}(xy)$ оценивается теперь числом $2l(\mathrm{bin}(l(p))) + 2 + l(p) + l(q)$; двоичная запись числа $l(p)$ имеет длину не больше $\log l(p) + 1$, поэтому D'-описание слова xy имеет длину не более $2\log l(p) + 4 + l(p) + l(q)$, откуда и вытекает утверждение теоремы. ▶

Альтернативный способ исправления допущенной ошибки состоит в том, чтобы модифицировать определение сложности, потребовав, чтобы описания были самоограниченными (self-delimiting); мы обсуждаем его в главе 4.

Заметим также, что в теореме можно поменять местами p и q и доказать, что $KS(xy) \leqslant KS(x) + KS(y) + 2\log_2 KS(y) + c$.

Насколько неравенство теоремы 4 близко к равенству? Может ли $KS(xy)$ быть существенно меньше суммы $KS(x) + KS(y)$? В полном согласии с нашей интуицией это возможно, если x и y содержат много общего. Например, при $x = y$ мы имеем $KS(xy) = KS(xx) = KS(x) + O(1)$, поскольку xx алгоритмически получается из x и обратно (теорема 3).

Чтобы уточнить это наблюдение, мы определим понятие количества информации, которая содержится в x, но отсутствует в y (для произвольных слов x и y). Эту величину называют *условной колмогоровской сложностью x при условии y* (говорят также «при известном y», «относительно y»). Её определение аналогично определению обычной (безусловной) сложности, но декомпрессор D имеет доступ не только к сжатому описанию, но и к слову y. Мы будем говорить об этом подробнее позже (глава 2), скажем лишь, что для условной сложности справедливо равенство

$$KS(xy) = KS(y) + KS(x \mid y) + O(\log n),$$

если x и y — слова сложности не более n. Читается оно так: количество информации в xy равно количеству информации в y плюс количество новой (отсутствующей в y) информации в x.

Разность $KS(x) - KS(x \mid y)$ естественно назвать *количеством информации об x в y*: эта разность показывает, насколькоz знание слова y упрощает описание слова x.

Понятие условной сложности позволяет придать смысл вопросу о том, сколько новой информации в ДНК одного организма по сравнению с ДНК другого: если d_1 — двоичное слово, кодирующее ДНК первого организма, а d_2 — двоичное слово, кодирующее ДНК второго, то искомая величина есть $KS(d_1 \mid d_2)$. Аналогичным образом можно спрашивать, какой процент информации был потерян при переводе романа на другой язык: в этом случае нас интересует отношение

$$KS(\text{оригинал} \mid \text{перевод}) / KS(\text{оригинал}).$$

Вопросы о количестве информации в различных объектах изучались и до алгоритмической теории информации, с помощью понятия шенноновской энтропии. Напомним её определение. Пусть случайная величина ξ принимает n значений с вероятностями $p_1, \ldots, p_n$. Тогда её *шенноновская энтропия* определяется формулой

$$H(\xi) = \sum p_i(-\log_2 p_i).$$

Говорят, что исход, имеющий вероятность p_i, несёт в себе $(-\log_2 p_i) = \log_2(1/p_i)$ битов информации; тогда $H(\xi)$ можно интерпретировать как среднее количество информации в исходе случайной величины.

Чтобы применить понятие шенноновской энтропии, скажем, к оценке количества информации в данном тексте на русском языке, нужно включить этот текст в какой-то ансамбль текстов, считая его «типичным» значением некоторой случайной величины. Это имеет смысл для короткой поздравительной телеграммы, но для текста романа трудно вообразить себе подходящий ансамбль. Столь же сложно

сделать это при оценке количества информации в ДНК: если считать ансамблем множество всех существовавших ДНК всех организмов, то число элементов этого ансамбля можно оценить сверху, скажем, числом 2^{1000}, и получится абсурдно малое число — меньше тысячи битов.

Видно, что в этих ситуациях колмогоровская сложность является более адекватным языком, чем шенноновская энтропия.

Сложность и случайность

Вернёмся к неравенству $KS(x) \leqslant l(x) + O(1)$ (теорема 2). Для большинства слов данной длины это неравенство близко к равенству. В самом деле, справедливо такое утверждение:

Теорема 5. *Для любого натурального числа n количество слов x, для которых $KS(x) < n$, меньше 2^n.*

◀ Пусть D — оптимальный способ описания, фиксированный при определении колмогоровской сложности. Тогда все слова вида $D(y)$ при $l(y) < n$ (и только они) имеют сложность менее n. А таких слов не больше, чем различных слов y, имеющих длину меньше n, которых всего

$$1 + 2 + 4 + 8 + \ldots + 2^{n-1} = 2^n - 1$$

штук (ведь слов длины k ровно 2^k). ▶

Отсюда легко заключить, что доля слов сложности меньше $n - c$ среди всех слов длины n меньше $2^{n-c}/2^n = 2^{-c}$. Например, доля слов сложности менее 90 среди всех слов длины 100 меньше 2^{-10}.

Таким образом, большинство слов несжимаемы или почти несжимаемы. Это можно выразить и так: почти наверняка случайно взятое слово данной длины окажется (почти) несжимаемым. Рассмотрим следующий мысленный (или даже реальный) эксперимент. Бросим монету, скажем, 80000 раз и сделаем из результатов бросаний файл длиной в 10000 байтов (8 битов образуют один байт). Можно смело держать пари, что ни один существовавший до момента бросания архиватор не сумеет сжать этот файл более чем на 10 байтов. (В самом деле, вероятность этого меньше 2^{-80} для каждого конкретного архиватора, а число различных архиваторов не так уж велико.)

Вообще естественно считать случайными несжимаемые слова: случайность есть отсутствие закономерностей, которые позволяют сжать слово. Конечно, чёткой границы между случайными и неслучайными объектами провести нельзя. Глупо интересоваться, какие именно из восьми слов длины 3 (то есть из слов 000, 001, 010, 011, 100, 101, 110, 111) случайны, а какие нет. Другой пример: начав со «случайного» слова длиной 10000, будем заменять в нём по очереди единицы на нули. В конце получится явно неслучайное слово (из одних нулей), но имеет ли смысл спрашивать, в какой именно момент слово перестало быть случайным? Вряд ли.

Более естественно определить «дефект случайности» двоичного слова x как разность $l(x) - KS(x)$. Используя эту терминологию, можно сказать так: теорема 2

утверждает, что дефект случайности почти что неотрицателен (не меньше некоторой константы), а теорема 5 гарантирует, что для случайно взятого слова данной длины n (мы считаем все слова длины n равновероятными) вероятность иметь дефект больше d не превосходит $1/2^d$.

Теперь закон больших чисел (утверждающий, что у большинства двоичных слов данной длины n доля единиц близка к $1/2$) можно попытаться перевести на язык колмогоровской сложности так: у любого слова с малым дефектом случайности доля единиц близка к $1/2$. Из этого «перевода» вытекает исходная формулировка закона (поскольку мы уже знаем, что слова с малым дефектом случайности составляют большинство); как мы впоследствии убедимся, в некотором смысле эти формулировки равносильны.

Если мы хотим провести чёткую границу между случайными и неслучайными объектами, мы должны от конечных объектов перейти к бесконечным. Определение случайности для бесконечных последовательностей нулей и единиц, данное учеником Колмогорова шведским математиком П. Мартин-Лёфом, подробно обсуждается нами в главе 3. Впоследствии К. Шнорр и Л. А. Левин нашли критерий случайности в терминах сложности: последовательность случайна тогда и только тогда, когда дефект случайности её начальных отрезков ограничен. (Правда, нужно использовать другой вариант колмогоровской сложности, так называемую *монотонную* сложность.)

Невычислимость KS и парадокс Берри

Прежде чем говорить о применениях колмогоровской сложности, скажем о препятствии, с которым сталкивается любое такое применение. Увы, функция KS невычислима: не существует алгоритма, который по данному слову x определяет его колмогоровскую сложность. Более того, не существует никакой нетривиальной (неограниченной) вычислимой нижней оценки для KS, как показывает следующая теорема.

Теорема 6. *Пусть k — произвольная (не обязательно всюду определённая) вычислимая функция из Ξ в $\mathbb{N}$. (Другими словами, вычисляющий её алгоритм применим к некоторым двоичным словам и даёт в качестве результатов натуральные числа.) Если k является нижней оценкой для колмогоровской сложности (то есть $k(x) \leqslant KS(x)$ для тех x, для которых $k(x)$ определено), то k ограничена: все её значения не превосходят некоторой константы.*

◄ Доказательство этой теоремы повторяет так называемый «парадокс Берри». Этот парадокс состоит в предложении рассмотреть

> *наименьшее число, которое нельзя определить фразой из не более чем тринадцати русских слов.*

Эта фраза как раз содержит тринадцать слов и определяет то самое число, которое нельзя определить, так что получается противоречие.

Следуя этой идее, будем искать *первое двоичное слово, колмогоровская сложность которого больше некоторого N*. С одной стороны, это слово по построению

имеет сложность больше N, с другой стороны, мы привели его короткое описание (оно включает в себя число N, но длина двоичной записи числа N гораздо меньше N). Разгадка парадокса в том, что это слово нельзя — мы получили бы противоречие, если бы могли при условии, что мы знаем, как такое слово искать. Вычислимая нижняя оценка колмогоровской сложности превращает этот парадокс в искомое доказательство.

Перейдём к формальному изложению. Пусть функция k является вычислимой нижней оценкой колмогоровской сложности. Рассмотрим функцию $B(N)$, аргументом которой является натуральное число N, вычисляемую следующим алгоритмом: «Параллельно вычислять $k(0), k(1), k(2) \ldots$ до тех пор, пока не обнаружится некоторое x, для которого $k(x) > N$. Первое из таких x и будет результатом».

Предположим, что вопреки утверждению теоремы функция k не ограничена. Тогда функция B определена для всех N, и $k(B(N)) > N$ по построению. По предположению k является нижней оценкой сложности, так что и $KS(B(N)) > N$. С другой стороны, $B(N)$ эффективно получается по двоичной записи $\mathrm{bin}(N)$ числа N, поэтому

$$KS(B(N)) \leqslant KS(\mathrm{bin}(N)) + O(1) \leqslant l(\mathrm{bin}(N)) + O(1) \leqslant \log_2 N + O(1)$$

(первое неравенство следует из теоремы 3, а второе — из теоремы 2; $O(1)$ обозначает ограниченное слагаемое). Получается, что

$$N < KS(B(N)) \leqslant \log_2 N + O(1),$$

что при бóльших N приводит к противоречию. ►

Применения колмогоровской сложности

Прежде всего оговоримся: речь пойдёт не о практических применениях, а о связи колмогоровской сложности с другими вопросами.

Бритва Оккама. Начнём с такого философского вопроса: что значит, что теория хорошо объясняет результаты эксперимента? Пусть имеется какой-то набор экспериментальных данных и разные теории, его объясняющие. Например, экспериментальными данными могут быть положения планет на небесной сфере. Их можно объяснять в духе Птолемея, рассматривая эпициклы и дифференты (и внося дополнительные поправки при необходимости), а можно ссылаться на законы современной механики. Как объяснить, чем современная теория лучше? Можно сказать так: современная теория позволяет вычислять положения планет с той же (и даже лучшей) точностью, имея меньше параметров. Условно говоря, достижение Кеплера состоит в том, что он нашёл более короткое описание для экспериментальных данных. Совсем грубо можно сказать, что экспериментаторы получают двоичные слова, после чего теоретики ищут для этих слов короткие описания (и тем самым верхние оценки на сложность), и лучше тот теоретик, у которого описание короче (оценка меньше).

Этот подход называют иногда «бритвой Оккама» по имени философа, который говорил, что не следует множить сущности без необходимости. Насколько такое толкование одобрил бы сам Оккам, сказать трудно. (Иногда говорят о MDL-принципе, от слов Minimal Description Length.)

Можно использовать ту же идею в более практической ситуации. Представим себе, что мы собираемся автоматически читать надписи на конвертах и ищем правило, отделяющее изображения нулей от изображений единиц. (Будем считать, что изображение дано в виде матрицы битов, записанной как двоичное слово.) У нас есть несколько тысяч образцов — картинок, про которые мы знаем, нуль это или единица. По этим образцам мы хотим сформулировать какое-то разумное разделяющее правило. Что означает в этом контексте слово «разумное»? Если мы просто перечислим все образцы того или другого типа, получится вполне действующее правило — действующее до тех пор, пока нам не принесут новый образец, — но оно будет слишком длинным. Естественно считать, что разумное правило должно иметь короткое описание, то есть по возможности меньшую колмогоровскую сложность.

Обоснование теории вероятностей. Речь идёт не о само́й математической теории вероятностей (которую можно рассматривать как часть теории меры и которая не требует никакого особого обоснования), а о механизме применения теории вероятностей. Представим себе, что мы бросаем монету тысячу раз (или, более реалистично, проверяем качество специального электронного датчика случайных чисел, дающего последовательность нулей и единиц). Если получится последовательность из тысячи нулей (или последовательность $0101010101\ldots$), то мы сочтём, что датчик плох. Но, собственно говоря, почему?

Часто говорят, что вероятность случайного появления тысячи нулей исчезающе мала (равна 2^{-1000}), если монета честная, и потому гипотеза честной монеты отвергается. С другой стороны, мы же не всегда бракуем датчик: возможна такая последовательность α из тысячи нулей и единиц, которая не вызовет у нас претензий к датчику. Заметим теперь, что вероятность появления последовательности α при бросании честной монеты также равна 2^{-1000} — так почему же мы в этом случае не возмущаемся? В чём разница между последовательностью из тысячи нулей и последовательностью α? Естественно усмотреть эту разницу в том, что последовательность из тысячи нулей имеет меньшую сложность, чем α.

Доказательство теорем теории вероятностей. Рассмотрим в качестве примера усиленный закон больших чисел. Он утверждает, что почти все бесконечные последовательности нулей и единиц (по равномерной бернуллиевой мере, соответствующей независимым бросаниям симметричной монеты) имеют предел частоты единиц, равный $1/2$.

Более подробно. Обозначим через Ω множество всех бесконечных последовательностей нулей и единиц. На этом множестве вводится равномерная бернуллиева мера. Это делается так: для каждого двоичного слова x рассмотрим множество Ω_x всех бесконечных продолжений этого слова. (Например, $\Omega_\Lambda = \Omega$.) Меру множества Ω_x положим равной $2^{-l(x)}$. (Например, мера множества Ω_{01}, состоящего из последовательностей, начинающихся на 01, равна $1/4$.)

Для каждой последовательности $\omega = \omega_0\omega_1\omega_2\ldots$ рассмотрим предел частоты единиц в начальных отрезках, то есть предел

$$\lim_{n\to\infty} \frac{\omega_0 + \omega_1 + \ldots + \omega_{n-1}}{n}.$$

Говорят, что последовательность удовлетворяет *усиленному закону больших чисел*, если этот предел существует и равен $1/2$. Например, последовательность $010101\ldots$ (с периодом 2) ему удовлетворяет, а последовательность $011011011\ldots$ (с периодом 3) — нет.

Усиленный закон больших чисел гласит, что множество последовательностей, не удовлетворяющих этому закону, имеет меру нуль. Напомним: множество $A \subset \Omega$ имеет меру нуль, если для всякого $\varepsilon > 0$ можно указать последовательность слов $x_0, x_1, x_2, \ldots$, для которой

$$A \subset \Omega_{x_0} \cup \Omega_{x_1} \cup \Omega_{x_2} \cup \ldots$$

и сумма ряда

$$2^{-l(x_0)} + 2^{-l(x_1)} + 2^{-l(x_2)} + \ldots$$

(состоящего из мер множеств Ω_{x_i}) не превосходит ε.

Усиленный закон больших чисел можно доказать с помощью алгоритмической теории информации, используя понятие случайной по Мартин-Лёфу последовательности (о котором мы уже упоминали). Доказательство (см. раздел 8.4) состоит из двух частей. Сначала мы доказываем, что всякая случайная по Мартин-Лёфу последовательность удовлетворяет этому закону. Это можно сделать с использованием критерия случайности Левина – Шнорра (если предел частот не существует или не равен $1/2$, то некоторые начальные отрезки имеют сложность меньше, чем должно быть для случайной последовательности; см. теорему 90 на с. 167) Вторая часть доказательства носит общий характер и не зависит от того, какой закон теории вероятностей рассматривается. Мы доказываем, что множество неслучайных (по Мартин-Лёфу) последовательностей имеет меру нуль. После этого мы говорим, что множество последовательностей, не удовлетворяющих закону больших чисел, содержится в множестве меры нуль и потому само имеет меру нуль.

Понятие случайной последовательности имеет и самостоятельный философский интерес. В начале XX века Рихард фон Мизес предложил положить это понятие (он называл его по-немецки Kollektiv) в основу теории вероятностей (в то время ещё не было идеи рассматривать теорию вероятностей в рамках теории меры). Подход Мизеса к определению случайной последовательности (он считал главным свойством так называемую «устойчивость частот») и его современные варианты мы рассматриваем в главе 9.

Нижние оценки в теории вычислительной сложности. Колмогоровская сложность оказалась полезным техническим средством для доказательства нижних оценок в теории сложности вычислений. Мы опишем её применение на модельном примере.

Рассмотрим следующую задачу: на ленте одноленточной машины Тьюринга написано некоторое слово x длины n. Требуется скопировать это слово справа от

него, то есть получить на ленте слово xx. С середины 1960-х годов известно, что это требует времени, пропорционального n^2. Точнее говоря, можно доказать, что для любой машины Тьюринга M, решающей эту задачу для всех слов x, найдётся такое число $\varepsilon > 0$, что для любого n существует слово x длины n, копирование которого (с помощью M) длится дольше εn^2.

Имеется следующее интуитивное объяснение этого факта. Машина Тьюринга имеет ограниченное число внутренних состояний, то есть может запомнить (помимо ленты) ограниченное число битов. Кроме того, скорость передвижения её по ленте ограничена: за один шаг она может сместиться не более чем на одну клетку. Следовательно, скорость переноса информации с помощью машины Тьюринга (единица измерения: *бит · клетка/шаг*) ограничена константой, зависящей от числа состояний машины (пропорциональной логарифму числа состояний). Копирование слова x длины n требует переноса n битов информации, содержащихся в слове x, на расстояние в n клеток, и потому число шагов пропорционально n^2.

Это рассуждение можно сделать вполне строгим, используя понятие колмогоровской сложности. Видно сразу же, что трудными для копирования будут слова, содержащие много информации (сложность которых близка к максимуму, то есть к n).

Мы рассмотрим этот пример подробнее в разделе 8.2 (с. 262).

Комбинаторный смысл утверждений о сложности. Мы приведём один пример такого рода (подробнее см. главу 10, с. 349). Существует неравенство для сложностей трёх слов и их комбинаций:

$$2\,KS(xyz) \leqslant KS(xy) + KS(xz) + KS(yz) + O(\log n)$$

для любых трёх слов x, y, z длины не больше n.

Оказывается, что это неравенство имеет естественные интерпретации, в которых о сложности уже ничего не говорится. В частности, из него можно вывести такой геометрический факт [52].

Пусть имеется тело в трёхмерном пространстве с перпендикулярными осями OX, OY и OZ, имеющее объём V. Рассмотрим три его ортогональные проекции на плоскости OXY, OXZ и OYZ. Пусть S_{xy}, S_{xz} и S_{yz} — площади этих проекций. Тогда имеет место такое неравенство:

$$V^2 \leqslant S_{xy} \cdot S_{xz} \cdot S_{yz}.$$

А вот алгебраическое следствие того же неравенства: для произвольной группы G и её подгрупп X, Y и Z справедливо неравенство

$$|X \cap Y \cap Z|^2 \geqslant \frac{|X \cap Y| \cdot |X \cap Z| \cdot |Y \cap Z|}{|G|}$$

(где $|H|$ обозначает число элементов в группе H).

Теорема о неполноте. Покажем, следуя Г. Чейтину, как можно использовать теорему 6, чтобы доказать известную теорему Гёделя о неполноте. Эта теорема

гласит, что не все истинные утверждения достаточно богатой математической теории (скажем, формальной арифметики или теории множеств) можно доказать в этой теории.

Будем считать, что в нашей теории можно записать утверждение $KS(x) > n$ для любого двоичного слова x и любого натурального числа n. (Утверждение $KS(x) > n$, построенное для данного слова x и для данного числа n, гласит, что выбранный декомпрессор D не даёт x ни на каком входе длины не более n; его легко записать в формальной арифметике и тем более в теории множеств.)

Будем перебирать все формальные выводы и искать среди них доказательства утверждений вида $KS(x) > n$ (для всевозможных слов x и чисел n). Найдя такое утверждение, мы смотрим, каково в нём значение n (чем оно больше, тем лучше). Если n превосходит прежний рекорд, то мы запоминаем в таблице рекордов это n и соответствующее ему значение x_n. Теперь есть две возможности. Либо таблица рекордов будет расти неограниченно, либо начиная с некоторого времени новых рекордов не будет и какое-то утверждение $KS(X) > N$ так и останется непревзойдённым. Во втором случае целый класс истинных утверждений, а именно, все истинные утверждения вида $KS(x) > n$ при $n > N$, являются недоказуемыми. (Напомним, что по теореме 5 такие истинные утверждения существуют, и их бесконечно много.)

В первом случае мы получим вычислимую последовательность слов $x_0, x_1, x_2 \ldots$ и чисел $n_0 < n_1 < n_2 < \ldots$, для которых доказуемы утверждения $KS(x_i) > n_i$. Мы предполагаем, что в теории доказуемы лишь истинные утверждения, так что и вправду $KS(x_i) > n_i$. Выбросив из этой последовательности повторения слов, можно считать, что все x_i различны (каждое слово может встречаться лишь конечное число раз, так как $n_i \to \infty$ при $i \to \infty$). После этого вычислимая функция k, для которой $k(x_i) = n_i$, будет неограниченной нижней оценкой для колмогоровской сложности, что противоречит теореме 6.

1. Простая колмогоровская сложность

1.1. Определение и основные свойства

Напомним (данное во введении) определение колмогоровской сложности. Этот вариант определения сложности был предложен в статье Колмогорова [63]. Чтобы отличить его от других, определённую таким образом сложность называют *простой*, или *обыкновенной*, колмогоровской сложностью. Другие виды сложности (префиксная, монотонная) будут рассматриваться, начиная с главы 4, так что пока мы можем говорить о колмогоровской сложности (без уточняющего эпитета), не опасаясь путаницы.

Способом описания, или *декомпрессором*, мы называли произвольное вычислимое частичное отображение D из множества двоичных слов Ξ в себя. (Вычислимость отображения D означает, что есть алгоритм, который применим к словам из области определения отображения D и только к ним; результат применения алгоритма к слову x есть $D(x)$.) Если $D(y) = x$, говорят, что y является *описанием* x *при способе описания* D.

Для каждого способа описания D мы определяем *сложность относительно этого способа описания*, полагая её равной длине кратчайшего описания:

$$KS_D(x) = \min\{l(y) \mid D(y) = x\}.$$

При этом минимум пустого множества считается равным $+\infty$.

Говорят, что способ описания D_1 *не хуже* способа описания D_2, если найдётся такая константа c, что $KS_{D_1}(x) \leqslant KS_{D_2}(x) + c$ для всех слов x. (Краткая запись: $KS_{D_1}(x) \leqslant KS_{D_2}(x) + O(1)$.)

Способ описания называют *оптимальным*, если он не хуже любого другого способа описания. Теорема Колмогорова – Соломонова (с. 11) утверждает, что существуют оптимальные способы описания. Её доказательство (подробно изложенное во введении) проходит так. Выберем какой-либо универсальный язык программирования. Пусть U — интерпретатор этого языка: $U(p, x)$ есть результат работы программы p на входе x (программа и вход — двоичные слова). Далее мы полагаем

$$D(\hat{p}x) = U(p, x),$$

где вычислимое отображение $p \mapsto \hat{p}$ выбрано так, чтобы по слову $\hat{p}$ можно было определить p, а также место, где $\hat{p}$ кончается. (В этом случае слово $\hat{p}$ не может быть началом слова $\hat{q}$ при $p \neq q$, и это свойство гарантирует, что D корректно определено.) Тогда для любого способа описания D' имеем

$$KS_D(x) \leqslant KS_{D'}(x) + l(\hat{p}),$$

где p — программа, соответствующая способу описания D'. (В самом деле, если y есть описание x относительно D', то $\hat{p}y$ есть описание x относительно D.)

Мы фиксируем некоторый оптимальный способ описания и сложность слова x относительно этого способа описания обозначаем $KS(x)$ (без индекса). В первой статье Колмогорова [63] использовалось обозначение $K(x)$; во второй [64] та же величина обозначалась $H(x)$. Сейчас в англоязычной литературе наиболее часто используется обозначение $C(x)$.

Сравнение оптимального способа описания с тождественным (при котором слово является своим собственным описанием) показывает, что $KS(x) \leqslant l(x) + O(1)$ (с. 14).

Сравнивая оптимальный способ описания D со способом $y \mapsto A(D(y))$, где A — произвольная вычислимая функция, находим, что $KS(A(x)) \leqslant KS(x) + O(1)$ (невозрастание сложности при алгоритмических преобразованиях, с. 14).

Последнее свойство позволяет говорить не только о сложности слов, но и о сложности других «конструктивных объектов» (натуральных чисел, рациональных чисел, перестановок, конечных множеств слов и т. п.) — любых объектов, которые можно естественным образом закодировать двоичными словами. А именно, мы определяем сложность объекта как сложность кодирующего слова при некотором (однозначном вычислимом) кодировании.

Поясним сказанное на примере натуральных чисел. Натуральное число n можно записать в двоичной системе — получится двоичное слово. Можно выбрать и другой способ представления натуральных чисел двоичными словами. Например, можно записать двоичные слова подряд (в порядке возрастания длин и в лексикографическом порядке для слов одной длины):

$$\Lambda, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, \ldots$$

и сопоставить их (в этом порядке) с натуральными числами $0, 1, 2, 3, \ldots$ (такое отождествление часто удобно, поскольку, в отличие от двоичной записи, оно является взаимно однозначным). Наконец, натуральное число n можно изображать словом из n единиц. Так вот, какой бы из этих способов ни выбрать для определения сложности натуральных чисел, мы получим функции сложности, отличающиеся не более чем на константу. В самом деле, существуют алгоритмы, переводящие натуральные числа из одной записи в другую, а применение алгоритма увеличивает сложность не более чем на константу.

Заметим, что колмогоровская сложность и так определена с точностью до ограниченного слагаемого, так что дополнительный произвол такого рода нам не страшен.

Отметим ещё, что двоичная запись натурального числа n имеет длину $\log n + O(1)$, и потому его сложность также не превосходит $\log n + O(1)$. (Здесь, напомним, $\log$ обозначает двоичный логарифм.)

Другое простое применение свойства невозрастания сложности при алгоритмических преобразованиях: покажем, что добавление или удаление нуля или единицы в конце слова меняет его сложность не более чем на константу. В самом деле,

функции $x \mapsto x0$, $x \mapsto x1$, а также функция, удаляющая последний бит слова, вычислимы.

Разумеется, то же самое верно для добавления бита в начале слова. Но изменение бита в произвольном месте слова может изменить его сложность более чем на константу. Например, если слово состоит из 2^n нулей, то его сложность не больше $KS(n) + O(1) \leqslant \log n + O(1)$. Заменив k-й нуль на единицу (при $k = 1, 2, \ldots, 2^n$), мы получим 2^n различных слов, поэтому сложность хотя бы одного из них не меньше n. (Как мы видели на с. 17, число слов сложности меньше n не превосходит числа всех описаний длины меньше n и потому меньше 2^n.)

Прибавление единицы к натуральному числу n меняет $KS(n)$ не более чем на константу, так что сложность, рассматриваемая как функция натурального аргумента, обладает «свойством Липшица» (это значит, что $|KS(m) - KS(n)| \leqslant c|m - n|$ для некоторого c и любых натуральных чисел m и n).

$\boxed{1}$ Докажите более сильное неравенство: $|KS(m) - KS(n)| \leqslant |m - n| + c$ (для некоторого c и для всех $m, n \in \mathbb{N}$) и даже $|KS(m) - KS(n)| \leqslant 2\log|m - n| + c$ (при $m \neq n$).

Вернёмся к утверждению о том, что число слов x с $KS(x) < n$ меньше 2^n. Заметим, что в этом утверждении нет никаких констант (что несколько необычно). Но всё равно зависимость от выбора оптимального способа описания в нём неявно присутствует: если мы заменим один способ описания на другой, то количество слов сложности меньше n по-прежнему будет меньше 2^n, но это будут другие слова!

$\boxed{2}$ Покажите, что число слов сложности меньше n заключено между 2^{n-c} и 2^n (при некотором c и всех n). [Указание: о верхней оценке 2^n мы уже говорили, а нижняя вытекает из того, что $KS(x) \leqslant l(x) + c$ при некотором c и потому все слова длины меньше $n - c$ имеют сложность меньше n.]

Покажите, что число слов сложности ровно n не превосходит 2^n, но для некоторых способов описания может быть сильно меньше (например, таких слов может не быть вовсе для бесконечно многих n). [Указание: модифицируем оптимальный способ описания, добавив 0 или 11 ко всем описаниям и сделав их длину чётной.]

$\boxed{3}$ Покажите, что среднее арифметическое сложностей всех слов длины n равно $n + O(1)$. [Указание: ряд $\sum k/2^k$ сходится, а доля слов длины n, сложность которых равна $n - k$, равна примерно 2^{-k}.]

Продолжая эту тему (связь оценок сложности с оценками количества), докажем такое важное утверждение:

Теорема 7. (а) *Множества $S_n = \{x \mid KS(x) < n\}$ образуют перечислимое семейство, причём $|S_n| < 2^n$ при всех n. (Здесь $|S_n|$ — число элементов в множестве S_n.)*

(б) *Если V_n $(n = 0, 1, 2, \ldots)$ — перечислимое семейство множеств, причём $|V_n| < 2^n$ при всех n, то найдётся такое c, что $KS(x) < n + c$ для любого n и для любого $x \in V_n$.*

Поясним, что такое перечислимое семейство множеств. Говорят, что множество (слов, натуральных чисел или иных конструктивных объектов) *перечислимо*, если

существует алгоритм, который порождает элементы этого множества (программа, печатающая их один из другим, возможно с перерывами, и никогда не останавливающаяся). Повторения разрешаются (но это не важно, так как можно вычёркивать повторяющиеся элементы.) Если множество конечно, программа может с некоторого момента ничего не печатать (хотя наблюдатель не узнает, наступил этот момент или нет).

Например, множество всех n, при которых в десятичном разложении $\sqrt{2}$ встречается ровно n девяток подряд, перечислимо. Перечисляющий его алгоритм таков: вычисляем последовательные цифры числа $\sqrt{2}$; как только встретилась группа из n девяток подряд, окружённая не-девятками, печатаем n и продолжаем работу.

Перечислимость семейства множеств V_n означает (по определению), что множество пар $\{\langle n, x\rangle \mid x \in V_n\}$ перечислимо. Это, вообще говоря, более сильное свойство, чем перечислимость каждого из множеств V_n. Если семейство перечислимо, то каждое из множеств перечислимо (отбираем пары с данным первым членом), но обратное, вообще говоря, неверно. Например, если все множества V_n конечны (как у нас), то все они перечислимы, но это не гарантирует перечислимости семейства. (Можно проверить, что перечислимость семейства равносильна возможности по n алгоритмически получать алгоритм, перечисляющий V_n.) О понятии перечислимого множества можно прочесть в любом учебнике по теории вычислимости, например, в [175].

◀ Покажем, что множество пар

$$\{\langle n, x\rangle \mid x \in S_n\} = \{\langle n, x\rangle \mid KS(x) < n\}$$

(первыми компонентами пар являются натуральные числа, вторыми — двоичные слова) перечислимо. В самом деле, пусть D — оптимальный способ описания, использованный при определении KS. Будем параллельно запускать D на всех двоичных словах, делая всё больше шагов на каждом слове и вовлекая всё новые и новые слова. (Например, на k-м этапе мы проводим по k шагов вычисления для первых k двоичных слов.) Как только одно из проводимых вычислений заканчивается и обнаруживается, что $D(y) = x$, перечисляющий алгоритм выдаёт на выход пару $\langle l(y) + 1, x\rangle$ (в самом деле, мы установили, что сложность слова x меньше $l(y) + 1$, поскольку это слово имеет описание y). После этого он выдаёт пары $\langle l(y) + 2, x\rangle, \langle l(y) + 3, x\rangle \ldots$ (чередуя их с другими парами, которые нужно выдать).

Для знакомых с теорией алгоритмов всё это объяснение можно заменить одной строчкой

$$KS(x) < n \;\Leftrightarrow\; \exists y\ (l(y) < n \wedge D(y) = x)$$

(множество в правой части перечислимо, так как график вычислимой функции D перечислим, а пересечение и проекция сохраняют перечислимость).

Более содержательно обратное утверждение. Пусть имеется перечислимое семейство конечных перечислимых множеств V_n, причём $|V_n| < 2^n$. Построим способ описания D_V следующим образом. Зарезервируем слова длины n для описания элементов множества V_n. Именно, будем считать k-е в лексикографическом порядке

слово длины n описанием k-го элемента множества V_n (в порядке появления элементов множества V_n при перечислении). Поскольку $|V_n| < 2^n$, слов длины n хватит (естественно, при повторном появлении слова в перечислении мы не выделяем нового описания). Построенный таким образом способ описания D_V будет вычислимым. В самом деле, чтобы вычислить $D_V(y)$, мы находим порядковый номер слова y среди всех слов длины $l(y)$. Пусть он равен k. После этого мы запускаем алгоритм, перечисляющий множество пар $\langle n, x\rangle$, у которых $x \in V_n$, и ожидаем появления k различных пар, первая компонента которых равна $l(y)$. Вторая компонента последней из них и будет $D_V(y)$.

По определению $KS_{D_V}(x) \leqslant n$ для любого $x \in V_n$. Сравнивая способ описания D_V с оптимальным, находим, что найдётся такое c, что $KS(x) < n + c$ для любого $x \in V_n$. Теорема 7 доказана. ▶

Интуитивный смысл доказанной только что теоремы можно объяснить так: она говорит, что утверждения «объектов определённого вида мало» (меньше 2^i) и «объекты этого вида просты» (имеют сложность меньше i) равносильны, если мы рассматриваем перечислимые семейства и измеряем сложность с точностью до аддитивной константы (а число элементов — с точностью до мультипликативной).

Эту же теорему можно сформулировать в других терминах. Пусть функция $f(x)$ определена на всех двоичных словах и принимает в качестве значений натуральные числа, а также специальный символ $+\infty$. Функция f называется *перечислимой сверху*, если существуют вычислимая функция $\langle x, k\rangle \mapsto F(x, k)$, определённая для всех слов x и всех натуральных чисел k, для которой

$$F(x, 0) \geqslant F(x, 1) \geqslant F(x, 2) \geqslant \ldots$$

и

$$f(x) = \lim_{k\to\infty} F(x, k).$$

при всех x. Значениями функции F также могут быть натуральные числа и $+\infty$. Наши требования гарантируют, что при любом k значение $F(x, k)$ является верхней оценкой для $f(x)$. Эта оценка уточняется с ростом k. При каждом x эта оценка в какой-то момент становится точной, но когда именно, мы можем не знать (если есть алгоритм, указывающий этот момент, то функция f вычислима).

Всякая вычислимая функция перечислима сверху.

Несложно проверить, что функция f перечислима сверху тогда и только тогда, когда множество

$$G_f = \{\langle x, n\rangle \mid f(x) < n\},$$

называемое иногда «надграфиком» функции f, перечислимо. (Это объясняет несколько странное название «перечислимая сверху».)

Убедимся в этом. Если функция f перечислима сверху и F — соответствующая функция двух аргументов, то

$$f(x) < n \ \Leftrightarrow\ \exists k\ F(x, k) < n.$$

Поэтому, вычисляя $F(x, k)$ параллельно для всех x и k, мы можем перечислять множество G_f. Обратно, если мы можем перечислять множество G_f, то $F(x, k)$

можно определить как наилучшую верхнюю оценку для f, которую мы можем дать после k шагов перечисления множества G_f. (В ходе этого перечисления мы устанавливаем, что $f(x) < n$ для некоторых x и n, тем самым получая верхние оценки для значения $f(x)$ при некоторых x; для остальных x мы полагаем $F(x, k) = +\infty$.)

Теперь мы можем переформулировать только что доказанную теорему 7 следующим образом:

Теорема 8. (а) *Функция KS перечислима сверху, причём $|\{x \mid KS(x) < n\}| < 2^n$ при всех n.*

(б) *Если функция K перечислима сверху и $|\{x \mid K(x) < n\}| < 2^n$ при всех n, то найдётся такое c, что $KS(x) < K(x) + c$ для всех x.*

Отметим, что в пункте (б) этой теоремы можно ослабить оценку и написать, что $|\{x \mid K(x) < n\}| = O(2^n)$.

Таким образом, колмогоровскую сложность можно определить как минимальную (с точностью до константы) перечислимую сверху функцию K, для которой $|\{x \mid K(x) < n\}| = O(2^n)$.

Можно сделать ещё один шаг и избавиться от требования минимальности, получив следующее «аксиоматическое» определение колмогоровской сложности [150].

Теорема 9. *Пусть K — функция с натуральными значениями, определённая на всех двоичных словах. Пусть при этом*:

(а) *K перечислима сверху*; [аксиома перечислимости]

(б) *для любой вычислимой функции A, аргументами и значениями которой являются двоичные слова, выполнено неравенство $K(A(x)) \leqslant K(x) + c$ при некотором c и всех x, для которых $A(x)$ определено*; [аксиома невозрастания сложности]

(в) *количество слов x, для которых $K(x) < n$, заключено между 2^{n-c_1} и 2^{n+c_2} при некоторых c_1, c_2 и при всех n.* [аксиома калибровки]

Тогда $K(x) = KS(x) + O(1)$ (другими словами, разность $K(x) - KS(x)$ ограничена с обеих сторон константами, не зависящими от x).

◄ Предыдущая теорема показывает, что $KS(x) \leqslant K(x) + O(1)$. Остаётся доказать, что $K(x) \leqslant KS(x) + O(1)$.

Лемма 1. *Существует константа c и вычислимая последовательность конечных множеств двоичных слов*

$$M_0 \subset M_1 \subset M_2 \subset \ldots$$

(говоря о вычислимости, мы имеем в виду, что множество задаётся списком своих элементов), в которой M_i содержит ровно 2^i слов и $K(x) \leqslant i + c$ для всех элементов $x \in M_i$ (при любом i).

Доказательство. Свойство (в) функции K гарантирует, что множество $A_i = \{x \mid K(x) < i + c\}$ содержит не менее 2^i элементов (при некотором c, например $c = c_1$, и при всех i). Свойство (а) гарантирует, что семейство множеств A_i перечислимо. Если оставить от множества A_i только 2^i элементов, которые появляются при перечислении первыми, то получится множество B_i, которое содержит ровно 2^i элементов. При этом список элементов множества B_i можно получить алгоритмически по i (дождавшись появления 2^i элементов при перечислении множества A_i —

наши предположения гарантируют, что это обязательно произойдёт). Правда, множества B_i не обязаны возрастать с ростом i, но это можно исправить, определив M_i индуктивно. Именно, M_0 полагаем равным B_0, а M_{i+1} есть объединение M_i с 2^i элементами B_{i+1}, не входящими в M_i. Лемма 1 доказана.

Лемма 2. *Существует константа c, при которой $K(x) \leqslant l(x) + c$ для любого слова x (напомним, $l(x)$ — длина слова x).*

Доказательство. Рассмотрим множества M_i из предыдущей леммы. Рассмотрим вычислимую функцию A, которая определена на объединении всех множеств M_i и отображает $M_{i+1} \setminus M_i$ на множество всех слов длины i. (В множестве $M_{i+1} \setminus M_i$ как раз 2^i слов, так что его можно взаимно однозначно отобразить на слова длины i.) По условию (б) мы знаем, что $K(A(y)) \leqslant K(y) + c'$ при некотором c' и всех x. Применяя это свойство для $y \in M_{i+1} \setminus M_i$, находим, что $K(x) < i + c$ для любого слова x длины i (при всех i и при некотором c, не зависящем от i). Лемма 2 доказана.

Теперь уже легко завершить доказательство теоремы. Пусть D — оптимальный способ описания. Если p — кратчайшее описание для слова x, то $K(x) = K(D(p)) \leqslant K(p) + O(1) \leqslant l(p) + O(1) = KS(x) + O(1)$. (Таким образом, мы дважды использовали свойство (б): один раз при доказательстве леммы 2 и второй раз только что, применяя его к вычислимой функции D.) ►

$\boxed{4}$ Покажите, что если в качестве описаний использовать слова в четырёхбуквенном алфавите (скажем, конечные последовательности цифр $0, 1, 2, 3$), то сложность (измеряемая как длина кратчайшего описания) будет равна половине обычной (с точностью до аддитивной константы).

$\boxed{5}$ (Продолжение) Сформулируйте и докажите аналогичное утверждение для n-буквенного алфавита.

$\boxed{6}$ Пусть $f\colon \mathbb{N} \to \mathbb{N}$ — всюду определённая вычислимая возрастающая функция, причём $\liminf f(n+1)/f(n) > 1$. Докажите, что если A_n — перечислимое семейство конечных множеств, причём число элементов в A_n не превосходит $f(n)$ при всех n, то найдётся такая константа c, что $KS(x) \leqslant \log f(n) + c$ для любого n и для любого $x \in A_n$.

$\boxed{7}$ Покажите, что при некотором c для всякого слова x любой длины n существует слово y той же длины, отличающееся от x не более чем в одной позиции, для которого $KS(y) \leqslant n - \log n + c$. [Указание. При $n = 2^k - 1$ можно взять в качестве y кодовое слово кода Хемминга (см, например, [133]). Если число n не имеет вида $2^k - 1$, применяем те же рассуждения к части x' слова x, имеющей удобную (и достаточно большую) длину.]

1.2. Алгоритмические свойства

Как мы видели, функция KS является перечислимой сверху, но не является вычислимой и даже не имеет вычислимых неограниченных нижних оценок (теорема 6, с. 18).

Заметим, что из этого вытекает, что никакой оптимальный способ описания не является всюду определённым (имеются слова, не являющиеся описаниями). В самом деле, если бы оптимальный способ описания D был всюду определён, то мы могли бы вычислить $KS_D(x)$, просто перепробовав все описания в порядке возрастания длин (до нахождения кратчайшего).

При этом возникает следующий любопытный парадокс. С точки зрения оптимальности наличие слов, не являющихся описаниями, явно невыгодно. В самом деле, если $D(y)$ не определено, можно рассмотреть другой способ описания D', для которого $D'(y)$ равно некоторому слову z, а в остальном D' совпадает с D. При замене D на D' сложность слова z может уменьшиться, а сложность остальных слов не изменится. Тем не менее для оптимального способа описания всегда есть слова, на которых он не определён!

Формального противоречия тут нет (такое доопределение сохраняет вычислимость, лишь если его сделать в одной точке или в конечном числе точек), но наблюдение это любопытно (его сделал Ю. И. Манин в своей книжке «Вычислимое и невычислимое» [96] — той самой, в которой он обсуждал возможности квантовых компьютеров задолго до того, как их начали изучать).

Отсюда следует, что область определения оптимального способа описания не может быть разрешимым множеством. (Множество слов называется *разрешимым*, если есть алгоритм, который по любому слову выясняет, принадлежит ли оно этому множеству.) В самом деле, если бы существовал алгоритм, выясняющий, определено ли $D(x)$ или нет, то D можно было бы продолжить до всюду определённого оптимального способа описания (положив, скажем $D(x) = 0$ для тех точек, где $D(x)$ было неопределённым).

Тем самым мы построили алгоритм с неразрешимой областью определения (это — центральный факт теории алгоритмов, см., например, [175]).

Вообще понятие колмогоровской сложности представляет интерес с точки зрения общей теории алгоритмов. В последние годы изучение этих связей дало множество интересных (и трудных) результатов, о которых можно прочесть в [127, 37]. Мы же рассмотрим только два вопроса такого рода — о простоте множества простых слов и о сложности больших чисел.

1.2.1. Простые слова и простые множества

Слово «простые» в этом разделе будет иметь совершенно разные значения в применении к словам и множествам. Говоря о простых словах, мы будем иметь в виду слова малой колмогоровской сложности. Понятие же простого множества, которое мы будем использовать, введено американским логиком Эмилем Постом и никакого отношения к колмогоровской сложности не имеет. (Трудно даже сказать, почему был выбран такой термин.)

Определение. Перечислимое множество A является *простым* (в смысле Поста), если его дополнение бесконечно, но не содержит бесконечного перечислимого подмножества.

Будем называть слово x «простым», если $KS(x) < l(x)/2$.

Теорема 10. *Множество всех «простых» слов является простым в смысле Поста.*

◀ Множество S всех «простых» слов перечислимо, поскольку функция KS вычислима сверху (используя всё более и более точные верхние оценки, мы рано или поздно сможем обнаружить и перечислить любое «простое» слово).

Как мы знаем, число слов сложности менее $n/2$ не превосходит $2^{n/2}$. Поэтому среди слов длины n «простые» слова составляют (ничтожное) меньшинство, так что дополнение к множеству S бесконечно.

Может ли дополнение к множеству S иметь бесконечное перечислимое подмножество? Нет. Предположим, что такое подмножество C существует, и покажем, что функция KS имеет неограниченную вычислимую нижнюю оценку. В самом деле, чтобы найти слово сложности больше t, достаточно перечислять множество C, пока не обнаружится слово c_t длины больше $2t$. Такое слово найдётся, поскольку C бесконечно; это слово должно иметь сложность больше t, иначе оно было бы «простым». Можно считать, что все c_t различны (отбраковывая уже использованные слова). Тогда функция, равная t на слове c_t, будет вычислимой неограниченной нижней оценкой для KS, а такое невозможно по теореме 6 (с. 18). ▶

Заметим, что в этой теорема граница $l(x)/2$ выбрана произвольно: с тем же успехом можно было бы называть слово «простым», если $KS(x) < l(x) - 1$ (или, скажем, $KS(x) < \log\log l(x)$).

1.2.2. Сложность бо́льших чисел

Будем рассматривать сложность $KS(m)$ как функцию натурального числа m, отождествляя каждое двоичное слово с его порядковым номером. Легко понять, что предел $KS(m)$ при $m \to \infty$ равен бесконечности, так как для любого c лишь конечное число объектов могут иметь сложность меньше c. Но это стремление к бесконечности не является эффективно вычислимым: нет алгоритма, который по данному числу n указывал бы то N, начиная с которого колмогоровская сложность становится больше n. (Мы убедились в этом, когда говорили о вычислимых нижних оценках для KS, см. с. 18.)

В этом разделе мы изучим вопрос о скорости стремления KS к бесконечности более подробно. Для этого, следуя Чейтину [22], рассмотрим функцию

$$B(n) = \max\{m \in \mathbb{N} \mid KS(m) \leqslant n\}$$

(наибольшее число, сложность которого не превосходит n). Эту функцию можно назвать «регулятором сходимости» $KS(m) \to \infty$ при $m \to \infty$, поскольку $K(x) > n$ при $x > B(n)$. (Формально говоря, для малых значений n может оказаться, что $KS(m) > n$ при любом натуральном m. В этом случае можно считать, например, что $B(n) = -1$.)

Функцию B можно назвать в каком-то смысле обратной к функции $KS_{\geqslant}(N) = \min\{KS(m) \mid m \geqslant N\}$. Значение $KS_{\geqslant}(N)$ можно интерпретировать как сложность задачи «указать какое-либо число, не меньшее N». Функция $KS_{\geqslant}$ медленно растёт, принимая значение n на участке $(B(n-1), B(n)]$.

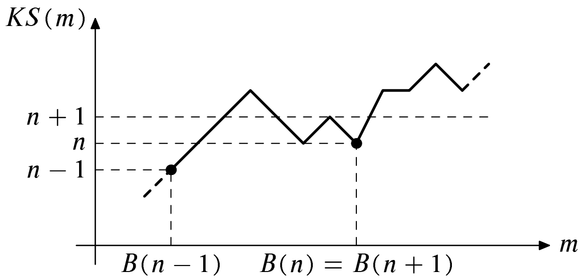


Рис. 1.1. К определению функции B: в точке $m = B(n-1)$ значение функции KS не превосходит $n-1$ (на рисунке показан случай равенства), а всюду правее оно больше $n-1$. В точке $m = B(n)$ значение функции KS не превосходит n (показан случай равенства), а правее оно больше n (на рисунке показан случай, когда оно даже больше $n+1$, и поэтому $B(n+1) = B(n)$). При $m \in (B(n-1), B(n)]$ значение $KS_{\geqslant}(m)$ равно n.

Медленный рост $KS_{\geqslant}$ соответствует быстрому росту функции B. О быстроте роста B говорит такой простой факт.

Теорема 11. *Пусть f — произвольная вычислимая функция с натуральными аргументами и значениями. Тогда $B(n) \geqslant f(n)$ для всех n, кроме конечного числа.*

(Заметим, что мы не требуем, чтобы $f(n)$ было определено при всех n. Говорится лишь, что для тех достаточно больших n, при которых $f(n)$ определено, выполнено неравенство $B(n) \geqslant f(n)$.)

◀ Алгоритмическое преобразование не увеличивает сложности, поэтому

$$KS(f(n)) \leqslant KS(n) + O(1) \leqslant \log n + c$$

для некоторого c и для всех n, при которых $f(n)$ определено. С другой стороны, $f(n) > B(n)$ влечёт $KS(f(n)) > n$ (по определению функции B). Получается, что для таких значений n выполняются неравенства

$$n < KS(f(n)) \leqslant \log n + c,$$

что возможно лишь для конечного числа значений n. ▶

Определение $B(n)$ можно переформулировать так: пусть D — оптимальный способ описания, использованный при определении сложности; тогда $B(n)$ есть максимальное значение D на всех словах длины не больше n:

$$B(n) = \max\{D(x) \mid l(x) \leqslant n\}.$$

Напомним, что мы отождествляем слова с натуральными числами и считаем, что значениями D являются натуральные числа. При этом максимум пустого множества следует считать равным -1.

Можно вместо D рассмотреть любую вычислимую функцию d, аргументами которой являются слова, а значениями — натуральные числа, и рассмотреть функцию

$$B_d(n) = \max\{d(x) \mid l(x) \leqslant n \text{ и } d(x) \text{ определено}\}.$$

Рассматриваемая нами функция B является наибольшей среди них в следующем смысле:

Теорема 12. *Для всякой функции d найдётся такая константа c, что*

$$B_d(n) \leqslant B(n+c)$$

при всех n.

◀ В самом деле, значение $d(x)$ на слове x длины не более n имеет сложность не более $n + c$ для некоторой константы c, поскольку применение вычислимой функции d увеличивает сложность не более чем на константу, а $KS(x) \leqslant n + O(1)$. Другими словами, $d(x)$ — одно из чисел, сложность которых не больше $n + c$, и не превосходит наибольшего такого числа. ▶

Это (тривиальное) наблюдение окажется нам полезным в таком частном случае. Пусть M — некоторый алгоритм. Мы говорим, что умеем решать *проблему остановки* для алгоритма M и какого-то множества входных слов, если по любому входному слову x из этого множества мы можем определить, останавливается ли M на x или не останавливается.

Классический результат теории алгоритмов (с которого началась эта теория) состоит в том, что для некоторых алгоритмов проблема остановки (для данного алгоритма и произвольного входа) неразрешима.

Нас сейчас будет интересовать эта проблема для слов ограниченной длины. Итак, пусть фиксирован алгоритм M. Рассмотрим функцию $t(x)$, которая есть время работы алгоритма M на слове x. (Если алгоритм M не останавливается на данном x, то $t(x)$ не определено, так что область определения t совпадает с областью определения алгоритма M.) Тогда $B_t(n)$ есть наибольшее время работы алгоритма M на словах длины не более n. Зная число $B_t(n)$ или любое большее его число, мы можем решать проблему остановки для машины M и любого входа x длины не более n: надо подождать указанное время; если алгоритм не остановился за это время, то он не остановится никогда.

Как мы видели, $B_t(n) \leqslant B(n + c)$ для некоторого c (зависящего от M), и зная $B(n + c)$ или любое большее число, мы можем решать проблему остановки для машины M и входов длины не более n. Сформулируем это замечание в виде теоремы:

Теорема 13. *Для всякого алгоритма M найдётся константа c и алгоритм A, который по любому n и по любому числу $t > B(n + c)$ выдаёт список всех слов длины не более n, на которых алгоритм M останавливается.*

Используя традиционную для теории алгоритмов терминологию, можно сказать, что согласно этой теореме проблема остановки для слов длины не более n *сводится* к задаче отыскания числа, большего $B(n + c)$.

Если в качестве M взять оптимальный декомпрессор D, то верно и обратное утверждение: зная n и умея решать проблему остановки D для слов длины не более n, мы можем составить список всех слов сложности не более n и тем самым определить $B(n)$. Заметим, что для решения проблемы остановки для слов длины не более n достаточно указать (помимо n) число слов такой длины, на которых D останавливается.

Это рассуждение можно продолжить, доказав такое утверждение:

Теорема 14. *Пусть $BB(n)$ — максимальное время работы оптимального декомпрессора на словах длины не более n. Тогда*

$$BB(n) \leqslant B(n+c) \quad \text{и} \quad B(n) \leqslant BB(n+c)$$

для некоторого c и всех n.

◀ Число $BB(n)$ имеет сложность не более $n + O(1)$, поскольку может быть получено по n и самому долго обрабатываемому описанию x длины не более n. Эта информация может быть записана в одном слове длины $n+1$, а именно, слове $0\ldots01x$ (в начале стоят $n - l(x)$ нулей). Поэтому $BB(n) \leqslant B(n+c)$ для некоторого c и всех n.

С другой стороны, покажем, что для любого $t > BB(n)$ его сложность $k = KS(t)$ не меньше $n - O(1)$. В самом деле, возьмём описание u числа t, имеющее длину k. Зная u и n, можно алгоритмически получить слово сложности больше n. Для этого мы восстанавливаем t и ждём t шагов для каждого описания длины не более n. Так мы обнаружим все слова сложности не более n, а затем возьмём первое слово, не попавшее в их число. Следовательно, для задания u и n нужно не меньше $n - O(1)$ битов. А между тем их можно задать $k + O(\log(n-k))$ битами, если к самоограниченному описанию числа $n-k$ приписать u. Отсюда получаем, что $k + O(\log(n-k)) \geqslant n - O(1)$, так что $(n-k) - O(\log(n-k)) \leqslant O(1)$ и $n - k \leqslant \leqslant O(1)$. ▶

Эта теорема показывает, что функцию B можно примерно (с точностью до $O(1)$-добавки в аргументе) охарактеризовать как «время обработки самого трудного входа» ограниченной длины. Родственная конструкция появлялась в теории сложности вычислений под именем *busy beaver function* (в честь «трудолюбивых бобров»). Точнее, там для данного n рассматривалась машина Тьюринга с n состояниями и двухбуквенным алфавитом (единица и пробел), которая останавливается на пустом входе, напечатав максимально возможное число единиц (не обязательно подряд).

Вообще можно сказать, что знание любого из следующих объектов (а также значения параметра n) позволяет найти любой другой (возможно, для чуть меньшего значения n):

(а) список всех слов сложности не более n с указанием их сложностей;

(б) число таких слов;

(в) значение $B(n)$;

(г) значение $BB(n)$;

(д) список всех слов длины не более n, на которых определён оптимальный декомпрессор («проблема остановки» оптимального декомпрессора на словах длины не более n);

(е) число таких слов;

(ё) самый долго обрабатываемый оптимальным декомпрессором вход длины не более n;

(ж) таблица T_n, в которой указаны значения сложности $KS(x)$ для всех слов x длины n;

(з) первое в словарном порядке слово γ_n длины n, имеющее сложность не менее n (напомним, что такое существует, так как число слов x с $KS(x) < n$ меньше 2^n).

Более точно, имеет место следующая теорема:

Теорема 15. *Все перечисленные объекты имеют сложность $n + O(1)$ и эквивалентны друг другу в следующем точном смысле: пусть X_n и Y_n — объекты, указанные в каких-либо двух пунктах из числа* (а) – (з). *Тогда найдётся константа c и алгоритм, позволяющий по X_n и n найти Y_{n-c}.*

◀ Проще всего доказать эквивалентность объектов (г), (д), (е) и (ё). Зная любой из них и число n, можно выяснить поведение оптимального декомпрессора на всех описаниях длины не более n, то есть выписать все его останавливающиеся вычисления на входах длины не более n. В самом деле, зная список (д), мы просто применяем оптимальный декомпрессор ко всем входам из этого списка (зная заранее, что вычисления остановятся). Зная (г), мы применяем оптимальный декомпрессор ко всем входам длины не более n, но даём ему проработать не более $BB(n)$ шагов (будучи уверенными, что не остановившиеся за это время вычисления никогда не остановятся). Зная самый долго обрабатываемый вход длины не более n, мы применяем к нему оптимальный декомпрессор и находим $BB(n)$, а дальше действуем, как в предыдущем случае. Наконец, зная число слов длины не более n, на которых оптимальный декомпрессор останавливается, мы применяем оптимальный декомпрессор параллельно ко всем словам длины не более n, пока не наберём нужное количество результатов (оставшиеся вычисления не закончатся никогда).

Обратно, зная n и поведение оптимального декомпрессора на всех описаниях длины не более n, легко найти любой из объектов (г), (д), (е) и (ё), а также объекты (а) – (в). По транзитивности (почти очевидной) мы заключаем, что (г) – (ё) все эквивалентны.

Теперь докажем, что (а) – (в) эквивалентны друг другу, а также (г) – (ё). Зная список всех слов сложности не больше n, можно найти их число (переход от (а) к (б)). Перейти от (б) к (а) не так просто: зная число слов сложности не больше n (и зная само n), можно дождаться появления всех таких слов, и получить их список (в котором можно найти наибольшее число, получив (в)), но этот список будет без указания сложностей. Поскольку мы уже знаем, как от (г) перейти к (а), нам достаточно перейти от (в) к (г), то есть от $B(n)$ к $BB(n)$ (с изменением аргумента на $O(1)$). По существу этот переход уже обсуждался. Зная $B(n)$, мы знаем верхнюю оценку для $BB(n-c)$, и остаётся лишь подождать указанное время для всех входов

длины не более $n - c$, чтобы узнать точное значение $BB(n - c)$. Итак, мы доказали эквивалентность всех объектов (а) – (ё).

Зная (а) – (ё), можно найти (ж) для чуть меньшего значения n — уменьшенного на константу c, для которой $KS(x) \leqslant l(x) + c$ при всех x. Ясно также, как от (ж) можно перейти к (з). Осталось показать, как от (з) перейти к (г). Как, зная первое в алфавитном порядке слово γ_n сложности не менее n среди слов длины n, найти $BB(n - O(1))$ или хотя бы получить верхнюю оценку для $BB(n - O(1))$? Это делается следующим образом.

Зная γ_n (и тем самым зная n), будем искать для всех предшествующих ему слов длины n их описания длины меньше n; раз такие описания существуют, то рано или поздно мы их найдём (пусть даже и не кратчайшие). Найдя их, рассмотрим максимальное время работы оптимального декомпрессора на этих (найденных нами) описаниях; пусть это время будет T. Мы хотим доказать, что $T > BB(n - c)$ для некоторой константы c, не зависящей от n. Допустим, что для данного c это неравенство не выполнено, то есть $T \leqslant BB(n - c)$. Это значит, что одно из слов длины не более $n - c$ обрабатывается больше времени, чем нужно для обнаружения описаний длины не более n у всех предшественников γ_n. Пусть это слово u, а длина его $n - d$. Зная u и d, мы можем восстановить n, а затем найти γ_n (за время обработки u найдутся все предшественники γ_n, так что γ_n будет первым отсутствующим). Чтобы задать u и d, достаточно $(n - d) + 2\log d + O(1)$ битов, откуда $n - d + 2\log d \geqslant KS(\gamma_n) - O(1) \geqslant n - O(1)$, так что $d = O(1)$.

Тем самым завершено доказательство эквивалентности объектов (а) – (ж) (в указанном выше смысле). Докажем теперь, что сложность любого из них равна $n + O(1)$. Пусть X_n — любой из этих объектов. Как мы доказали, X_n можно получить по γ_{n+c} и n (число n равно $l(\gamma_{n+c}) - c$ и потому может быть опущено), следовательно, $KS(X_n) \leqslant KS(\gamma_{n+c}) + O(1) \leqslant n + O(1)$.

С другой стороны, пусть сложность X_n равна $n - d$. Слово γ_{n-c} по определению имеет сложность не меньше $n - c$ и может быть получено по кратчайшему описанию X_n длины $n - d$ и по d (заметим, что по этим данным можно восстановить n, сложив длину кратчайшего описания с d). Следовательно, $n - c \leqslant KS(\gamma_{n-c}) \leqslant (n - d) + 2\log d + O(1)$, то есть $d \leqslant 2\log d + c + O(1)$ и $d = O(1)$. ►

При доказательстве теоремы 15 мы предполагали, что фиксирован некоторый оптимальный декомпрессор (вообще говоря, все рассматриваемые в ней объекты зависят от его выбора). Это, однако, не играет роли, как показывает следующая задача:

$\boxed{8}$ Докажите, что утверждение теоремы 15 остаётся верным, если разрешить в разных пунктах (а) – (ж) использовать разные оптимальные декомпрессоры.

$\boxed{9}$ Покажите, что все объекты теоремы 15 вычислимо находятся по n, если разрешить использовать оракул $\mathbf{0}'$, то есть разрешить алгоритмам спрашивать у оракула, остановится ли данная программа (эта программа уже не обращается сама к оракулу). Поэтому сложность с оракулом $\mathbf{0}'$ для всех этих объектов есть $O(\log n)$.

Утверждение этой задачи можно обратить, и использовать сведения о сложных объектах для решения проблемы остановки. Мы установили, что существу-

ет константа c и алгоритм A, который по слову γ_n решает проблему остановки оптимального декомпрессора для всех слов длины не больше $n - c$. (Аналогичное рассуждение годится и для любого алгоритма, а не только для оптимального декомпрессора.) Это означает, что если у нас есть «оракул», который для каждого n указывает слово γ_n, то с его помощью можно решать проблему остановки. Вместо этого можно также использовать оракул для множества «сжимаемых» слов (тех слов x, для которых $KS(x) < l(x)$): умея проверять сжимаемость слова, мы можем найти γ_n перебором.

Как говорят, проблема остановки *сводится по Тьюрингу* к множеству сжимаемых слов. Отсюда очевидно следует, что она сводится по Тьюрингу к «надграфику» функции KS, то есть множеству $\{\langle x, k\rangle \mid KS(x) < k\}$. Как говорят, множество сжимаемых слов и надграфик KS являются *полными по Тьюрингу* перечислимыми множествами (это означает, что к любому из них сводится по Тьюрингу проблема остановки).

10 Оцените сверху (хоть как-нибудь) число вопросов, которые придётся задать оракулу для множества $\{\langle x, k\rangle \mid KS(x) < k\}$, чтобы решить проблему остановки оптимального декомпрессора для всех слов длины не более n.

Следующая задача показывает, что функция B растёт так быстро, что её нельзя существенно увеличить, применив какую-то вычислимую функцию (увеличение покрывается изменением аргумента на константу).

11 Докажите, что если f — вычислимая функция с натуральными аргументами и значениями, то найдётся такая константа c, что для всех n, для которых $f(B(n))$ определено, выполнено неравенство $B(n + c) \geqslant f(B(n))$. [Указание: число $f(B(n))$ имеет сложность не больше $n + O(1)$.]

12 Говорят, что множество U является *r-отделимым* [118], если всякое перечислимое множество V, не пересекающееся с U, можно отделить от U разрешимым множеством, то есть найдётся разрешимое множество R, содержащее V и не пересекающееся с U.

(а) Докажите, что множество пар $\{\langle x, k\rangle \mid KS(x) < k\}$ (которое можно назвать «надграфиком» функции KS) является r-отделимым множеством. Множество сжимаемых слов тоже является r-отделимым. [Указание: если этот надграфик не пересекается с некоторым перечислимым множеством V, то вторые компоненты пар из V ограничены, иначе мы получили бы неограниченную нижнюю оценку для KS. Значит, V целиком содержится в некоторой полосе, а пересечение надграфика с этой полосой конечно.]

(б) Докажите, что если множество U_1 m-сводится к множеству U_2 (существует всюду определённая вычислимая функция f, для которой $U_1 = f^{-1}(U_2)$) и U_2 является r-отделимым, то U_1 также r-отделимо. [Указание. Если перечислимое V не пересекается с U_1, то $f(V)$ является перечислимым множеством, не пересекающимся с U_2. Если R отделяет $f(V)$ от U_2, то $f^{-1}(R)$ будет разрешимым множеством, отделяющим V от U_1.]

(в) Выведите из этого, что существуют перечислимые множества, не обладающие свойством r-отделимости и потому не m-сводящиеся к надграфику функции KS. [Указание: существуют перечислимые неотделимые множества.]

$\boxed{\mathbf{13}}$ Покажите, следуя [59], что задачи «по любому натуральному n указать слово сложности не меньше n» и «по любому алгоритму без входа указать слово, отличающееся от его выхода (если таковой существует)» эквивалентны: имея оракул, выполняющий требуемое для одной из задач, можно алгоритмически решать другую. [Указание: видя алгоритм, можно оценить сверху сложность его возможного выхода, она не сильно превосходит сложность самого алгоритма. Напротив, указать слово большой сложности означает указать слово, отличное от выходов алгоритмов из списка; считая, что эти алгоритмы выдают кортежи, мы с помощью оракула построить кортеж, который в i-й позиции отличается от выхода i-го алгоритма списка.]

$\boxed{\mathbf{14}}$ (Продолжение) Покажите, что обе эти задачи эквивалентны (в том же смысле) задаче вычисления функции без неподвижной точки: по любому алгоритму построить другой алгоритм, вычисляющий не ту же функцию, что первый.

$\boxed{\mathbf{15}}$ (Продолжение) Покажите, что перечислимый оракул позволяет решать эти задачи в том и только том случае, когда с его помощью можно решать проблему остановки (теорема Арсланова, доказанная им без использования колмогоровской сложности). [Указание: пусть оракул для перечислимого множество A позволяет получать слова сколь угодно большой сложности. Посмотрим, сколько шагов нужно, чтобы в перечислении A появились все элементы, используемые при получении слова сложности более n. Любое T, большее этого числа, имеет сложность порядка n, так как T-приближение A можно использовать вместо A. С другой стороны, имея оракул для A, можно найти T для данного n.]

Мы привели лишь несколько самых простых примеров, связывающих колмогоровскую сложность с теорией вычислимости (или, как раньше говорили, «теорией рекурсии»). В последние десятилетия эта область активно развивалась и продолжает развиваться, но мы ограничимся приведёнными примерами. Множество интересных результатов приведено в недавних монографиях Ниса [127], а также Доуни и Хиршфельдта [37].

Теорема 15 указывает среди всех объектов сложности n некоторый выделенный (с точностью до описанной эквивалентности). Это отчасти парадоксально: хотелось бы думать, что все «случайные» слова длины n (слова длины n и сложности примерно n) более или менее одинаковы (если бы какое-то из них выделялось чем-то особенным, то эту особенность можно было бы использовать, чтобы задать его короче, и тем самым слово было бы не случайным). Тем не менее мы указали некоторое специальное случайное слово γ_n, как же так? Разрешение этого парадокса состоит в том, что индивидуальные особенности этого слова могут быть использованы для его простого задания, но лишь с оракулом $\mathbf{0}'$.

Мы вернёмся ещё к этому вопросу, когда будем говорить о числе Ω в разделе 5.7, а также в разделе 14.3, говоря о двухчастных описаниях.

Отметим ещё, что хотя все объекты из теоремы 15 эквивалентны в указанном в ней смысле (просто получаются друг из друга), но по длине они отличаются радикально: от длины n в пунктах (б), (е), (ё) и (з) до длины $\log B(n) \approx B(n - O(1))$ в пункте (в).

2. Сложность пары и условная сложность

2.1. Сложность пары

Мы уже говорили, что вычислимое кодирование позволяет говорить не только о сложности слов, но и о сложности других конструктивных объектов. Сейчас нас интересуют (упорядоченные) пары слов. Пару слов x, y можно кодировать, например, словом $[x, y] = \bar{x}01y$, где $\bar{x}$ означает слово x с удвоенными битами. Можно использовать и любое другое кодирование, важно только, чтобы оно было вычислимым и чтобы $[x, y] \neq [x', y']$, если $x \neq x'$ или $y \neq y'$. От одного такого кодирования можно алгоритмически переходить к другому, поэтому теорема 3 (с. 14) о невозрастании сложности при алгоритмическом преобразовании показывает, что смена кодирования изменяет сложность не более чем на константу.

Фиксируем какое-либо кодирование пар ($[x, y]$ обозначает слово, кодирующее пару слов x, y) и назовём сложностью пары слов x, y число $KS([x, y])$. Обозначение: $KS(x, y)$. Вот несколько очевидных свойств:

- $KS(x, x) = KS(x) + O(1)$;
- $KS(x, y) = KS(y, x) + O(1)$;
- $KS(x) \leqslant KS(x, y) + O(1)$; $KS(y) \leqslant KS(x, y) + O(1)$.

Следующая теорема оценивает сложность пары, если известны сложности её компонент:

Теорема 16.

$$KS(x, y) \leqslant KS(x) + 2 \log KS(x) + KS(y) + O(1);$$
$$KS(x, y) \leqslant KS(x) + \log KS(x) + 2 \log\log KS(x) + KS(y) + O(1);$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

(Последовательность утверждений теоремы можно продолжать неограниченно. Кроме того, можно поменять местами x и y.)

◄ Это рассуждение (для первого из неравенств) по существу проводилось во введении (теорема 4 на с. 15); правда, там мы говорили о сложности конкатенации слов xy, а не о сложности пары. Повторим его для пар.

Назовём вычислимое отображение $x \mapsto \hat{x}$ множества двоичных слов в себя *беспрефиксным кодированием*, если ни для каких двух различных слов x и y слово $\hat{x}$ не является началом слова $\hat{y}$. (Отсюда, в частности, следует, что $\hat{x} \neq \hat{y}$ при $x \neq y$.) Смысл этого определения таков: любое слово вида $\hat{x}z$ можно однозначно разрезать на части и найти слова x и z.

Пример беспрефиксного кодирования: $x \mapsto \bar{x}01$, где $\bar{x}$ означает строку x с удвоенными битами. Здесь признаком окончания слова являются биты 01. Это кодирование не очень экономно (увеличивает длину вдвое). Более экономно такое кодирование:

$$x \mapsto \hat{x} = \overline{\text{bin}(l(x))}01x$$

(здесь $\text{bin}(l(x))$ — двоичная запись длины слова x). Для него

$$l(\hat{x}) = l(x) + 2\log l(x) + O(1).$$

Этот приём можно «итерировать»: начав с произвольного беспрефиксного кодирования $x \mapsto \hat{x}$, можно построить новое (также беспрефиксное) кодирование

$$x \mapsto \widehat{\text{bin}(l(x))}x.$$

В самом деле, если слово $\widehat{\text{bin}(l(x))}x$ является началом слова $\widehat{\text{bin}(l(y))}y$, то одно из слов $\widehat{\text{bin}(l(x))}$ и $\widehat{\text{bin}(l(y))}$ является началом другого, поэтому $\text{bin}(l(x)) = \text{bin}(l(y))$. Отсюда мы заключаем, что x есть начало y, а потом — что $x = y$. (Другими словами, сначала мы однозначно декодируем длину слова, пользуясь тем, что она записана в беспрефиксном коде, а потом уже однозначно определяем само слово.)

Такая итерация даёт беспрефиксное кодирование, при котором

$$l(\hat{x}) = l(x) + \log l(x) + 2\log\log l(x) + O(1),$$

затем

$$l(\hat{x}) = l(x) + \log l(x) + \log\log l(x) + 2\log\log\log l(x) + O(1),$$

и так далее.

Вернёмся к доказательству теоремы. Пусть D — оптимальный способ описания, используемый при определении сложности. Рассмотрим способ описания D', задаваемый так:

$$D'(\hat{p}q) = [D(p), D(q)],$$

где $\hat{p}$ — беспрефиксный код слова p, а квадратные скобки означают кодирование пар, использованное при определении сложности пары. Беспрефиксность кодирования $p \mapsto \hat{p}$ гарантирует корректность этого определения ($\hat{p}$ однозначно вычленяется из $\hat{p}q$).

Если p и q — кратчайшие описания слов x и y, то слово $\hat{p}q$ будет описанием слова $[x, y]$, и его длина как раз даёт оценку, указанную в теореме. (Чем более экономно беспрефиксное кодирование, тем лучше получается оценка; описанный выше итеративный метод построения беспрефиксных кодов даёт необходимые оценки.) ▶

Из теоремы 16 следует, что

$$KS(x, y) \leqslant KS(x) + KS(y) + O(\log n)$$

для слов x и y длины не более n. Как говорят, сложность пары не превосходит суммы сложностей её членов с точностью до логарифма (длин).

16 Дайте естественное определение сложности тройки слов. Покажите, что $KS(x,y,z) \leqslant KS(x) + KS(y) + KS(z) + O(\log n)$, если слова x, y, z имеют длину не больше n.

Возникает естественный вопрос: нельзя ли усилить оценку теоремы 16 и доказать, что $KS(x,y) \leqslant KS(x) + KS(y) + O(1)$?

Следующее простое рассуждение показывает, что это невозможно. В самом деле, из этой оценки вытекало бы, что $KS(x,y) \leqslant l(x) + l(y) + O(1)$. Рассмотрим какое-то N и всевозможные $n = 0, 1, 2, \ldots, N$. Для каждого такого n имеется 2^n слов x длины n, а также 2^{N-n} слов y длины $N - n$. Комбинируя такие x и y, мы для данного n получаем 2^N пар $\langle x, y\rangle$, а всего (для всех n от 0 до N) получаем $(N+1)2^N$ пар. Если бы сложность всех этих пар $\langle x, y\rangle$ не превосходила $l(x) + l(y) + O(1) = N + O(1)$, то получилось бы $(N+1)2^N$ различных слов $[x, y]$, имеющих сложность не более $N + O(1)$, а таких слов, как мы знаем (теорема 7, с. 26), лишь $O(2^N)$.

17 Докажите, что не найдётся такого c, что

$$KS(x,y) \leqslant KS(x) + \log KS(x) + KS(y) + c$$

при всех x и y. [Указание. Замените в правой части неравенства сложности на длины и подсчитайте количество пар.]

18 (а) Покажите, что если $x \mapsto \hat{x}$ — беспрефиксное кодирование, то

$$\sum_{x \in \Xi} 2^{-l(\hat{x})} \leqslant 1$$

(здесь Ξ — множество всех двоичных слов).

(б) Покажите, что если беспрефиксное кодирование увеличивает длину слова не более чем на $f(n)$ (где n — исходная длина), то есть $l(\hat{x}) \leqslant l(x) + f(l(x))$, то ряд $\sum_n 2^{-f(n)}$ сходится.

Эта задача объясняет, почему понадобился коэффициент 2 при логарифме в доказательстве теоремы 16 (с. 40): ряды

$$\sum \frac{1}{n^2}, \quad \sum \frac{1}{n(\log n)^2}, \quad \sum \frac{1}{n \log n (\log\log n)^2}, \quad \ldots$$

сходятся, в то время как ряды

$$\sum \frac{1}{n}, \quad \sum \frac{1}{n \log n}, \quad \sum \frac{1}{n \log n \log\log n}, \quad \ldots$$

расходятся.

Следующая задача даёт общий ответ на вопрос, какие функции могут быть использованы для оценок теоремы 16.

19 Пусть $f\colon \mathbb{N} \to \mathbb{N}$ — неубывающая вычислимая функция. Тогда следующие три свойства равносильны:

(а) $KS(x,y) \leqslant KS(x) + KS(y) + f(KS(x)) + O(1)$;

(б) $KS(x,y) \leqslant l(x) + l(y) + f(l(x)) + O(1)$;

(в) $\sum_n 2^{-f(n)} < \infty$.

[Указание. Из (а) очевидно следует (б); обратная импликация получается переходом к кратчайшим описаниям. Чтобы вывести (а) из (в), можно подсчитать число пар с $l(x) + f(l(x)) + l(y) < n$; можно воспользоваться также свойствами префиксной сложности из главы 4, см. задачу 107. Наконец, чтобы вывести (в) из (б), заметим, что правая часть в (б) будет меньше $n + O(1)$, если $l(x) = k$, а $l(y) = n - k - f(k)$ (при $k + f(k) \leqslant n$). Поэтому число таких пар не меньше $\sum 2^k 2^{n-k-f(k)} = 2^n \sum_k 2^{-f(k)}$, где сумма берётся по всем k, для которых $k + f(k) \leqslant n$.]

20 Докажите, что все неравенства в теореме 16 перестанут быть верными, если заменить в них коэффициент 2 на 1 (но можно заменить его на $1 + \varepsilon$ при любом $\varepsilon > 0$). [Указание. См. предыдущую задачу.]

21 Докажите, что

$$KS(x,y) \leqslant KS(x) + \log KS(x) + KS(y) + \log KS(y) + O(1).$$

22 (Продолжение) Докажите, что верно даже более сильное неравенство:

$$KS(x,y) \leqslant KS(x) + KS(y) + \log(KS(x) + KS(y)) + O(1)$$

(где сумму под логарифмом можно заменить максимумом, так как они отличаются не более чем вдвое).

23 Докажите, что $KS(x, KS(x)) = KS(x) + O(1)$. [Указание. Очевидно, что $KS(x, KS(x)) \geqslant KS(x) + O(1)$. С другой стороны, кратчайшее описание p слова x задаёт одновременно и x, и $KS(x)$, так что $KS(x, KS(x)) \leqslant l(p) + O(1) = KS(x) + O(1)$.]

24 Покажите, что если $KS(x) \leqslant n$ и $KS(y) \leqslant n$, то $KS(x,y) \leqslant 2n$.

2.2. Условная сложность

Посылая файл по электронной почте, можно сэкономить, если послать не сам файл, а его сжатый вариант (кратчайшее описание). Экономия может быть ещё больше, если получатель уже имеет старую версию файла — в таком случае достаточно описать внесённые изменения. Эти соображения приводят к следующему определению *условной сложности слова x при известном слове y*.

Назовём *способом условного описания* произвольную вычислимую функцию D двух аргументов (аргументы и значения функции D являются двоичными словами). Первый аргумент мы будем называть *описанием*, второй — *условием*. Если $D(y,z) = x$, мы говорим, что слово y является *описанием* слова x *при известном* z (ещё говорят «при условии z», «относительно z»). Сложность $KS_D(x \mid z)$ определяется как длина кратчайшего описания слова x при известном слове z:

$$KS_D(x \mid z) = \min\{l(y) \mid D(y,z) = x\}.$$

Говорят, что способ (условного) описания D_1 *не хуже* способа D_2, если найдётся такая константа c, что

$$KS_{D_1}(x\,|\,z) \leqslant KS_{D_2}(x\,|\,z) + c$$

для любых слов x и z. Способ (условного) описания D называется *оптимальным*, если он не хуже любого другого способа (условного) описания.

Теорема 17. *Существует оптимальный способ условного описания.*

◀ Этот вариант теоремы Колмогорова – Соломонова доказывается точно так же, как и безусловный (см. теорему 1, с. 11).

Именно, фиксируем некоторый способ программирования для функций двух аргументов, при котором программы записываются в виде двоичных слов, и положим

$$D(\hat{p}y, z) = p(y, z),$$

где $p(y,z)$ обозначает результат применения программы p ко входам y и z, а $\hat{p}$ есть беспрефиксный код слова p. Теперь легко проверить, что если D' — произвольный способ условного описания, а p — соответствующая ему программа, то

$$KS_D(x\,|\,z) \leqslant KS_{D'}(x\,|\,z) + l(\hat{p}). \quad ▶$$

Как и прежде, мы фиксируем некоторый оптимальный способ D условного описания и опускаем индекс D в KS_D. Соответствующую функцию мы называем *условной колмогоровской сложностью*; как и безусловная, она определена с точностью до ограниченного слагаемого.

Вот несколько простых свойств условной колмогоровской сложности:

Теорема 18.

$$\begin{aligned} KS(x\,|\,y) &\leqslant KS(x) + O(1);\\ KS(x\,|\,x) &= O(1);\\ KS(f(x,y)\,|\,y) &\leqslant KS(x\,|\,y) + O(1);\\ KS(x\,|\,y) &\leqslant KS(x\,|\,g(y)) + O(1). \end{aligned}$$

Здесь g и f — произвольные вычислимые функции (одного и двух аргументов); имеется в виду, что указанные в теореме неравенства выполнены, если $f(x,y)$ (соответственно $g(y)$) определено.

◀ Безусловный способ описания можно рассматривать и как условный (не зависящий от второго аргумента), отсюда получаем первое неравенство. Второе утверждение получается, если рассмотреть способ описания $D(p,z) = z$. Третье неравенство доказывается так: пусть D — оптимальный способ условного описания; рассмотрим другой способ

$$D'(p,y) = f(D(p,y),y)$$

и сравним его с оптимальным. Доказательство последнего неравенства аналогично, только нужно определить способ описания так:

$$D'(p,y) = D(p,g(y)). \quad ▶$$

$\boxed{25}$ Покажите, что условная сложность «непрерывна по второму аргументу»: $KS(x\,|\,y0) = KS(x\,|\,y) + O(1)$; $KS(x\,|\,y1) = KS(x\,|\,y) + O(1)$.

Используя эту «непрерывность», покажите, что для любого слова x и для любого неотрицательного $l \leqslant KS(x)$ найдётся условие y, при котором $KS(x\,|\,y) = l + O(1)$.

$\boxed{26}$ Покажите, что при фиксированном y функция $x \mapsto KS(x\,|\,y)$ отличается от KS не более чем на константу, зависящую от y (и не превосходящую $2\,KS(y) + O(1)$).

$\boxed{27}$ Докажите, что $KS([x, z]\,|\,[y, z]) \leqslant KS(x\,|\,y) + O(1)$ для любых x, y, z (квадратные скобки означают вычислимое кодирование пар).

$\boxed{28}$ Пусть фиксирован некоторый разумный язык программирования. (Формально говоря, надо, чтобы соответствующая ему нумерация была главной, то есть чтобы была возможна эффективная трансляция программ с других языков [175].) Покажите, что условная сложность $KS(x\,|\,y)$ равна (с точностью до $O(1)$) минимальной сложности программы, которая даёт x на входе y. [Указание. Если D — оптимальный способ условного описания, то сложность программы, которая получается из D фиксацией первого аргумента p, не превосходит $l(p) + O(1)$. С другой стороны, если программа p переводит y в x, то $KS(x\,|\,y) = KS(p(y)\,|\,y) \leqslant KS(p) + O(1)$.]

К этой интерпретации условной сложности (как минимальной сложности объекта с определённым свойством) мы вернёмся в главе 13.

Если в определении предыдущей задаче ограничиться всегда останавливающимися программами (программами для всюду определённых функций), то получится другой вид условной сложности, который можно назвать *тотальной* сложностью (по-английски всюду определённые функции называются *total*).

$\boxed{29}$ Покажите, что определение тотальной условной сложности $KT(x\,|\,y)$ как минимальной сложности программы, переводящей y в x и останавливающейся на любом входе, корректно (не зависит с точностью до $O(1)$ от выбора разумного способа программирования). Докажите, что $KS(x\,|\,y) \leqslant KT(x\,|\,y) \leqslant KS(x)$ с точностью до $O(1)$.

$\boxed{30}$ Докажите, что тотальная сложность может быть сильно больше условной: для любого n найдутся два слова x и y длины n, для которых $KS(x\,|\,y) = O(1)$, но $KT(x\,|\,y) \geqslant n$. [Указание. Будем перечислять все программы сложности меньше n, определённые на всех словах длины n, храня два слова x и y длины n и поддерживая такое свойство: ни одна из перечисленных программ не переводит y в x. Когда обнаруживается новая программа, нарушающая это условие, мы выбираем новое значение y и подбираем подходящее значение x. Этот процесс вычислим по n (длине y) и задаёт частичную функцию на словах длины n, так что $KS(x\,|\,y) = O(1)$ для каждой выбранной пары значений.]

$\boxed{31}$ Покажите, что если $KT(x\,|\,y) \leqslant n$ и $KT(y\,|\,x) \leqslant n$, то существует программа вычислимой перестановки множества двоичных слов, переводящая x в y и имеющая сложность не более $2n + O(1)$. [Указание. Можно закодировать программы всюду определённых функций f и g, переводящих x и y и обратно, одним словом

длины $2n$. Без ограничения общности можно считать также, что x и y начинаются (скажем) на 0. Тогда рассмотрим отношение $R(u, v)$ между словами, начинающимися на 0, считая $R(u, v)$ истинным, если $f(u) = v$ и $g(v) = u$. Это разрешимое взаимно однозначное соответствие между разрешимыми множествами слов, которое легко продолжить до перестановки множества всех слов, так как дополнения этих разрешимых множеств бесконечны.]

$\boxed{32}$ Покажите, что оценку предыдущей задачи нельзя существенно улучшить: для любого k найдутся два слова x и y длины $n = 2k + O(1)$, для которых $KS(x), KS(y) \leqslant k + O(1)$ (и тем более $KT(x\,|\,y), KT(y\,|\,x) \leqslant k + O(1)$), но любая перестановка множества слов длины n, отображающая x в y, имеет сложность не менее $2k$. [Указание. Для начала выберем (произвольно) 2^k слов y длины n и какое-нибудь одно слово x. Будем перечислять все перестановки сложности менее $2k$. Когда и если все пары выбранных слов будут связаны какой-либо из перечисленных перестановок, выбираем новое слово x, которое связано менее чем с половиной из уже выбранных слов y имеющимися перестановками, и добавим его. После этого понадобится $\Omega(2^k)$ новых перестановок, чтобы связать его со всеми выбранными y. Поэтому будет выбрано не более $2^{2k}/\Omega(2^k) = O(2^k)$ слов x, и сложность выбранных слов x и y будет не больше $k + O(1)$. Выбор слова x с указанным свойством возможен, так как каждое из выбранных слов y связано с небольшой долей всех возможных x, и можно переставить порядок усреднения при подсчёте долей. Заметим ещё, что это рассуждение позволяет взять одно из слов в заданном множестве из 2^k элементов.]

Многие свойства безусловной сложности легко переносятся и на условную. Вот некоторые из них (доказательства повторяют соответствующие рассуждения для безусловной сложности):

- функция $KS(x\,|\,y)$ перечислима сверху (это означает, что множество троек $\langle x, y, n\rangle$, для которых $KS(x\,|\,y) < n$, перечислимо);
- при данных y и n множество тех слов x, для которых $KS(x\,|\,y) < n$, содержит менее 2^n элементов, поэтому
- для всякого y и для всякого n найдётся слово x длины n, сложность которого при известном y не меньше n.

$\boxed{33}$ Докажите, что для любых слов y и z и для любого n найдётся слово x длины n, у которого $KS(x\,|\,y) \geqslant n - 1$ и $KS(x\,|\,z) \geqslant n - 1$. [Указание: плохие слова каждого типа образуют менее половины.]

Теорема 19. *Если $\langle x, y\rangle \mapsto K(x\,|\,y)$ — перечислимая сверху функция, причём для любых y и n множество*

$$\{x \mid K(x\,|\,y) < n\}$$

содержит менее 2^n элементов, то $KS(x\,|\,y) \leqslant K(x\,|\,y) + c$ при некотором c и при любых x и y.

Эта теорема доказывается аналогично теореме 8.

В теореме о сложности пары (теорема 16, с. 40) также можно заменить сложность на условную:

Теорема 20.

$$KS(x,y) \leqslant KS(x) + 2\log KS(x) + KS(y\,|\,x) + O(1).$$

◀ Пусть D_1 — оптимальный способ безусловного описания, а D_2 — оптимальный способ условного описания. Построим способ описания D', положив

$$D'(\hat{p}q) = [D_1(p), D_2(q, D_1(p))].$$

Здесь $\hat{p}$ — беспрефиксный код слова p, а квадратные скобки обозначают вычислимое кодирование пар, используемое при определении сложности пары. Если p — кратчайшее описание слова x, а q — кратчайшее описание слова y при известном x, то слово $\hat{p}q$ будет описанием слова $[x,y]$ относительно D', откуда

$$KS(x,y) \leqslant KS_{D'}(x,y) + O(1) \leqslant l(\hat{p}) + l(q) + O(1).$$

Осталось заметить, что можно выбрать беспрефиксное кодирование так, чтобы $l(\hat{p})$ не превосходило $l(p) + 2\log l(p) + O(1)$ (см. доказательство теоремы 16 о сложности пары, с. 40). ▶

Как и раньше, можно заменить $2\log KS(x)$ на $\log KS(x) + 2\log\log KS(x)$ и так далее. Можно также добавочный член перенести на условную сложность, написав

$$KS(x,y) \leqslant KS(x) + KS(y\,|\,x) + 2\log KS(y\,|\,x) + O(1).$$

В доказательстве при этом следует заменить $D'(\hat{p}q)$ на $D'(\hat{q}p)$.

$\boxed{34}$ Докажите «неравенство треугольника»:

$$KS(x\,|\,z) \leqslant KS(x\,|\,y) + 2\log KS(x\,|\,y) + KS(y\,|\,z) + O(1)$$

для любых трёх слов x, y, z.

Если не вдаваться в тонкости различных вариантов оценки добавочного члена с логарифмом, результат теоремы 20 можно сформулировать так: для слов x и y длины не более n имеет место неравенство

$$KS(x,y) \leqslant KS(x) + KS(y\,|\,x) + O(\log n).$$

Оказывается, что это неравенство на самом деле представляет собой равенство (с той же логарифмической точностью).

Теорема 21 (Колмогорова – Левина).

$$KS(x,y) = KS(x) + KS(y\,|\,x) + O(\log n)$$

для слов x, y длины не больше n.

◀ В одну сторону неравенство уже доказано. Осталось доказать, что $KS(x,y) \geqslant \geqslant KS(x) + KS(y\,|\,x) + O(\log n)$, если x и y — слова длины не более n.

Пусть x, y — произвольные слова длины не более n. Обозначим сложность $KS(x,y)$ пары $\langle x, y\rangle$ через a. Пусть A — множество всех пар слов, у которых сложность не больше a. Число элементов в множестве A не больше $O(2^a)$ (точнее, меньше 2^{a+1}), и пара $\langle x, y\rangle$ — один из этих элементов.

Для каждого слова t рассмотрим сечение A_t множества A:

$$A_t = \{u \mid \langle t, u\rangle \in A\}$$

(рис. 2.1). Сумма мощностей всех множеств A_t при всех t равна мощности множества A, то есть не превосходит $O(2^a)$. Поэтому бо́льших сечений у множества A немного, что мы сейчас и используем.

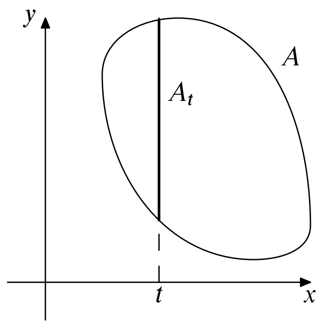


Рис. 2.1. Сечение A_t множества A простых пар.

Пусть $m = \lfloor \log_2 |A_x| \rfloor$, где x — первая компонента исходной пары. Другими словами, пусть число элементов в A_x заключено между 2^m и 2^{m+1}. Докажем, что $KS(y\,|\,x)$ не сильно превосходит m, а $KS(x)$ не сильно превосходит $a - m$. Начнём с первого.

Зная a, можно перечислять множество A. Если мы к тому же знаем x, то можно оставлять от A только пары, у которых первая координата равна x, и получить перечисление множества A_x. Чтобы задать y, помимо a и условия x, достаточно указать порядковый номер элемента y в этом перечислении. Для этого достаточно $m + O(1)$ битов, и вместе с a получается $m + O(\log n)$ битов: заметим, что $a = KS(x,y)$ не превосходит $O(n)$, если x и y — слова длины не более n, а потому для задания a достаточно $O(\log n)$ битов. Итак,

$$KS(y\,|\,x) \leqslant m + O(\log n).$$

Перейдём ко второй оценке. Множество B всех t, для которых $|A_t| \geqslant 2^m$, содержит не более $2^{a+1}/2^m$ элементов (иначе сумма $|A| = \sum |A_t|$ была бы больше 2^{a+1}). Множество B можно перечислять, зная a и m. (В самом деле, надо перечислять пары в множестве A; как только найдётся 2^m пар с одинаковой первой координатой, эта координата помещается в перечисление множества B.) Тем самым

исходное слово x (как и любой другой элемент множества B) можно задать, указав $(a - m) + O(\log n)$ битов ($a - m$ битов уходят на порядковый номер слова x в B, а $O(\log n)$ битов позволяют дополнительно указать a и m). Отсюда

$$KS(x) \leqslant (a - m) + O(\log n),$$

и остаётся сложить два полученных неравенства. ▶

Доказанную только что теорему можно рассматривать как перевод на язык колмогоровской сложности такого комбинаторного факта. Пусть дано множество A пар слов. Его мощность не превосходит произведения мощности проекции на первую координату и мощности наибольшего сечения A_t (первая координата равна t, вторая произвольна). Это соответствует неравенству $KS(x, y) \leqslant KS(x) + KS(y \,|\, x) + O(\log n)$. Обратное неравенство интерпретируется сложнее. Пусть дано множество A пар слов, а также числа p и q, для которых $|A| \leqslant pq$. Тогда можно разбить A на две части P и Q с такими свойствами: проекция P на первую координату содержит не более p элементов, а все сечения Q_x множества Q (первая координата равна x, вторая произвольна) содержат не более q элементов. (В самом деле, если отнести к P те сечения, которые содержат больше q элементов, то их число не превосходит p.) Подробнее об этом см. в главе 10.

Заметим, что на самом деле для доказательства важны не длины слов x и y, а их сложности. По существу мы доказали такой факт:

Теорема 22 (Колмогорова – Левина, вариант со сложностью).

$$KS(x, y) = KS(x) + KS(y \,|\, x) + O(\log KS(x, y))$$

для любых слов x, y.

$\boxed{35}$ Оцените более точно константы в проведённом доказательстве и покажите, что

$$KS(x) + KS(y \,|\, x) \leqslant KS(x, y) + 3 \log KS(x, y) + O(\log\log KS(x, y)).$$

$\boxed{36}$ Докажите, что если $KS(x, y \,|\, k, l) < k + l$, то $KS(x \,|\, k, l) < k + O(1)$ или $KS(y \,|\, x, k, l) < l + O(1)$. [Указание: это в точности соответствует рассуждению в доказательстве теоремы 22.]

$\boxed{37}$ Покажите, что в теореме Колмогорова – Левина члены порядка $O(\log n)$ неизбежны (причём в обе стороны): при любом n найдутся слова x и y длины не более n, для которых

$$KS(x, y) \geqslant KS(x) + KS(y \,|\, x) + \log n - O(1),$$

а также слова x и y длины не более n, для которых

$$KS(x, y) \leqslant KS(x) + KS(y \,|\, x) - \log n + O(1).$$

[Указание. В первом случае можно воспользоваться замечанием после теоремы 16 (с. 42). Во втором случае заметим, что сложность пары $KS(x, l(x)) = KS(x) + O(1)$

при любом x, а $KS(x\,|\,l(x))$ может быть равно $l(x)+O(1)$ и $KS(x)+O(1)$. Остаётся выбрать случайную длину между $n/2$ и n, а потом случайное слово этой длины.]

$\boxed{38}$ Покажите, что изменение одного бита в слове длины n меняет его сложность не более чем на $\log n + O(\log\log n)$. Докажите то же самое для условной сложности $KS(x\,|\,n)$.

Как мы видели в задаче 7 (с. 30), для любого слова x длины n найдётся слово x' той же длины, отличающееся от x только в одной позиции, для которого $KS(x') < n - \log n + O(1)$ (и потому $KS(x'\,|\,n) < n - \log n + O(1)$). В частности, если x несжимаемо при известном n, то при достаточно больших n сложность $KS(x'\,|\,n)$ строго меньше сложности $KS(x\,|\,n)$. И наоборот, если сложность $KS(x\,|\,n)$ достаточно маленькая, точнее $KS(x\,|\,n) \leqslant \alpha n$ для некоторой положительной константы α, то её можно увеличить путём изменения некоторого бита x: для всякого достаточно большого n для всех x длины n с $KS(x\,|\,n) \leqslant \alpha n$ существует x' длины n, отличающееся от x только в одной позиции, для которого $KS(x'\,|\,n) > KS(x\,|\,n)$. Но доказательство этого факта требует более сложного комбинаторного рассуждения (см. [15]), чем уменьшение сложности.

$\boxed{39}$ Фиксируем некоторый (безусловный) способ описания D. Докажите, что для некоторой константы c и для всех n и k выполнено такое свойство: если какое-то слово x имеет 2^k описаний длины не более n, то $KS(x\,|\,k) \leqslant n - k + c$. [Указание. Пусть k фиксировано. Для каждого n рассмотрим слова x, имеющие не менее 2^k описаний длины не более n. Число таких слов (при данном k) не превосходит 2^{n-k}, и можно воспользоваться теоремой 19, с. 46.]

С помощью этой задачи можно доказать такое утверждение о безусловной сложности:

$\boxed{40}$ Пусть D — фиксированный (безусловный) способ описания. Тогда найдётся такое число c, что для любого слова x число кратчайших D-описаний слова x не превосходит c. [Указание. В условиях предыдущей задачи $KS(x) \leqslant n - k + 2\log k + O(1)$, поэтому при $KS(x) = n$ значение k ограничено.]

$\boxed{41}$ Докажите, что найдётся константа c с таким свойством: если для данных x и n вероятность события $KS(x\,|\,y) \leqslant k$ (для случайно взятого слова y длины n; все такие слова считаем равновероятными) не меньше 2^{-l}, то $KS(x\,|\,n,l) \leqslant k + l + c$. [Указание. Соединим каждое слово y длины n со всеми словами x, сложность которых относительно y не превосходит k, получим двудольный граф с $O(2^{n+k})$ рёбрами, и в нём число вершин x, из которых выходит не менее 2^{n-l} рёбер, есть $O(2^{k+l})$. Обратите внимание, что в $KS(x\,|\,n,l)$ в условие не входит k — это не опечатка!]

Эта задача позволяет ответить на такой вопрос: чему в среднем равна сложность $KS(x\,|\,y)$ для данного x и случайно выбранного слова y данной длины n? Ясно, что $KS(x\,|\,y) \leqslant KS(x\,|\,n) + O(1)$ (поскольку $n = l(y)$ восстанавливается по y). Оказывается, что для большинства слов y данной длины эта оценка точная:

$\boxed{42}$ Докажите, что найдётся такая константа c, что для любого слова x и для любых натуральных чисел n и d доля тех слов y длины n, у которых $KS(x\,|\,y) <$

$< KS(x\,|\,n) - d$ (среди всех слов длины n), не превосходит $cd^2/2^d$. Выведите отсюда, что среднее арифметическое $KS(x\,|\,y)$ по всем словам y данной длины n равно $KS(x\,|\,n) + O(1)$ (константа в $O(1)$ не зависит от x и n).

43 Докажите, что если $KS(x\,|\,k) \leqslant k$, то $KS(x) \leqslant k + O(1)$. [Указание. См. теорему 7. Можно заметить также, что если описание x при известном k имеет длину k, то можно считать k известным, а если оно короче, то разницу можно задать дополнительно, так как есть запас.]

Близкое (но не тождественное!) утверждение:

44 Докажите, что $KS(x) = KS(x\,|\,KS(x)) + O(1)$. [Указание. Пусть x имеет короткое описание q при известном $KS(x)$. Тогда для восстановления x достаточно задать q и разницу длин $KS(x) - l(q)$, и получается более короткое описание слова x, чем это возможно.]

45 Докажите, что при любом n среди слов длины n найдётся слово x, для которого $KS(KS(x)\,|\,x) = \log n - O(1)$. (Заметим, что это максимальное значение, так как $KS(x) \leqslant n$ для слов x длины n.)

Это утверждение (в несколько другом варианте) доказал П. Гач [43]. Наглядное доказательство в игровых терминах недавно предложили Е. Калинина и Б. Бауэнс, идея его такова. Рассмотрим прямоугольную доску ширины 2^n и высоты n, на которые два игрока (Белые и Чёрные) по очереди ставят пешки своего цвета. В отличие от шахмат, в одной и той же клетке могут помещаться и белая, и чёрная пешки (хватит места для одной пешки каждого цвета). На каждом ходу можно поставить несколько пешек в разные клетки (а можно и ни одной не ставить); поставленные пешки сдвигать и убирать нельзя. Ещё Чёрным разрешается закрашивать клетки. Ограничения:

(а) каждый из игроков может поставить не более 2^i пешек на горизонтали i (горизонтали нумеруются снизу вверх от 0 до $n-1$);

(б) Чёрные могут в каждой вертикали закрасить не более половины клеток.

Белая пешка считается убитой, если её клетка закрашена или если ниже в той же вертикали есть чёрная пешка. Игра не имеет официального конца (хотя по существу конечна), Белые выигрывают, если в пределе у них есть неубитая пешка. В этой игре Белые могут выиграть так: поставив пешку в верхней горизонтали, они ждут, пока Чёрные её не убьют. Если это произошло из-за того, что Чёрные поставили пешку ниже, Белые переходят к следующей вертикали (скажем, начав слева и идя направо). Если же Чёрные закрасили клетку, то Белые ставят пешку на клетку ниже и снова ждут, пока она не будет убита, и так далее. (Без ограничения общности можно считать, что ответные ходы Чёрных убивают пешку Белых; поскольку выигрыш определяется предельной позицией, другие ходы можно мысленно отложить.) Поскольку закрасить можно лишь половину клеток вертикали, Чёрным придётся рано или поздно поставить пешку. Во все вертикали они поставить пешки не могут, так как сумма 2^i по всем горизонталям на 1 меньше числа вертикалей. Ограничение на число пешек в горизонтали Белые не нарушат, так как во всех уже отыгранных вертикалях под их пешкой в горизонтали i есть чёрная

пешка в горизонтали $j < i$, а сумма 2^j по всем $j < i$ меньше 2^i, и есть место ещё для одной пешки Белых.

Теперь пусть Чёрные играют так: клетка (x, i) закрашивается, если обнаружилось, что $KS(i \mid x) < \log n - 1$; чёрная пешка ставится в клетку (x, i), если обнаружилось, что у x есть условное описание (при известном n) длины i. Легко понять, что при этом Чёрные не нарушат ограничений. Неубитая белая пешка в клетке (x, i) будет означать, что $KS(x \mid n) \geqslant i$ и что $KS(i \mid x) \geqslant \log n - 1$; поскольку действия Белых тоже вычислимы, то при этом $KS(x \mid n) \geqslant i + O(1)$, поскольку множество белых пешек в горизонтали i можно перечислять и оно содержит не более 2^i элементов.

Это рассуждение доказывает не совсем то, что требовалось: мы доказали, что $KS(KS(x \mid n) \mid x) \geqslant \log n - 1$; чтобы получить требуемое, достаточно немного изменить рассуждение и играть параллельно на досках для всех значений n.

46 Докажите, что (для некоторой константы c) для любого слова x и любого числа n найдётся такое слово y длины n, что

$$KS(xy) \geqslant KS(x \mid n) + n - c.$$

[Указание. Для данного n количество слов x, при которых $KS(xy) < k$ при всех y длины n, не превосходит $2^k/2^n$, и это свойство перечислимо, так что можно применить теорему 19 (с. 46).]

47 Пусть f — функция с натуральными аргументами и значениями, причём $f(n) + \varepsilon h \leqslant f(n + h) \leqslant f(n) + (1 - \varepsilon)h$ для некоторого $\varepsilon > 0$ и для любых натуральных n и h. Покажите, что существует бесконечная последовательность ω, у которой сложность начального отрезка длины n равна $f(n) + O(1)$ при любом n.

[Указание. Будем строить такую последовательность добавлением блоков длины h: очередной блок, добавляемый к началу длины n. увеличивает сложность либо больше чем на $f(n + h) - f(h)$, либо меньше; выбор будет зависеть от того, ниже или выше границы мы находимся сейчас. Для поиска усложняющего блока можно применить предыдущую задачу, второй блок можно взять, скажем, из одних нулей. Заметим, что h будет одно и то же для всей последовательности, и поэтому достаточно контролировать сложность начальных отрезков из целого числа блоков.]

48 Докажите, что бесконечная последовательность $x_0x_1x_2\ldots$ нулей и единиц вычислима тогда и только тогда, когда величина $KS(x_0 \ldots x_{n-1} \mid n)$ (условная сложность её начальных отрезков при известной длине) ограничена сверху.

[Указание. Отметим в бесконечном двоичном дереве всех слов перечислимое множество S вершин (слов), у которых условная сложность при известной длине ограничена некоторой фиксированной константой. Все «горизонтальные» сечения множества S ограничены по мощности. Нам нужно вывести из этого, что любая бесконечная ветвь, целиком проходящая по S, вычислима. Будем предполагать, что S образует поддерево (оставив от него только те вершины, для которых путь из корня также лежит в S.) Фиксируем некоторую бесконечную ветвь ω в поддереве S и на каждом уровне n подсчитаем числа l_n и r_n вершин из S слева и справа от ω. Пусть L и R — верхние пределы чисел l_n и r_n, и пусть N — номер уровня,

начиная с которого эти верхние пределы не превышаются. Зная L, R и N, можно вычислять сколь угодно большие начальные отрезки ветви ω: надо искать ветви, слева от которых есть не менее L элементов из S на некотором уровне (большем N) и справа есть не менее R элементов из S на некотором (возможно, другом, также большем N) уровне; как только такая ветвь нашлась, её начальный отрезок до нижнего из этих двух уровней совпадает с ω.]

49 Докажите, что в предыдущей задаче ограниченность $KS(x_0 \ldots x_{n-1} \mid n)$ можно заменить более слабым условием: $KS(x_0 \ldots x_{n-1}) \leqslant \log n + c$ для некоторого c и для всех n. [Указание: возникает перечислимое множество S слов (вершин бесконечного двоичного дерева), в котором число вершин на уровнях ниже N есть $O(N)$. Если бы числа вершин на всех уровнях были бы ограничены, то задача свелась бы к предыдущей. Этого можно добиться, оставив от множества только те вершины x, для которых существует продолжение длины $2l(x)$, все начала которого принадлежат S.]

Задача 48 подсказывает разные возможности определения сложности вычислимых последовательностей $x = (x_0 x_1 x_2 \ldots)$ нулей и единиц:

- наименьшая сложность программы вычисления $x_0 \ldots x_{n-1}$ по данному n (с точностью до константы это то же самое, что наименьшая сложность программы вычисления x_n по данному n), эту сложность мы будем обозначать через $KS(x)$,
- наименьшая сложность программы, которая для всех достаточно больших n вычисляет $x_0 \ldots x_{n-1}$ по данному n, обозначаемая через $KS_\infty(x)$ (для конечного количества n программа может на входе n зависать или давать неверный ответ),
- $\max\{KS(x_0 \ldots x_{n-1} \mid n) \mid n = 0, 1, \ldots\}$, обозначаемая через $M(x)$, и
- $\limsup_{n\to\infty} KS(x_0 \ldots x_{n-1} \mid n)$, обозначаемая через $M_\infty(x)$.

Между этими величинами имеются очевидные неравенства:

$$M_\infty(x) \leqslant M(x) \leqslant KS(x)$$

(с точностью до постоянного слагаемого) и

$$M_\infty(x) \leqslant KS_\infty(x) \leqslant KS(x)$$

(с той же точностью).

50 Покажите, что существуют вычислимые последовательности, для которых $KS_\infty(x)$ значительно меньше $M(x)$ (а значит и $KS(x)$). Точнее, можно указать последовательность x^m вычислимых последовательностей, для которой $KS_\infty(x^m)$ есть $O(\log m)$, а $M(x^m)$ не меньше m. [Указание. Рассмотрите последовательность $x^m = y_m 000\ldots$, где y_m обозначает лексикографически первое слово длины m и сложности не менее m при известном m.]

51 Покажите, что также бывают вычислимые последовательности для которых $M(x)$ значительно меньше $KS(x)$. А именно, можно указать последовательность x^m

вычислимых последовательностей, для которой $M(x^m)$ есть $O(\log m)$, а $KS(x^m)$ не меньше m. [Указание. Рассмотрите последовательность $x^m = (1^{BB(m)}000\ldots)$; число единиц перед нулями равно $BB(m)$, определение этой функции см. на с. 35.]

$\boxed{52}$ Покажите, что $KS_\infty(x) \leqslant 2M_\infty(x) + O(1)$. [Указание. Годится рассуждение из задачи 48.]

Можно показать, что константу 2 в предыдущей задаче уменьшить нельзя, но это уже выходит за рамки нашей книги. Интересующегося читателя мы отсылаем к работе [40].

$\boxed{53}$ Рассмотрим слова длины n, у которых сложность не меньше n. (Такие слова естественно назвать *несжимаемыми*.) (**а**) Покажите, что количество таких слов не меньше 2^{n-c} и не больше $2^n - 2^{n-c}$ (для некоторого c). (**б**) Покажите, что сложность количества несжимаемых слов длины n равна $n - O(1)$ (отсюда следуют утверждения предыдущего пункта!). (**в**) Покажите, что если слово x длины $2n$ является несжимаемым, то его половины x_1 и x_2 (длины n) имеют сложность $n - O(1)$. (**г**) Покажите, что если слово x длины n является несжимаемым, то любое его подслово длины k имеет сложность не меньше $k - O(\log n)$. (**д**) Покажите, что для любой константы $c < 1$ любое несжимаемое слово достаточно большой длины n содержит подслово из $\lfloor c \log_2 n \rfloor$ нулей. [Указание. (а) Всего описаний длины меньше n будет $2^n - 1$, но часть их уходит на более короткие слова: любое слово длины $n - d$ при некотором d имеет сложность меньше n, что доказывает первое утверждение. Чтобы доказать второе, заметим, что слова длины n, начинающиеся с k нулей, можно задать $2\log k + (n - k)$ битами. (б) Если число несжимаемых слов длины n можно описать двоичным словом t длины $n - k$, то по t и $\log k$ битов можно восстановить и n, и список всех несжимаемых слов длины n, поэтому первое из несжимаемых слов будет иметь сложность меньше, чем следует. (в) Если одну из половин можно описать короче, то и всё слово можно задать короче, начав с (беспрефиксного кода) разности между длиной и сложностью сжимаемой половины. (г) Если подслово имеет меньшую сложность, то и всё слово можно описать короче (задав подслово, его положение и остальные биты). (д) Подсчитаем число слов длины n, не содержащих подряд k нулей; рекуррентное соотношение показывает, что это число растёт с увеличением n примерно как геометрическая прогрессии, знаменатель которой есть наибольший действительный корень уравнения $x = 2 - (1/x^k)$, и можно оценить сложность таких слов.]

$\boxed{54}$ Покажите, что (при некотором значении c) для любой бесконечной последовательности $x_0x_1x_2\ldots$ нулей и единиц найдётся бесконечно много значений n, для которых $KS(x_0x_1\ldots x_{n-1}) \leqslant n - \log n + c$.

Покажите, что найдётся последовательность $x_0x_1x_2\ldots$ и константа c, для которых $KS(x_0x_1\ldots x_{n-1}) \geqslant n - 2\log n - c$ при всех n.

[Указание. Ряд $\sum 1/n$ расходится, а ряд $\sum 1/n^2$ сходится. Подробнее см. теоремы 95 и 99.]

$\boxed{55}$ Для данного слова x длины n определим величины $d(x) = n - KS(x)$ и $d_c(x) = n - KS(x \mid n)$. Покажите, что они связаны неравенством

$$d_c(x) - 2\log d_c(x) - O(1) \leqslant d(x) \leqslant d_c(x) + O(1).$$

[Указание. Надо доказать, что если $KS(x\,|\,n) = n - d$, то $KS(x) \leqslant n - d + 2\log d + O(1)$. В самом деле, взяв условное описание p длины $n - d$ и добавив в его начало самоограниченную запись d, мы получим слово, по которому можно восстановить сначала d, потом n, и наконец x.]

56 Докажите, что $d(xy) = d(x) + d(y\,|\,x) + O(\log d(xy))$ для любых двух слов x, y длины n, где $d(u) = l(u) - KS(u)$. [Указание: можно воспользоваться задачей 36.]

(Интуитивный смысл разности между длиной слова и его сложностью как дефекта случайности обсуждается для разных видов сложности в главе 5 и в главе 14.)

57 Покажите, что для достаточно большого значения константы c перечислимое множество пар слов (x, y), для которых $KS(x\,|\,y) < c$, является полным по Тьюрингу (с помощью оракула для этого множества можно алгоритмически решать проблему остановки). [Указание: воспользуйтесь задачей 15.]

2.3. Количество информации

Мы знаем (теорема 18), что условная сложность $KS(y\,|\,x)$ не превосходит безусловной сложности $KS(y)$ (с точностью до константы). Разность $KS(y) - KS(y\,|\,x)$ показывает, насколько знание слова x упрощает описание слова y. Поэтому её называют *количеством информации в слове x о слове y*. Обозначение: $I(x:y)$.

Теорема 18 показывает, что информация $I(x:y)$ неотрицательна (с точностью до константы): существует такое c, что $I(x:y) \geqslant c$ при всех x и y.

58 Пусть дана произвольная вычислимая функция f. Тогда для некоторой константы c выполнено неравенство $I(f(x):y) \leqslant I(x:y) + c$.

59 Докажите обобщение этого свойства на вероятностные алгоритмы. А именно, пусть дана произвольная вычислимая функция $f(x, r)$. Пусть r выбирается с равномерным распределением среди строк некоторой длины n. Тогда для всех l вероятность события $I(f(x, r):y) > I(x:y) + l$ не превосходит $2^{-l+O(KS(n)+KS(l))}$. Это свойство информации называют «невозрастанием информации при алгоритмических преобразованиях». Пользуясь метафорой Левина, смысл этого неравенства можно выразить словами «пытка неинформированного свидетеля не даст информации о преступлении» [82]. [Указание. Используйте условный вариант задачи 41.]

Вспомнив, что

$$KS(x, y) = KS(x) + KS(y\,|\,x) + O(\log KS(x, y))$$

(теорема 22, с. 49), можно выразить условную сложность через безусловные:

$$KS(y\,|\,x) = KS(x, y) - KS(x) + O(\log KS(x, y)).$$

Тогда для информации получается выражение:

$$I(x:y) = KS(y) - KS(y\,|\,x) = KS(x) + KS(y) - KS(x, y) + O(\log KS(x, y)).$$

Отсюда сразу вытекает

Теорема 23 (симметрия информации).

$$I(x:y) = I(y:x) + O(\log KS(x,y)).$$

Эта теорема гарантирует, что разница между $I(x:y)$ и $I(y:x)$ логарифмически мала по сравнению с $KS(x,y)$. Как показывает следующая задача, эта разница может быть сравнима с самими значениями $I(x:y)$ и $I(y:x)$, если те много меньше $KS(x,y)$.

$\boxed{\mathbf{60}}$ Покажите, что если x — слово длины n, для которого $KS(x \mid n) \geqslant n$, то $I(x:n) = KS(n) + O(1)$, в то время как $I(n:x) = O(1)$.

Симметрия (пусть и не полная, а лишь с точностью до логарифмического слагаемого) позволяет называть $I(x:y)$ (или $I(y:x)$) *взаимной информацией* слов x и y. Соотношения между взаимной информацией, условными сложностями и сложностью пары можно изобразить на символической картинке (рис. 2.2).

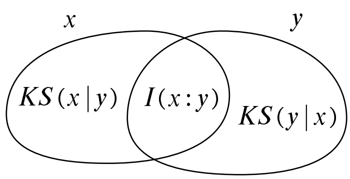


Рис. 2.2. Взаимная информация и условная сложность.

На ней показано, что слова x и y имеют $I(x:y) \approx I(y:x)$ битов общей информации. Добавив $KS(x \mid y)$ битов (информация, которая есть в x, но не в y, левая область), мы получаем

$$I(y:x) + KS(x \mid y) \approx (KS(x) - KS(x \mid y)) + KS(x \mid y) \approx KS(x)$$

битов (как и должно быть в слове x). Аналогичным образом центральная часть вместе с $KS(y \mid x)$ справа дают $KS(y)$. Наконец, все три области вместе складываются в

$$KS(x \mid y) + I(x:y) + KS(y \mid x) = KS(x) + KS(y \mid x) = KS(x \mid y) + KS(y) = KS(x,y)$$

(все равенства верны с точностью $O(\log n)$, если слова x и y имеют длину не больше n).

В некоторых случаях эту картинку можно понимать буквально. Возьмём, например, «несжимаемое» слово $r = r_1 \ldots r_n$ длины n, для которого $KS(r_1 \ldots r_n) \geqslant n$. Тогда сложность любого его подслова u равна $l(u)$ с точностью до $O(\log n)$. В самом деле, поскольку u является подсловом r, то $r = tuv$ для некоторых слов t и v. Тогда $l(r) \leqslant KS(r) \leqslant KS(t) + KS(u) + KS(v) \leqslant l(t) + l(u) + l(v) = l(r)$ (с логарифмической точностью) и потому все неравенства обращаются в равенства (с той же точностью).

Если теперь взять два перекрывающихся подслова x и y (рис. 2.3), то $KS(x)$ будет равно длине x, $KS(y)$ будет равно длине y (с точностью до $O(\log n)$).

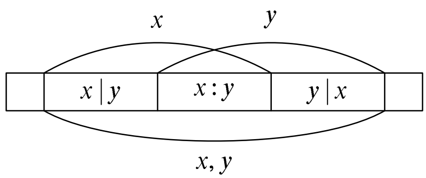


Рис. 2.3. Общая информация в перекрывающихся подсловах.

Сложность пары $KS(x, y)$ будет равна длине объединения отрезков (поскольку пара x, y отличается от этого объединения лишь информацией о длинах, которая требует $O(\log n)$ битов).

Следовательно, и условные сложности $KS(x \mid y)$, $KS(y \mid x)$, и общая информация $I(x:y)$ равны (с логарифмической точностью) длинам соответствующих отрезков.

Однако такая ситуация имеет место далеко не всегда. Как мы впоследствии убедимся (см. главу 11), далеко не для всяких двух слов x и y, имеющих большу́ю взаимную информацию $I(x:y)$, эту взаимную информацию можно «материализовать» в виде слова z, для которого $KS(z \mid x) \approx 0$, $KS(z \mid y) \approx 0$ и $KS(z) \approx I(x:y)$. (В нашем последнем примере таким z было пересечение подслов x и y.)

61 Докажите, что для любого слова x длины не больше n математическое ожидание количества общей информации в x и случайном слове длины n есть $O(\log n)$.

Рассмотрим теперь сложностные характеристики трёх слов. Здесь важным приёмом является *релятивизация* — перенесение свойств безусловной сложности на условную. Поясним это на примере.

Теорема о сложности пары (с. 40) утверждает, что

$$KS(x, y) \leqslant KS(x) + 2\log KS(x) + KS(y) + O(1).$$

Если все сложности заменить на условные (при известном z), получится неравенство

$$KS(x, y \mid z) \leqslant KS(x \mid z) + 2\log KS(x \mid z) + KS(y \mid z) + O(1),$$

где под условной сложностью пары x, y при известном z понимается сложность её кода: $KS(x, y \mid z) = KS([x, y] \mid z)$. Как и для безусловной сложности пары, выбор кодирования несуществен (меняет сложность на $O(1)$).

Это неравенство доказывается точно так же, как и неравенство без z: мы соединяем описания p для x (при известном z) и q для y (при известном z) в единое описание $\hat{p}q$ относительно нового способа (условного) описания.

Поучительно исключить из этого неравенства условные сложности, заменив их на безусловные: $KS(x, y \mid z) = KS(x, y, z) - KS(z)$, а также $KS(x \mid z) = KS(x, z) - KS(z)$ и $KS(y \mid z) = KS(y, z) - KS(z)$ (с логарифмической точностью). При

этом мы приходим к такой теореме:

Теорема 24.

$$KS(x,y,z) + KS(z) \leqslant KS(x,z) + KS(y,z) + O(\log n)$$

для слов x, y, z сложности не более n.

Это неравенство иногда называют *базисным.*

Релятивизацию можно применить и к теореме 21 (с. 47), связывающей сложность пары и условную сложность. Получается такое утверждение (мы формулируем его с логарифмической точностью):

Теорема 25.

$$KS(x,y\,|\,z) = KS(x\,|\,z) + KS(y\,|\,x,z) + O(\log n),$$

если слова x, y, z имеют сложность не больше n.

Здесь $KS(x\,|\,y,z)$ понимается как сложность x относительно кода пары $\langle y,z\rangle$, то есть как $KS(x\,|\,[y,z])$. Как и раньше, вычислимое кодирование пар может быть любым (сложность изменится не более чем на константу).

◄ Можно воспроизвести доказательство теоремы 21, заменив везде описания без условий на описания при известном z. При этом $KS(y\,|\,x)$ превратится в $KS(y\,|\,x,z)$. Можно сказать, что мы теперь имеем трёхмерное пространство с координатами x, y, z и проводим все прежние рассуждения одновременно во всех плоскостях, параллельных плоскости xy.

Если это кажется слишком сложным, можно просто выразить все условные сложности через безусловные: в левой части получим

$$KS(x,y\,|\,z) = KS(x,y,z) - KS(z),$$

а в правой

$$KS(x\,|\,z) + KS(y\,|\,x,z) = KS(x,z) - KS(z) + KS(y,x,z) - KS(x,z),$$

так что после сокращения левая и правая части становятся равными (с логарифмической точностью). (Отметим в скобках, что это более простое рассуждение даёт худшие значения констант в $O(\log n)$-обозначениях.) ►

$\boxed{62}$ Проверьте, что в теореме 25 достаточно потребовать, чтобы условные сложности $KS(x\,|\,z)$ и $KS(y\,|\,x,z)$ не превосходили n.

Можно «релятивизовать» определение информации, определив $I(x:y\,|\,z)$ как разность $KS(y\,|\,z) - KS(y\,|\,x,z)$. Как и для случая безусловной информации, эта величина неотрицательна (с точностью до $O(1)$). Выразив условную сложность через безусловную (с логарифмической точностью), можно записать неотрицательность $I(x:y\,|\,z)$ как

$$KS(y\,|\,z) - KS(y\,|\,x,z) = KS(y,z) - KS(z) - KS(y,x,z) + KS(x,z) \geqslant 0,$$

приходя к неравенству теоремы 24.

Вообще большинство известных равенств и неравенств, касающихся безусловных сложностей, условных сложностей и информации (с логарифмической точностью), являются прямыми следствиями «базисных неравенств» — теорем 21 и 24. Первые примеры неравенств, которые не выводятся из базисных, были найдены сравнительно недавно [193, 194], и они довольно сложные. Мы обсудим их в разделе 10.13, а пока рассмотрим несколько простых следствий базисных неравенств.

Независимые слова. Будем называть слова x и y «независимыми», если $I(x:y) \approx \approx 0$. (Степень точности должна ещё уточняться, но мы предполагаем, что членами порядка $O(\log n)$ мы пренебрегаем, если n — максимальная из длин (или сложностей) используемых слов.)

Независимые слова можно рассматривать как аналог независимых случайных величин в теории вероятностей. Есть такая теорема: если случайная величина ξ независима с парой величин $\langle\alpha,\beta\rangle$, то она независима и с величинами α и β по отдельности.

Для колмогоровской сложности аналогичное утверждение (если слово x независимо с парой $\langle y,z\rangle$, то оно независимо и с её отдельными компонентами) естественно выражается неравенством

$$I(x:\langle y,z\rangle) \geqslant I(x:y)$$

(и аналогичным неравенством для z вместо y). Это неравенство справедливо (как обычно, с логарифмической точностью), и в этом легко убедиться, переписав его в терминах безусловных сложностей:

$$KS(x) + KS(y,z) - KS(x,y,z) \geqslant KS(x) + KS(y) - KS(x,y),$$

что после приведения подобных членов даёт базисное неравенство теоремы 24.

Сложность пар и троек. Напротив, для доказательства следующей теоремы (упоминавшейся на с. 22) полезно заменить безусловные сложности на условные:

Теорема 26.

$$2\,KS(x,y,z) \leqslant KS(x,y) + KS(x,z) + KS(y,z) + O(\log n),$$

если x,y,z — слова сложности не более n.

◄ Перенося $KS(x,y)$ и $KS(x,z)$ в левую часть, и заменяя разности $KS(x,y,z) - KS(x,y)$ и $KS(x,y,z) - KS(x,z)$ на условные сложности $KS(z\,|\,x,y)$ и $KS(y\,|\,x,z)$, мы получаем такое неравенство:

$$KS(z\,|\,x,y) + KS(y\,|\,x,z) \leqslant KS(y,z) + O(\log n).$$

Остаётся переписать правую часть неравенства как $KS(y) + KS(z\,|\,y)$ и заметить, что $KS(z\,|\,x,y) \leqslant KS(z\,|\,y)$ и $KS(y\,|\,x,z) \leqslant KS(y)$ (с логарифмической точностью). ►

А можно вместо этого формально сложить два неравенства (базисное и для сложности пары):

$$KS(x,y,z) + KS(y) \leqslant KS(x,y) + KS(y,z) + O(\log n),$$
$$KS(x,y,z) \leqslant KS(y) + KS(x,z) + O(\log n),$$

после чего сократить $KS(y)$ в обеих частях и получить требуемое. (К сожалению, это доказательство, как и предыдущее, несимметрично.) Мы вернёмся к этому неравенству и его геометрическим следствиям в главе 10.

Различные сложностные характеристики трёх слов можно систематизировать следующим образом. Имеются семь основных таких характеристик (три сложности одиночных слов, три сложности пар и сложность всей тройки). Другие характеристики (условная сложность, информация) выражаются через них. Чтобы лучше представить себе, какие ограничения накладываются на эти семь сложностей, удобно перейти к новым координатам. Введём переменные $a_1, a_2, \ldots, a_7$, соответствующие семи областям на рис. 2.4.

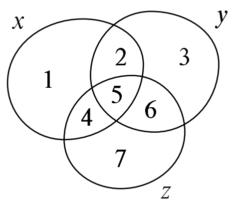


Рис. 2.4. Новые координаты $a_1, a_2, \ldots, a_7$.

Будем считать, что

$$\begin{aligned}
KS(x) &= a_1 + a_2 + a_4 + a_5,\\
KS(y) &= a_2 + a_3 + a_5 + a_6,\\
KS(z) &= a_4 + a_5 + a_6 + a_7,\\
KS(x,y) &= a_1 + a_2 + a_3 + a_4 + a_5 + a_6,\\
KS(x,z) &= a_1 + a_2 + a_4 + a_5 + a_6 + a_7,\\
KS(y,z) &= a_2 + a_3 + a_4 + a_5 + a_6 + a_7,\\
KS(x,y,z) &= a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7.
\end{aligned}$$

Легко проверить, что эти равенства задают обратимое линейное преобразование семимерного пространства (каждому набору из семи сложностей соответствует единственное значение переменных $a_1, \ldots, a_7$).

Различные условные сложности и взаимные информации (с логарифмической точностью) выражаются через безусловные, поэтому можно их выразить и в новых координатах. Например, $I(x:y) = KS(x) + KS(y) - KS(x,y) = a_2 + a_5$, а $KS(x\,|\,y) = KS(x,y) - KS(y) = a_1 + a_4$.

Каков смысл наших новых координат? Легко проверить, что $a_1 = KS(x\,|\,y,z)$ (с логарифмической точностью). Аналогичный смысл имеют a_3 и a_7. Координата a_2 равна (с той же точностью) $I(x:y\,|\,z)$; аналогичный смысл имеют a_4 и a_6 (см. рис. 2.5). Отсюда, в частности, вытекает, что для любых слов x, y, z соответствующие значения координат $a_1, a_2, a_3, a_4, a_6, a_7$ неотрицательны (с точностью до $O(\log n)$, если слова x, y, z имеют сложность не больше n).

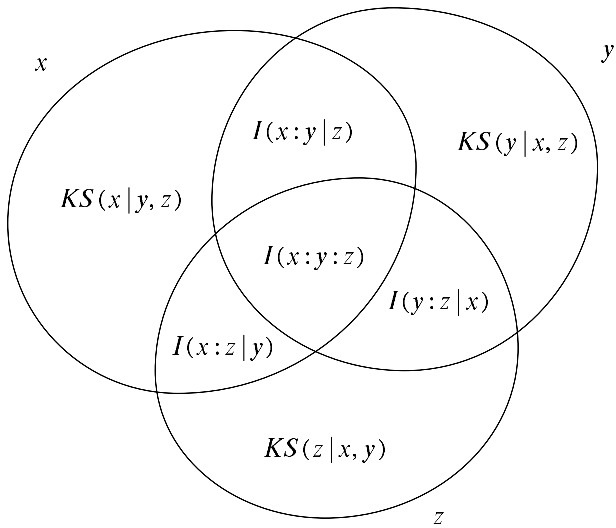


Рис. 2.5. Сложностной смысл новых координат.

Ситуация с координатой a_5 сложнее. Её хотелось бы интерпретировать как «общую информацию, содержащуюся в трёх словах x, y, z». Иногда для неё используется обозначение $I(x:y:z)$. Однако смысл такого выражения далеко не очевиден, особенно если иметь в виду, что a_5 может быть отрицательным. Рассмотрим такой пример. Пусть x и y — две половины несжимаемого слова длины $2n$. Тогда $KS(x) = n$, $KS(y) = n$, $KS(x, y) = 2n$ и $I(x:y) = 0$ (с логарифмической точностью). Рассмотрим слово z длины n, которое является побитовой суммой x и y по модулю 2. Тогда любое из трёх слов x, y, z может быть восстановлено по двум другим, следовательно, сложности $KS(x, y), KS(y, z)$ и $KS(x, z)$ одинаковы и равны $2n$ (с логарифмической точностью), и сложность $KS(x, y, z)$ также равна $2n$. Сложность слова z равна n (она не больше n, так как его длина равна n, но и не меньше, так как иначе вместе с y не получилось бы пары сложности $2n$).

После этого можно вычислить значения $a_1, \ldots, a_7$ для этого примера (рис. 2.6).

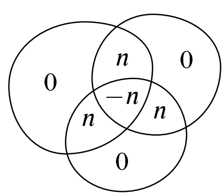


Рис. 2.6. Два независимых несжимаемых слова длины n и их сумма mod 2.

Таким образом координата a_5 может быть отрицательной, однако суммы $a_5 + a_2$, $a_5 + a_4$ и $a_5 + a_6$, будучи взаимными информациями пар слов, должны быть неотрицательны. (В нашем примере эти суммы равны нулю.)

Этот пример соответствует простейшему способу «разделения секрета» z между двумя лицами: если одному человеку сообщить x, а другому y, то ни один из них по отдельности не будет ничего знать о z (поскольку $I(x:z) \approx 0$ и $I(y:z) \approx 0$), но вместе они могут восстановить z, побитово сложив x и y.

Можно проверить, что указанными неравенствами (неотрицательность всех a_i, кроме a_5, и неотрицательность трёх названных сумм) исчерпываются все линейные неравенства, имеющиеся для сложностных характеристик трёх слов. Мы вернёмся к этому вопросу в главе 10.

В качестве примера использования таких диаграмм докажем снова неравенство

$$2\,KS(x,y,z) \leqslant KS(x,y) + KS(x,z) + KS(y,z).$$

В наших новых переменных оно записывается как $a_2 + a_4 + a_5 + a_6 \geqslant 0$ (это легко проверить в уме, подсчитав, сколько раз входит каждое из a_i в левую и правую часть неравенства). Теперь остаётся заметить, что $a_2 + a_5 \geqslant 0$, $a_4 \geqslant 0$ и $a_6 \geqslant 0$.

63 Докажите, что $I(xy:z) = I(x:z) + I(y:z\,|\,x) + O(\log n)$ для слов x, y, z сложности не более n. [Указание: это легко объяснить на диаграмме.]

Эта задача показывает, что информация в xy о z складывается из двух частей: информации в x о z и информации в y о z (при известном x). Можно рассматривать её как аналог равенства $KS(x,y) = KS(x) + KS(y\,|\,x)$, только теперь вместо сложности стоит информация о слове z. В качестве следствия получаем, что если xy независимо с z, то x независимо с z, а также y независимо с z при известном x. (Под независимостью понимается малость взаимной информации.) Симметричное рассуждение показывает, что y независимо с z, а также x независимо с z при известном y.

64 Покажите, что свойства «x и y независимы» и «x и y независимы при известном z» не следуют одно из другого: любое из них может быть истинно, в то время как второе ложно.

65 Будем говорить, что слова x, y, z, t образуют *марковскую цепь* (по аналогии с соответствующим понятием теории вероятностей), если $I(x:z\,|\,y)$ и $I(\langle x,y\rangle : t\,|\,z)$ малы (разумеется, точный смысл определения зависит от выбора границы малости). Покажите, что в этом случае эти же слова в обратном порядке образуют марковскую цепь, то есть что $I(t:y\,|\,z)$ и $I(\langle t,z\rangle : x\,|\,y)$ малы. [Указание. $I(\langle x,y\rangle : t\,|\,z) = I(y:t\,|\,z) + I(x:t\,|\,y,z)$, и равенство нулю левой части означает обращение в нуль обоих слагаемых справа; второе слагаемое симметрично относительно обращения порядка.]

3. Случайность по Мартин-Лёфу

В этой главе мы прерываем изложение основных свойств колмогоровской сложности, чтобы определить другое базисное понятие алгоритмической теории информации — понятие случайной по Мартин-Лёфу, или «типичной», последовательности. (Материал этой главы не использует предыдущей и не используется в следующей; он понадобится в главе 5, где будет дан критерий случайности в терминах колмогоровской сложности.)

Мы начнём с напоминания основных фактов о мерах на последовательностях нулей и единиц.

3.1. Пространство Ω и меры

Рассмотрим пространство $\Omega = \mathbb{B}^{\mathbb{N}}$, элементами которого являются бесконечные последовательности нулей и единиц. Его называют *канторовским* пространством. Для каждого двоичного слова x мы рассмотрим множество Ω_x всех продолжений этого слова. Например, Ω_{00} есть множество всех последовательностей, начинающихся с двух нулей, а $\Omega_\Lambda = \Omega$ (здесь Λ обозначает пустое слово).

Множества вида Ω_x мы будем называть *интервалами*. Интервалы и всевозможные их объединения называют *открытыми* подмножествами пространства Ω. Эта топология соответствует стандартной метрике, в которой расстояние между двумя последовательностями тем меньше, чем больше у них совпадающее начало:

$$d(\omega, \omega') = 2^{-n},$$

где n — наименьшее число, для которого $\omega_n \neq \omega'_n$. (Через ω_n мы обозначаем член с номером n, начиная нумерацию с нуля: $\omega = \omega_0\omega_1\omega_2\ldots$)

$\boxed{\textbf{66}}$ Покажите, что это пространство гомеоморфно канторовскому множеству, которое получится, если из отрезка выбросить среднюю треть, из двух оставшихся частей также выбросить по средней трети и так далее.

Нас, однако, будет интересовать не столько топология, сколько мера. Семейство подмножеств пространства Ω называют *σ-алгеброй*, если оно замкнуто относительно конечных и счётных пересечений и объединений, а также перехода к дополнению.

Минимальную σ-алгебру, содержащую все множества Ω_x (и тем самым все открытые множества), называют алгеброй *борелевских* множеств.

Рассмотрим произвольную σ-алгебру, содержащую все интервалы. Пусть на ней задана функция μ, которая ставит в соответствие каждому множеству из σ-алгебры

неотрицательное число, причём выполнено свойство *σ-аддитивности*:

> если множество A есть объединение конечного или счётного числа непересекающихся множеств $A_0, A_1, A_2, \ldots$, и все эти множества принадлежат σ-алгебре, на которой определена функция μ, то
>
> $$\mu(A) = \mu(A_0) + \mu(A_1) + \mu(A_2) + \ldots$$
>
> (в правой части стоит конечная сумма или сходящийся бесконечный ряд с неотрицательными членами).

Такие функции называют *мерами* на пространстве Ω; множества, на которых мера определена, называют *измеримыми* (по этой мере), и $\mu(A)$ называют мерой множества A.

Если мера всего пространства Ω равна 1, то меру называют *распределением вероятностей*, элементы σ-алгебры называют *событиями*, а число $\mu(A)$ называют *вероятностью* события A.

Любая мера монотонна (если $A \subset B$, то $\mu(A) \leqslant \mu(B)$). В самом деле, $\mu(B) - \mu(A) = \mu(B \setminus A) \geqslant 0$.

Другое важное свойство мер — непрерывность: если множество B есть объединение возрастающей последовательности множеств

$$B_0 \subset B_1 \subset B_2 \subset \ldots,$$

то $\mu(B)$ равно пределу $\mu(B_i)$ при $i \to \infty$. (В самом деле, применяем счётную и конечную аддитивность к множествам $A_i = B_i \setminus B_{i-1}$.) Аналогичное свойство непрерывности верно и для убывающих последовательностей множеств.

Для каждой меры μ на Ω можно рассмотреть функцию p на двоичных словах, определённую равенством $p(x) = \mu(\Omega_x)$. Эта функция принимает неотрицательные значения, при этом выполнено свойство *аддитивности*:

$$p(x) = p(x0) + p(x1)$$

для любого слова x. (В самом деле, интервал Ω_x есть дизъюнктное объединение двух его половин Ω_{x0} и Ω_{x1}.)

Теория меры (теорема о продолжении меры по Лебегу) позволяет выполнить и обратный переход. Именно, для каждой функции p на двоичных словах, принимающей неотрицательные действительные значения и обладающей свойством аддитивности, можно построить меру μ, для которой $\mu(\Omega_x) = p(x)$ при всех x. Эта мера обладает дополнительным свойством (которое мы будет предполагать в дальнейшем: если $\mu(A) = 0$ и $B \subset A$, то B измеримо (отсюда уже следует, что $\mu(B) = 0$.

Мы не приводим конструкции продолжения меры по Лебегу (она есть в любом учебнике по теории меры, например, [66] или [50]), но приведём явное описание измеримых множеств меры нуль, возникающих при этом построении.

Напомним, что у нас фиксирована неотрицательная функция p на словах, обладающая свойством аддитивности. Будем называть число $p(x)$ мерой интервала Ω_x. Говорят, что множество $A \subset \Omega$ является *нулевым*, если для всякого $\varepsilon > 0$ можно

найти конечное или счётное покрытие множества A интервалами с суммой мер не больше ε.

Другими словами, множество A нулевое, если существует функция $\langle\varepsilon, i\rangle \mapsto$ $\mapsto x(\varepsilon, i)$ (первый аргумент — положительное действительное число, второй — натуральное число, значениями являются двоичные слова), для которой

- $A \subset \Omega_{x(\varepsilon,0)} \cup \Omega_{x(\varepsilon,1)} \cup \Omega_{x(\varepsilon,2)} \ldots$ и
- $p(x(\varepsilon,0)) + p(x(\varepsilon,1)) + p(x(\varepsilon,2)) + \ldots \leqslant \varepsilon$

при любом положительном ε. При этом возможность конечного покрытия мы учитываем, не требуя, чтобы $x(\varepsilon, i)$ было определено при всех ε и i (неопределённые значения пропускаются в объединении и в сумме).

Так вот, измеримыми множествами меры нуль при продолжении по Лебегу будут как раз нулевые в описанном смысле множества, так что мы употребляем выражения «нулевое множество» и «множество меры нуль» как синонимы. (По большей части нам будет достаточно приведённого выше определения нулевого множества.)

Несколько простых наблюдений, которые нам пригодятся в дальнейшем:

- В определении нулевого множества можно было бы ограничиться рациональными значениями ε (или только значениями вида 2^{-k}).
- Всякое подмножество нулевого множества является нулевым.
- Конечное или счётное объединение нулевых множеств является нулевым. (В самом деле, чтобы найти покрытие объединённого множества с суммой мер меньше ε, достаточно соединить покрытия его частей с суммой мер меньше $\varepsilon/2, \varepsilon/4, \varepsilon/8, \ldots$)
- Пусть функция p такова, что каждая точка имеет меру нуль (это будет так, если для любой бесконечной последовательности $\omega = \omega_0\omega_1\omega_2\ldots$ предел мер $p(\omega_0\ldots\omega_n)$ при $n \to \infty$ равен нулю). Тогда любое конечное или счётное множество является нулевым.

Равномерная мера на пространстве Ω получится, если положить меру интервала Ω_x равной $2^{-l(x)}$:

$$p(x) = 2^{-n} \text{ для всех слов } x \text{ длины } n.$$

Возникающая при этом мера тесно связана с обычной мерой на действительной прямой (точнее, на отрезке $[0,1]$): мера множества $A \subset \Omega$ равна мере множества действительных чисел, которое получится, если каждую последовательность из A считать бесконечной двоичной дробью. (Замечание в скобках: соответствие между двоичными дробями и числами не совсем взаимно однозначно, так как двоично-рациональные числа имеют два представления: например, $0{,}01111\ldots = 0{,}10000\ldots$ Но это затрагивает лишь счётное множество чисел, которое пренебрежимо с точки зрения теории меры.) В самом деле, числа, двоичные записи которых начинаются на данную строку x длины n, заполняют промежуток длины 2^{-n}. Отсюда легко заключить, что для любого отрезка $I \subset [0,1]$ равномерная мера множества последовательностей, являющихся двоичными разложениями чисел из I, равна длине отрезка I.

С точки зрения теории вероятностей равномерная мера соответствует независимым бросаниям честной монеты. В самом деле, при независимых бросаниях честной монеты все возможные последовательности длины n равновероятны и имеют вероятность 2^{-n}. Множество Ω_x есть событие «полученная в результате бросаний последовательность начинается на x» и имеет меру $2^{-l(x)}$.

Можно рассматривать и несимметричную монету, по-прежнему предполагая бросания независимыми. Соответствующая мера (распределение вероятностей) называется *бернуллиевой* с параметрами q, p (вероятности выпадения нуля и единицы; предполагается, что $p, q \geqslant 0$ и $p + q = 1$).

При этом вероятность появления последовательности, начинающейся на слово x, равна $q^u p^v$, где u и v — число нулей и единиц в этом слове. Другими словами, мы рассматриваем функцию

$$x \mapsto q^{u(x)} p^{v(x)},$$

где через $u(x)$ и $v(x)$ обозначено число нулей и единиц в слове x. Легко проверить, что эта функция аддитивна.

3.2. Усиленный закон больших чисел

Чтобы проиллюстрировать все введённые понятия в действии, сформулируем и докажем *усиленный закон больших чисел*.

Пусть $p + q = 1$, $p, q \geqslant 0$. Рассмотрим множество A_p всех бесконечных последовательностей $\omega_0\omega_1\omega_2\ldots$ нулей и единиц, в которых предел частоты единиц существует и равен p, то есть

$$\lim_{n\to\infty} \frac{\omega_0 + \omega_1 + \ldots + \omega_{n-1}}{n} = p.$$

Теорема 27. *Множество A_p имеет меру* 1 *относительно бернуллиева распределения вероятностей* (*с параметрами p и q*).

Другими словами, дополнение множества A_p, то есть множество тех последовательностей, в которых частота единиц не стремится ни к какому пределу или имеет предел, отличный от p, является нулевым.

◀ Мы докажем эту теорему для равномерной меры (то есть при $p = q = 1/2$) с помощью явного подсчёта. Для произвольного p доказательство будет кратко намечено в одном из упражнений (см. также раздел 9.6).

Сначала рассмотрим «допредельную» ситуацию, зафиксировав некоторое число n. Все последовательности из n нулей и единиц равновероятны, и нам нужно доказать, что большинство из них содержит примерно $n/2$ единиц. Пусть выбрана некоторая граница $\varepsilon > 0$. Посмотрим, сколько последовательностей имеет более $(1/2 + \varepsilon)n$ единиц. Другими словами, мы должны просуммировать в треугольнике Паскаля часть n-й строки, которая начинается несколько правее середины. Всего в этой части не более n слагаемых и они убывают, поэтому сумму можно оценить как первое слагаемое, умноженное на n. (На самом деле нам не нужна больша́я точность, и полиномиальные от n множители нам не повредят, так что ими мы сразу пренебрегаем. В частности, мы опускаем множитель n.)

Первое слагаемое равно

$$\frac{n!}{k!\,(n-k)!},$$

где k — ближайшее справа к $(1/2+\varepsilon)n$ целое число. Воспользуемся формулой Стирлинга:

$$m! = \sqrt{(2\pi + o(1))m}\left(\frac{m}{e}\right)^m,$$

где e — основание натуральных логарифмов. Отбрасывая полиномиальные множители и используя обозначения $u = k/n$, $v = (n-k)/n$, получаем

$$\frac{n!}{k!\,(n-k)!} \approx \frac{(n/e)^n}{(k/e)^k((n-k)/e)^{n-k}} = \frac{n^n}{k^k(n-k)^{n-k}} \approx \\ \approx \frac{n^n}{(un)^{un}(vn)^{vn}} = \frac{1}{u^{un}v^{vn}} = 2^{H(u,v)n},$$

где

$$H(u,v) = -u\log u - v\log v.$$

Число $H(u,v)$ называется *шенноновской энтропией* случайной величины с двумя значениями, принимаемыми с вероятностями u и v. (Подробно шенноновская энтропия обсуждается в главе 7.) На рисунке 3.1 показан соответствующий график

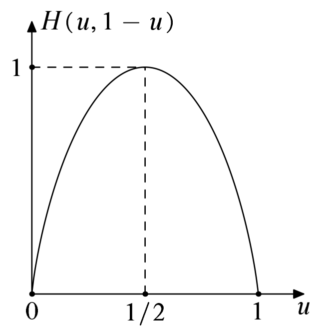


Рис. 3.1. Шенноновская энтропия как функция u.

(напомним, что $v = 1-u$). Легко проверить, что величина $H(u,1-u)$ достигает максимума (равного 1) в единственной точке $u = 1/2$.

Вернёмся к нашей задаче: число последовательностей длины n, у которых частота единиц больше $(1/2+\varepsilon)$, не превосходит $\mathrm{poly}(n)2^{H(1/2+\varepsilon,1/2-\varepsilon)n}$, то есть не превосходит $2^{cn+o(n)}$, где c — некоторая константа, меньшая единицы (зависящая от ε). Тем самым доля таких последовательностей экспоненциально убывает с ростом n. Точно такую же долю составляют последовательности, у которых доля единиц меньше $(1/2-\varepsilon)$.

Подведём промежуточный итог. Для любого фиксированного $\varepsilon > 0$ мы доказали такое утверждение:

Лемма. *Доля последовательностей длины n, у которых частота единиц отклоняется от $1/2$ более чем на ε, среди всех последовательностей длины n не превосходит δ_n, где δ_n экспоненциально убывает с ростом n.*

Эта лемма (если ограничиться утверждением о сходимости и не накладывать никаких ограничений на скорость сходимости к нулю) называется *законом больших чисел* (без слова «усиленный»). Нам же для продолжения доказательства важно знать, что ряд $\sum_n \delta_n$ сходится.

Мы хотим доказать, что множество $A_{1/2}$ тех последовательностей, у которых предел частоты единиц в начальных отрезках равен $1/2$, имеет меру 1. Другими словами, мы хотим показать, что его дополнение (обозначим его B) является нулевым множеством.

По определению предела B есть объединение по всем $\varepsilon > 0$ множеств B_ε, где B_ε есть множество тех последовательностей, у которых частота единиц бесконечное число раз бывает больше $1/2 + \varepsilon$ или меньше $1/2 - \varepsilon$.

Очевидно, достаточно рассматривать счётное число различных значений ε (скажем, только рациональные), а счётное объединение нулевых множеств есть нулевое множество. Осталось доказать, таким образом, что множество B_ε является нулевым при любом фиксированном ε.

Множество B_ε состоит из тех последовательностей, у которых есть сколь угодно длинные плохие начальные отрезки (если считать плохим двоичное слово, в котором частота единиц отличается от $1/2$ более чем на ε). Поэтому при любом N множество B_ε покрывается интервалами вида Ω_x, если взять все плохие x длины N и более. Но суммарная (равномерная) мера всех этих интервалов не превосходит

$$\delta_N + \delta_{N+1} + \delta_{N+2} + \ldots,$$

и эта сумма может быть сделана сколь угодно малой, так как ряд $\sum_i \delta_i$ сходится.

(Это рассуждение называется в теории вероятностей *леммой Бореля – Кантелли.* В общем виде эта лемма гласит, что если сумма мер множеств $A_0, A_1, \ldots$ конечна, то множество тех точек, которые принадлежат бесконечно многим множествам A_i, является нулевым.) ►

Оценить число плохих слов длины n можно и не используя формулы Стирлинга. Будем рассматривать плохие слова, у которых частоты слишком велики (больше $1/2 + \varepsilon$). Рассмотрим на множестве всех слов данной длины n два распределения вероятностей. Первое из них (назовём его L) — равномерное: все слова имеют вероятность 2^{-n}. Второе (назовём его S) отдаёт предпочтение единицам и соответствует n независимым бросаниям монеты, у которой вероятность единицы есть $p = 1/2 + \varepsilon$. При этом вероятность слова из u нулей и v единиц равна $q^u p^v$ (где $q = 1/2 - \varepsilon$ — вероятность нуля). Чем больше в слове x единиц, тем больше отношение вероятностей $S(x)/L(x)$. Легко подсчитать, что для всех плохих слов это отношение не меньше $2^n/2^{H(p,q)n}$, и потому суммарная L-мера всех

плохих слов во столько же раз меньше их суммарной S-меры, которая не превосходит 1. Отсюда получаем, что доля плохих слов не больше $2^{H(p,q)n}/2^n$, то есть мы доказали лемму другим способом. Это доказательство технически проще (хотя до него труднее догадаться). Другое его достоинство в том, что с его помощью можно доказать усиленный закон больших чисел не только для равномерной меры, но и для произвольной бернуллиевой меры (для произвольного p).

$\boxed{\mathbf{67}}$ Проведите подробно такое доказательство. [Указание. Пусть p_0 и q_0 — фиксированные положительные числа, причём $p_0+q_0=1$. Тогда выражение $-p_0\log p - q_0\log q$, где p,q — произвольные положительные числа и $p+q=1$, достигает минимума при $p=p_0$, $q=q_0$. См. также раздел 9.6, с. 308.]

Усиленный закон больших чисел часто формулируют таким образом: «в случайной (по равномерной мере) последовательности частота единиц стремится к $1/2$» (аналогично для других бернуллиевых мер). При этом слова «случайная последовательность» используются не буквально, а как оборот речи: выражение «случайная последовательность обладает свойством α» воспринимается как единое целое и означает, что множество всех последовательностей, не обладающих свойством α, является нулевым.

Возникает естественный вопрос: нельзя ли придать этому обороту буквальный смысл. Именно, выберем какую-либо меру на пространстве Ω, скажем, равномерную. Хотелось бы выделить в пространстве Ω некоторое подмножество, называемое множеством случайных последовательностей, причём так, чтобы для любого свойства α следующие два утверждения были бы эквивалентны:

- все случайные последовательности обладают свойством α;
- множество последовательностей, не обладающих свойством α, имеет меру нуль.

Другими словами, меру 1 должны иметь те и только те множества, которые содержат внутри себя все случайные последовательности. Ещё одна переформулировка: множество случайных последовательностей должно быть минимальным по включению множеством меры 1, а множество неслучайных последовательностей должно быть наибольшим по включению нулевым множеством.

Легко понять, что этот план невыполним: каждая точка пространства Ω образует одноэлементное множество, являющееся нулевым. Вместе с тем объединение этих одноэлементных множеств покрывает всё пространство.

В 1965 году ученик Колмогорова шведский математик Пер Мартин-Лёф обнаружил, что положение можно спасти, если рассматривать не все нулевые множества, а только «эффективно нулевые». Удивительным образом среди них есть наибольшее по включению, и потому можно определить понятие случайной последовательности так, чтобы свойство α выполнялось для всех случайных последовательностей тогда и только тогда, когда множество последовательностей, не обладающих этим свойством, является эффективно нулевым. В следующем разделе мы изложим конструкцию Мартин-Лёфа.

3.3. Эффективно нулевые множества

Пусть фиксирована некоторая мера на пространстве Ω; меру интервала Ω_x мы обозначаем $p(x)$.

Говорят, что множество $A \subset \Omega$ является эффективно нулевым (по данной мере), если по любому $\varepsilon > 0$ можно эффективно указать покрытие множества A последовательностью интервалов, суммарная мера которых не превосходит ε.

Это определение требует некоторых уточнений. Во-первых, мы будем рассматривать не все действительные ε, а только рациональные (иначе непонятно, в какой форме алгоритму можно подать на вход число ε). С другой стороны, надо уточнить, в какой форме алгоритм выдаёт последовательность интервалов. Вот одно из возможных уточнений:

Определение. Множество $A \subset \Omega$ называется *эффективно нулевым* (по данной мере), если существует вычислимая функция x двух аргументов (первый — положительное рациональное число, второй — натуральное число), значениями которой являются двоичные слова, причём

1) $A \subset \Omega_{x(\varepsilon,0)} \cup \Omega_{x(\varepsilon,1)} \cup \Omega_{x(\varepsilon,2)} \ldots;$

2) $p(x(\varepsilon,0)) + p(x(\varepsilon,1)) + p(x(\varepsilon,2)) + \ldots \leqslant \varepsilon$

при любом рациональном $\varepsilon > 0$. При этом мы не требуем, чтобы функция x была всюду определена; если $x(\varepsilon, i)$ не определено, то соответствующий член (в обоих условиях) пропускается.

$\boxed{68}$ Покажите, что определение не изменится, если предполагать, что алгоритм получает на вход рациональное $\varepsilon > 0$, а на выходе перечисляет некоторое множество двоичных слов (печатая их одно за другим с произвольными перерывами), которые образуют покрытие интервалами с *суммарной мерой* (мерой объединения) не больше ε. (Обратите внимание, что суммарная мера может быть меньше суммы мер интервалов, если они пересекаются.)

$\boxed{69}$ Покажите, что определение не изменится, если мы будем рассматривать не все рациональные ε, а только числа вида 2^{-k} при натуральных k. Покажите, что определение не изменится, если заменить знак $\leqslant$ во втором условии на строгое неравенство.

$\boxed{70}$ Покажите, что определение не изменится, если требовать, чтобы для каждого ε функция $i \mapsto x(\varepsilon, i)$ была бы определена на некотором начальном отрезке натурального ряда (возможно, бесконечном).

$\boxed{71}$ Покажите, что определение не изменится, если заменить перечислимое семейство интервалов на разрешимое. [Указание: интервал можно разбить на несколько более мелких, поэтому можно считать, что длины интервалов в перечислении образуют невозрастающую последовательность, а в этом случае множество интервалов разрешимо.]

Приведём несколько примеров эффективно нулевых множеств относительно равномерной меры на Ω.

Рассмотрим одноэлементное множество, единственным элементом которого является последовательность из одних нулей. Оно является эффективно нулевым:

чтобы найти покрытие из интервалов суммарной меры меньше данного $\varepsilon > 0$, возьмём натуральное k, для которого $2^{-k} < \varepsilon$, и покрытие из единственного интервала $\Omega_{00\ldots0}$ (в индексе стоит слово из k нулей).

Формально говоря, $x(\varepsilon, 0) = 0^k$, где 0^k обозначает последовательность из k нулей, а k — наименьшее натуральное число, для которого $2^{-k} < \varepsilon$. Значения $x(\varepsilon, i)$ при $i \neq 0$ не определены.

В этом примере последовательность из одних нулей можно заменить на любую вычислимую последовательность нулей и единиц; нужно только вместо 0^k рассмотреть начальный отрезок этой последовательности длины k.

Но вычислимость существенна, как показывает следующая задача:

$\boxed{72}$ Покажите, что существует последовательность $\omega \in \Omega$, для которой одноэлементное множество $\{\omega\}$ не является эффективно нулевым. [Указание. Рассмотрим все вычислимые функции x, удовлетворяющие второму условию из определения эффективно нулевого множества. Их счётное число. Для каждой из них рассмотрим наибольшее множество A, удовлетворяющее первому условию этого определения (пересечение покрытий по всем ε). Это множество будет (эффективно) нулевым. Объединение счётного числа таких множеств будет нулевым, и потому можно взять ω вне этого объединения.

Замечание. Утверждение этой задачи очевидно следует из теоремы Мартин-Лёфа о наибольшем эффективно нулевом множестве (теорема 28, с. 73), которую мы докажем в этом разделе (и приведённое выше указание к задаче использует по существу тот же метод доказательства). Мы увидим, что множество $\{\omega\}$ является эффективно нулевым тогда и только тогда, когда последовательность ω «не случайна в смысле Мартин-Лёфа».

Легко построить и невычислимую последовательность ω, для которой множество $\{\omega\}$ является эффективно нулевым. Возьмём любую последовательность $\omega = 0?0?0?0\ldots$ (нули чередуются с произвольными цифрами). Покажем, что множество $\{\omega\}$ является эффективно нулевым.

В самом деле, чтобы построить покрытие с суммарной мерой 2^{-n}, можно взять все конечные слова длины $2n$, в которых n нулей чередуется с n произвольными цифрами (как в ω). Таких слов 2^n, а мера каждого интервала 2^{-2n}, поэтому суммарная мера равна 2^{-n}.

На самом деле в этом примере мы доказали, что множество всех последовательностей, у которых на чётных местах нули, является эффективно нулевым. Тем самым и любое его подмножество (в том числе одноэлементное) является эффективно нулевым.

Возвращаясь к определению эффективно нулевого множества, заметим, что требования в его определении можно разделить в следующем смысле. Будем называть вычислимую функцию x корректной, если она удовлетворяет требованию (2) (про сумму мер). Напомним, что требование (1) определения означает, что множество A при любом рациональном $\varepsilon > 0$ лежит в объединении множеств

$$\Omega_{x(\varepsilon,0)} \cup \Omega_{x(\varepsilon,1)} \cup \Omega_{x(\varepsilon,2)} \ldots$$

Таким образом, корректная функция x «обслуживает» все множества, лежащие внутри

$$\bigcap_{\varepsilon>0}(\Omega_{x(\varepsilon,0)}\cup\Omega_{x(\varepsilon,1)}\cup\Omega_{x(\varepsilon,2)}\ldots)=\bigcap_{\varepsilon>0}\bigcup_{i}\Omega_{x(\varepsilon,i)}.$$

Мы видим, что каждой (вычислимой) корректной функции x соответствует некоторое эффективно нулевое множество (задаваемое только что написанной формулой), и что эффективно нулевыми являются все эти множества (для всех корректных функций), а также все их подмножества — и этим исчерпываются все эффективно нулевые множества.

Прежде чем сформулировать теорему Мартин-Лёфа, дадим определение *вычислимой меры* на пространстве Ω.

Действительное число α называют *вычислимым*, если существует алгоритм, вычисляющий приближения к α с любой заданной точностью $\varepsilon>0$. Точнее говоря, α вычислимо, если найдётся вычислимая функция $\varepsilon\mapsto a(\varepsilon)$, определённая на всех положительных рациональных числах и принимающая рациональные значения, для которой

$$|\alpha-a(\varepsilon)|<\varepsilon$$

при любом рациональном $\varepsilon>0$.

$\boxed{73}$ Покажите, что определение не изменится, если дополнительно потребовать, чтобы выдаваемые приближения были «приближениями с недостатком», то есть чтобы $a(\varepsilon)<\alpha$ для всех ε. [Указание: потеряв вдвое в точности, можно любое приближение превратить в приближение с недостатком (или с избытком).]

$\boxed{74}$ Покажите, что сумма, разность, произведение и частное двух вычислимых действительных чисел вычислимы.

$\boxed{75}$ Покажите, что числа e (основание натуральных логарифмов) и π вычислимы.

$\boxed{76}$ Покажите, что элементарные функции (корень, синус, логарифм, экспонента и т. д.) сохраняют вычислимость, то есть что их значения в вычислимых точках вычислимы. (При этом мы предполагаем, естественно, что основание логарифмов и показательной функции вычислимо.)

Мера μ на пространстве Ω называется вычислимой, если меры всех интервалов являются вычислимыми действительными числами, и, более того, алгоритм приближения к $\mu(\Omega_x)$ можно эффективно указать по x. Более формально:

Определение. Мера μ на пространстве Ω называется *вычислимой*, если существует вычислимая функция $\langle x,\varepsilon\rangle\mapsto a(x,\varepsilon)$, определённая для всех слов x и всех положительных рациональных чисел ε, для которой

$$|\mu(\Omega_x)-a(x,\varepsilon)|<\varepsilon$$

для любых x и ε.

Вообще говоря, это определение не предполагает, что мера всего пространства равна единице, но на практике оно нам понадобится только для таких μ (то есть для распределений вероятностей).

Теорема 28 (Мартин-Лёфа). *Пусть μ — вычислимая мера на Ω. Тогда существует наибольшее по включению эффективно нулевое относительно меры μ множество.*

Переформулировка: объединение всех эффективно нулевых (относительно данной вычислимой меры) множеств является эффективно нулевым множеством.

◀ Как мы говорили, каждой корректной функции x соответствует эффективно нулевое множество. Таких множеств счётное число, и любое эффективно нулевое множество содержится в одном из них, поэтому мы немедленно заключаем, что объединение всех эффективно нулевых множеств является нулевым множеством.

Проблема лишь в том, чтобы доказать, что оно является *эффективно* нулевым. Для этого мы будем перечислять все корректные функции, а затем применим эффективный вариант теоремы о счётном объединении нулевых множеств.

По техническим причинам нам будет удобно изменить определение корректной функции. Именно, мы будем называть вычислимую функцию x корректной, если всякая конечная частичная сумма ряда

$$p(x(\varepsilon,0)) + p(x(\varepsilon,1)) + p(x(\varepsilon,2)) + \ldots$$

строго меньше ε. (Напомним, что $p(x) = \mu(\Omega_x)$.) Это требование чуть более сильное (если все частичные суммы ряда строго меньше ε, то его сумма не превосходит ε, но не наоборот). Но легко понять, что определение эффективно нулевого множества не изменится, так как там всегда можно заменить ε, скажем, на $\varepsilon/2$.

В дальнейшем, говоря о корректных функциях (они понадобятся нам только в этом доказательстве), мы имеем в виду это усиленное требование корректности.

Следующая лемма утверждает, что можно эффективно перечислять все корректные функции x.

Лемма. *Существует вычислимая* (*частичная*) *функция трёх аргументов*

$$\langle q,\varepsilon,i\rangle \mapsto X(q,\varepsilon,i)$$

(*q и i — натуральные числа, ε — положительное рациональное число*)*, которая при любом фиксированном q даёт корректную функцию X_q двух аргументов и все корректные* (*вычислимые*) *функции двух аргументов могут быть получены таким образом.*

Доказательство. Расположим все программы для функций двух аргументов (корректных и некорректных) в вычислимую последовательность, и «программой номер q» будем называть q-й член этой последовательности.

Мы определим $X(q,\varepsilon,i)$ как результат применения программы номер q к входам ε, i, если выполнены некоторые условия; если нет, то $X(q,\varepsilon,i)$ не определено. Условия гарантируют, что все X_q корректны и что корректные функции остались нетронутыми.

Чтобы вычислить $X(q,\varepsilon,i)$, мы параллельно запускаем программу номер q на всех парах

$$(\varepsilon,0),(\varepsilon,1),\ldots,$$

сначала делая один шаг первого вычисления, потом два шага первых двух вычислений и т. д. Как только вычисление на паре (ε, i) заканчивается и выдаёт некоторое слово в качестве результата, мы начинаем проверку корректности. Проверка эта состоит в том, что для всех обнаруженных слов z мы начинаем всё с большей и большей точностью вычислять соответствующие значения $p(z)$, пока не убедимся, что сумма этих значений меньше ε. Точнее говоря, проверка заканчивается успешно, когда мы обнаруживаем, что сумма текущих приближений меньше ε на величину, меньшую суммарной погрешности приближений к $p(z)$. Здесь важно, что мера μ вычислима, так что мы можем вычислять приближения к $p(z)$ с любой заданной точностью для любого слова z.

Возможно, эта проверка никогда и не закончится (такое возможно, если сумма мер уже обнаруженных интервалов не меньше ε). (Возможен и отрицательный результат проверки — когда при некоторой точности вычислений обнаруживается, что сумма мер больше ε; мы будем считать, что и в этом случае проверка не заканчивается.)

Теперь $X(q, \varepsilon, i)$ определяется как результат применения программы номер q к входу (ε, i), если этот результат был получен и прошёл проверку корректности в ходе описанного нами процесса.

Если программа номер q на самом деле вычисляет корректную функцию, то все тесты корректности (успешно) завершатся и X_q будет совпадать с этой функцией.

С другой стороны, во всех случаях функция X_q окажется корректной. В самом деле, из последовательности обнаруживаемых значений q-й программы на парах $(\varepsilon, 0), (\varepsilon, 1), \ldots$ будет оставлен некоторый начальный (в порядке обнаружения) отрезок, где суммы мер ещё меньше ε; для всех следующих пар проверка никогда не кончится.

Лемма доказана.

$\boxed{77}$ Объясните, почему нам понадобилось изменить определение корректности: где не проходит доказательство при старом определении? [Указание: если ряд содержит конечное число ненулевых членов, сумма которых в точности равна ε, мы этого никогда не узнаем.]

Продолжаем доказательство теоремы Мартин-Лёфа. Пусть X — функция из леммы. При каждом $q = 0, 1, 2, \ldots$ рассмотрим эффективно нулевое множество Z_q, соответствующее корректной функции X_q. Мы уже знаем, что всякое эффективно нулевое множество содержится в одном из Z_q; остаётся доказать, что объединение $Z_0 \cup Z_1 \cup \ldots$ является эффективно нулевым множеством.

Это делается по той же схеме, что и доказательство теоремы о счётном объединении нулевых множеств. Именно, чтобы найти покрытие суммарной меры не больше ε для $\bigcup_q Z_q$, мы объединим покрытие размера $\varepsilon/2$ для Z_0, покрытие размера $\varepsilon/4$ для Z_1, и так далее.

Формально говоря, мы рассматриваем функцию $x(\varepsilon, i)$, определённую так:

$$x(\varepsilon, [q, k]) = X(q, \varepsilon/2^{q+1}, k),$$

где $[q,k]$ обозначает номер пары q,k при каком-то взаимно однозначном кодировании пар натуральных чисел натуральными числами. (Это кодирование должно быть вычислимым, но в остальном может быть произвольным.) ▶

Теперь мы можем дать определение случайной по Мартин-Лёфу последовательности. Пусть фиксирована некоторая вычислимая мера μ на пространстве Ω.

Определение. Последовательность ω называется *случайной по Мартин-Лёфу* относительно меры μ, если она не содержится в наибольшем эффективно нулевом множестве относительно этой меры.

Переформулировка: последовательность случайна по Мартин-Лёфу, если она не содержится ни в каком эффективно нулевом множестве.

Ещё одна переформулировка: последовательность ω случайна по Мартин-Лёфу, если множество $\{\omega\}$ не является эффективно нулевым.

Отступление о терминологии. Понятие случайной по Мартин-Лёфу последовательности формализует интуитивную идею «типической» (или «типичной») последовательности. Неформально говоря, последовательность типична, если она не обладает никакими особенными свойствами (в целом для последовательностей нехарактерными). Подобным образом понимается слово «типичный» в жаргонном выражении «Онегин — типичный представитель лишних людей» (или, если взять более современный пример, «Вова — типичный лох»). «Особенное» свойство — это свойство, которым обладает лишь пренебрежимо малая часть рассматриваемых объектов. Скажем, свойство последовательности «начинаться с нуля» не является особенным, так как им обладает половина последовательностей. А свойство «каждый второй член равен нулю» является особенным.

Эта неформальная идея получает уточнение в конструкции Мартин-Лёфа: особенным свойством считается принадлежность эффективно нулевому множеству, и таким образом типическими последовательностями становятся последовательности, не принадлежащие никакому эффективно нулевому множеству, то есть случайные по Мартин-Лёфу.

Было бы правильно использовать для таких последовательностей термин «типические», оставив слово «случайные» для интуитивного понятия, допускающего различные уточнения (одним из которых является типичность по Мартин-Лёфу). Однако попытки ввести новую, более правильную, терминологию часто приводят лишь к увеличению путаницы (это замечание авторы самокритично относят и к своим попыткам такого рода). Тем более что путаница и так велика, и в разных текстах слова «случайная последовательность» имеют разный смысл.

В этой книге мы будем говорить «случайная по Мартин-Лёфу» («МЛ-случайная», или «типическая») последовательность, оставив слово «случайная» (без уточнений) для интуитивного представления о случайности.

Следующее утверждение, несмотря на всю его очевидность, полезно осознать (и прочувствовать его парадоксальность):

Теорема 29. *Множество $A \subset \Omega$ является эффективно нулевым тогда и только тогда, когда все его элементы не случайны по Мартин-Лёфу* (*«нетипичны»*).

В частности, множество всех нетипичных последовательностей является (наибольшим) эффективно нулевым множеством, а множество всех типических последовательностей имеет меру 1.

◀ В самом деле, элементы эффективно нулевого множества нетипичны по определению; с другой стороны, если все элементы множества A нетипичны, то A содержится в наибольшем эффективно нулевом множестве и потому является эффективно нулевым. ▶

Парадокс здесь в том, что свойство множества «быть нулевым» скорее означает, что у него «мало элементов», чем ограничивает природу этих элементов. Всякая точка на отрезке (или всякая последовательность в Ω) образует нулевое множество, какой бы она ни была.

С другой стороны, получается, что свойство «быть эффективно нулевым» можно сформулировать в терминах ограничений на элементы множества — запрещается включать в него случайные по Мартин-Лёфу последовательности. А неслучайных (по Мартин-Лёфу) элементов можно включать сколько угодно.

В частности, вспомним, что всякая вычислимая последовательность образует эффективно нулевое одноэлементное множество (по равномерной мере). Отсюда сразу же следует такое утверждение:

Теорема 30. *Множество всех вычислимых последовательностей нулей и единиц является эффективно нулевым подмножеством Ω относительно равномерной меры.*

Любопытно отметить, что по существу это наблюдение было сделано ещё до Мартин-Лёфа, при изучении конструктивного варианта математического анализа («пример Заславского» [192]). Там речь шла о действительных числах, а не о последовательностях нулей и единиц.

В следующем разделе мы продолжим обсуждение свойств последовательностей, МЛ-случайных по равномерной мере. А сейчас мы приведём любопытный критерий МЛ-случайности последовательности, приведённый со ссылкой на Соловея в [24].

Теорема 31. *Последовательность ω не МЛ-случайна по вычислимой мере μ тогда и только тогда, когда существует вычислимая последовательность интервалов с конечной суммой мер, покрывающая эту последовательность бесконечное число раз, то есть вычислимая последовательность слов $x_0, x_1, x_2, \ldots$, для которой*

$$\sum_i \mu(\Omega_{x_i}) < \infty$$

и $\omega \in \Omega_{x_i}$ при бесконечно многих i.

◀ Если последовательность ω не МЛ-случайна, то множество $\{\omega\}$ имеет покрытия с суммой мер не больше ε для любого положительного рационального ε. Соединим такие покрытия для $\varepsilon = 1, 1/2, 1/4, 1/8, \ldots$ Соединённое покрытие имеет сумму мер не больше 2 и покрывает ω бесконечно много раз (на каждом уровне хотя бы по разу).

Напротив, пусть есть некоторое покрытие точки ω интервалами, соответствующими словам $x_0, x_1, x_2, \ldots$, и сумма мер этих интервалов не превосходит какой-то

границы c, которую удобно считать рациональным числом. Чтобы найти покрытие точки ω с суммой не больше ε, рассмотрим множество тех точек из Ω, которые покрыты не менее N раз, где целое положительное N выбрано так, чтобы c/N было меньше ε. Легко понять, что это множество (N-кратно покрытые точки) представимо в виде объединения непересекающихся интервалов, и эти интервалы можно перечислять алгоритмически, глядя на последовательность $x_0, x_1, x_2, \ldots$ и зная N. Тем самым множество $\{\omega\}$ является эффективно нулевым, а последовательность ω — неслучайной по Мартин-Лёфу. ▶

Замечание. Эта теорема является «конструктивизацией» леммы Бореля – Кантелли (если сумма мер множеств $A_0, A_1, \ldots$ конечна, то множество точек, принадлежащих бесконечно многим из A_i, является нулевым), и приведённое нами рассуждение является конструктивизацией доказательства леммы Бореля – Кантелли. Однако тут нужна осторожность: для конструктивизации годится не всякое доказательство. Обычное доказательство этой леммы (ряд сходится, значит, хвосты его стремятся к нулю и покрывают интересующее нас множество) не годится, так как мы не умеем по ε эффективно находить хвост, меньший ε.

При определении случайности по мере мы предполагали её вычислимость, поскольку эта вычислимость использовалась в доказательстве существования максимального эффективно нулевого множества. В принципе можно обойтись и без максимального множества, сказав, что случайными (по произвольной мере μ на Ω) будут те последовательности, которые не лежат ни в каком нулевом множестве. Для этого понятия случайности Б. Кьёс-Хансен предложил термин *Hippocratic randomness* [58] (а П. Гач — более нейтральный термин «слепая (безоракульная) случайность», *blind randomness*). Это не единственный возможный вариант определения случайности относительно невычислимых мер: отметим (без подробностей), что можно использовать также *равномерные тесты случайности* (uniform randomness tests, подход Левина – Гача, см. обзор [8]).

78 Докажите, что для перечислимой сверху меры существует максимальное эффективно нулевое множество. (При этом мы не предполагаем, что мера всего пространства равна 1; в последнем случае из перечислимости сверху следует вычислимость.)

79 Приведите пример (невычислимой) меры, для которой максимального эффективно нулевого множества не существует.

[Указание. Можно построить меру, относительно которой любая вычислимая последовательность будет образовывать эффективно нулевое множество (и даже некоторый её начальный отрезок будет иметь меру нуль), одновременно всякий алгоритм, претендующий на порождение эффективно нулевых множеств, будет либо давать интервал недопустимо большой меры, либо не будет покрывать некоторую вычислимую последовательность. Мера строится диагональным процессом, в котором мы по очереди рассматриваем все вычислимые последовательности и все кандидаты в порождающие алгоритмы.]

80 Приведите пример невычислимой меры, для которой существует наибольшее эффективно нулевое множество. [Указание: можно рассмотреть невычислимую

меру, настолько близкую к равномерной, чтобы меры всех множеств отличались не больше чем в константу раз.]

3.4. Свойства случайных по Мартин-Лёфу последовательностей

Усиленный закон больших чисел также даёт пример эффективно нулевого множества (относительно равномерной меры).

Теорема 32. *Множество последовательностей нулей и единиц, для которых $1/2$ не является пределом последовательности частот, является эффективно нулевым по равномерной мере.*

◀ Достаточно показать, что при любом рациональном $\varepsilon > 0$ множество тех последовательностей, у которых частота бесконечное число раз становится больше $1/2 + \varepsilon$ (или меньше $1/2 - \varepsilon$), является эффективно нулевым. Для этого заметим, что оценка меры, приведённая при доказательстве закона больших чисел в предыдущем разделе (теорема 27, с. 66), была эффективной: покрытие состояло из продолжений всех достаточно длинных слов с большим отклонением частоты, а его мера эффективно оценивалась сверху остаточным членом в сумме бесконечно убывающей геометрической прогрессии. ▶

Переформулируем доказанное утверждение в терминах МЛ-случайных последовательностей:

Теорема 33. *Пусть $\omega = \omega_0\omega_1\ldots$ — случайная в смысле Мартин-Лёфа последовательность относительно равномерной меры. Тогда*

$$\lim_{n\to\infty} \frac{\omega_0 + \omega_1 + \ldots + \omega_{n-1}}{n} = \frac{1}{2}.$$

Аналогичное утверждение справедливо и для неравномерных бернуллиевых мер. Пусть даны положительные числа p, q, для которых $p + q = 1$, причём p и q вычислимы. Рассмотрим бернуллиеву меру с параметрами q, p (соответствующую последовательности независимых испытаний с вероятностью успеха p, см. с. 66). Легко проверить, что эта мера вычислима (поскольку вычислимы p и q).

Теорема 34. *Любая МЛ-случайная относительно бернуллиевой меры с вычислимыми параметрами q, p последовательность имеет предел частот, равный p.*

◀ В самом деле, оценка вероятностей последовательностей с бо́льшими уклонениями частот от p (производимая с помощью сравнения меры с другой мерой, в которой p сдвинуто, см. задачу 67, с. 69), даёт явную оценку и покрытие, поэтому получается эффективно нулевое множество. ▶

Ещё несколько свойств типичности (МЛ-случайности) по равномерной мере:

Теорема 35. *Пусть ω — типическая по равномерной мере последовательность. Тогда последовательность полученная из ω удалением, добавлением или изменением конечного числа членов, также является типической.*

◀ Достаточно доказать, что дописывание в начало последовательности нуля или единицы, а также удаление первого её члена не сказывается на типичности.

В самом деле, пусть последовательность ω нетипична, то есть образует эффективно нулевое множество: по всякому ε можно построить покрытие интервалами с суммой мер не больше ε. Припишем к словам, задающим эти интервалы, нуль в начале. Получится покрытие для последовательности 0ω вдвое меньшей меры. Это рассуждение показывает, что если ω не типична, то и 0ω не типична. (Аналогично для 1ω.)

С другой стороны, если у всех слов, задающих интервалы покрытия, удалить первый бит, то получится семейство интервалов вдвое большей меры, покрывающее последовательность ω', получающуюся удалением из ω первого бита. Значит, ω' не типична. ▶

81 Покажите, что замена всех нулей на единицы и наоборот сохраняет типичность последовательности (по равномерной мере).

Типичность также сохраняется при переходе к вычислимой подпоследовательности, как показывает следующая задача.

82 Пусть $n_0, n_1, n_2, \ldots$ — вычислимая последовательность различных натуральных чисел ($n_i \neq n_j$ при $i \neq j$). Покажите, что если последовательность $\omega = \omega_0\omega_1\omega_2\ldots$ типична, то и её подпоследовательность

$$\omega|n = \omega_{n_0}\omega_{n_1}\omega_{n_2}\ldots$$

типична. [Указание. Из каждого интервала Ω_x в покрытии для $\omega|n$ получается конечное семейство интервалов, у которых члены с номерами $n_0, n_1, \ldots, n_{i-1}$ совпадают с битами слова x (здесь i — длина слова x), а остальные биты любые. Общая мера этих интервалов равна мере интервала Ω_x.]

Мы продолжим эту тему (о выборе подпоследовательности из случайной последовательности) в главе о частотном подходе к определению случайности, предложенном фон Мизесом (глава 9, с. 292).

83 Пусть последовательность ω типична (МЛ-случайна) по равномерной мере. Разделим её на блоки по две цифры и заменим блоки 00 на нули, а блоки 01, 10 и 11 — на единицы. Докажите, что полученная последовательность типична по бернуллиевой мере с параметрами $1/4, 3/4$. [Указание. Описанное преобразование задаёт отображение $F\colon \Omega \to \Omega$. Прообраз любого открытого множества U при этом открыт, и равномерная мера этого прообраза равна $(1/4, 3/4)$-мере множества U.]

84 (Продолжение) Докажите, что любая типическая по $(1/4, 3/4)$-мере последовательность может быть получена описанным способом из последовательности, типической по равномерной мере.

[Указание. Для любого открытого множества $B \subset \Omega$ рассмотрим множество B' всех тех последовательностей ω, у которых $F^{-1}(\{\omega\}) \subset B$ (те последовательности, которые не имеют прообраза вне B, или дополнение к образу дополнения B). Образ компактного множества компактен, поэтому B' открыто. Проверьте, что

если B есть объединение перечислимого семейства интервалов, то и B' имеет такой вид, и бернуллиева мера B' не превосходит равномерной меры B. См. также доказательство более общего утверждения (теорема 123, с. 206).]

Интересно понять, какова может быть «сложность» МЛ-случайной (по равномерной мере) последовательности с точки зрения теории вычислимых функций. Мы уже видели, что такая последовательность не может быть вычислимой. Она также не может быть характеристической функцией перечислимого множества.

Теорема 36. *Пусть A — перечислимое множество натуральных чисел. Тогда его характеристическая последовательность $a_0a_1a_2\ldots$, где $a_i = 0$ при $i \notin A$ и $a_i = 1$ при $i \in A$, не является МЛ-случайной по равномерной мере.*

◀ Пусть k — произвольное натуральное число. Будем перечислять множество A, следя за первыми k членами его характеристической последовательности. По мере перечисления среди этих членов появляются всё новые и новые единицы и мы получаем всё новые и новые варианты k-битового начала характеристической последовательности (какой-то из них окажется окончательным, но у нас нет способа узнать об этом). Но так или иначе этих вариантов не больше $k+1$ штук (число единиц может меняться от 0 до k). И если мы все эти слова включим в покрытие, то общая мера интервалов покрытия будет не больше $(k+1)/2^k$ и таким образом может быть сделана сколь угодно малой. (Заметим, что в определении эффективно нулевого множества как раз требуется перечислять элементы покрытия, не указывая момента окончания такого перечисления.) ▶

Возникает вопрос, можно ли вообще в каком-то смысле указать конкретную случайную по Мартин-Лёфу последовательность. Следующий результат (приводимый для читателей, знакомых с началами теории вычислимых функций, о которых можно прочесть, например, в [175]), показывает, что случайную последовательность можно найти в классе $\Sigma_2 \cap \Pi_2$ арифметической иерархии (другое описание последовательностей этого класса — вычислимые относительно $\mathbf{0}'$ последовательности).

Теорема 37. *Существует вычислимая относительно $\mathbf{0}'$ последовательность, МЛ-случайная по равномерной мере.*

◀ Достаточно показать, что для любого перечислимого множества слов x_0, x_1, $x_2, \ldots$, у которого $\sum 2^{-l(x_i)} < 1/2$, существует вычислимая относительно $\mathbf{0}'$ последовательность, для которой ни одно из слов x_i не является началом. (Максимальное эффективно нулевое множество имеет покрытие с суммой мер меньше $1/2$, и всякая последовательность вне этого покрытия будет МЛ-случайной.)

Посмотрим, как интервалы Ω_{x_i} распределяются между двумя половинами множества Ω (какие из них начинаются на нуль, а какие на единицу).

В сумме мера тех и других не превосходит $1/2$, значит, мера интервалов, приходящихся на одну из половин, не больше $1/4$. Правда, наблюдая за последовательностью x_i, мы не можем с уверенностью выбрать нужную половину (вдруг потом объявится короткое слово, которое попадёт в эту половину).

Но если у нас есть оракул для $\mathbf{0}'$, то с его помощью такой выбор можно сделать (поскольку превышение меры над $1/4$ есть перечислимое событие). Соответствующую половину разделим на две четверти и посмотрим, в какой из них мера

интервалов покрытия будет не больше $1/8$. И так далее. Тем самым мы получим вычислимую (относительно $\mathbf{0}'$) последовательность нулей и единиц с таким свойством: над любым её началом интервалы покрытия заполняют не более половины всех продолжений. В частности, никакое начало этой последовательности не может войти в покрытие, что нам и требовалось. ►

Похожее, но более геометрически наглядное рассуждение получится, если говорить о точках отрезка $[0,1]$ вместо двоичных последовательностей. Такую точку называют *МЛ-случайной относительно меры Лебега на* $[0,1]$, если её двоичная запись МЛ-случайна относительно равномерной меры на Ω. (Наличие двух двоичных записей тут не важно, так как в этой ситуации обе они не случайны.) Случайность в этом смысле можно эквивалентно определить и геометрически. Множество $X \subset [0,1]$ объявим *эффективно нулевым подмножеством* $[0,1]$, если существует алгоритм, который по рациональному $\varepsilon > 0$ перечисляет покрытие этого множества рациональными интервалами, сумма длин которых меньше ε. Случайными будут точки, не покрытые никакими эффективно нулевыми множествами. Всё это является простой переформулировкой соответствующих понятий для канторовского пространства (поскольку оно отличается от $[0,1]$ лишь наличием двух копий для двоично-рациональных точек, что не сказывается на нулевых множествах). В частности, верно такое свойство:

$\boxed{\mathbf{85}}$ Покажите, что существует максимальное эффективно нулевое множество, и оно состоит из точек, двоичные записи которых не случайны в канторовском пространстве.

Теперь можно указать «конкретную» случайную точку. Возьмём перечислимое семейство интервалов суммарной длины меньше 1, покрывающее все неслучайные точки. Они в объединении дают открытое множество, не совпадающее со всем отрезком. Дополнение этого открытого множества будет замкнутым множеством. Наименьшая точка этого множества будет (по построению) случайной.

$\boxed{\mathbf{86}}$ Докажите, что наименьшая точка отрезка $[0,1]$, не покрытая некоторым перечислимым семейством интервалов, является перечислимой снизу, то есть пределом возрастающей вычислимой последовательности рациональных чисел (и потому вычислима с оракулом $\mathbf{0}'$).

Ещё одна конструкция последовательности с таким свойством (задающей случайное перечислимое снизу число) приведена в разделе 5.7 на с. 179.

Приведённое нами доказательство теоремы 37 представляет собой релятивизованный вариант такой задачи:

$\boxed{\mathbf{87}}$ Пусть имеется вычислимая последовательность слов $x_0, x_1, x_2, \ldots$, причём сумма ряда

$$\sum_i 2^{-l(x_i)}$$

меньше 1 и является вычислимым действительным числом. Тогда существует вычислимая последовательность нулей и единиц, для которой ни одно из слов x_i не является началом.

[Указание. Пусть сумма этого ряда меньше рационального числа S, меньшего единицы. Тогда можно по индукции построить вычислимую последовательность $\omega_0\omega_1\omega_2\ldots$ с таким свойством: доля множества $U = \cup\Omega_{x_i}$ среди продолжений слова $\omega_0\ldots\omega_k$ составляет менее S.]

Последняя задача связана с модификацией определения случайности по Мартин-Лёфу, предложенной Шнорром [143]. А именно, в определении эффективно нулевого множества можно дополнительно потребовать, чтобы для каждого $\varepsilon > 0$ ряд из мер покрывающих интервалов (сумма которого не должна превосходить ε) вычислимо сходился. Это означает, что для каждого $\varepsilon > 0$ и $\delta > 0$ можно алгоритмически указать конечное число членов ряда $\sum_i p(x(\varepsilon, i))$, которые все определены и приближают сумму всего ряда с ошибкой не более δ. (Поскольку члены ряда неотрицательны, это эквивалентно тому, что его сумма является вычислимым действительным числом, которое можно указать по ε.) Эффективно нулевые множества, для которых выполнено это дополнительное условие, будем называть *эффективно нулевыми по Шнорру*. (Шнорр называет их *total rekursive Nullmenge*, см. определение 8.1 в [143], в отличие от эффективно нулевых в смысле Мартин-Лёфа, которые называются у Шнорра *rekursive Nullmenge*, см. там же определение 4.1.)

$\boxed{\mathbf{88}}$ Покажите, что если в определении эффективно нулевого множества требовать, чтобы суммарная мера покрывающих интервалов *в точности равнялась* ε, то получится определение, эквивалентное определению Шнорра. (Можно также вместо суммарной меры покрывающих интервалов говорить о мере их объединения.)

Задача 87 показывает, что для любого эффективно нулевого по Шнорру множества существует вычислимая последовательность, ему не принадлежащая. (Для простоты мы ограничиваемся случаем равномерной меры, хотя это и не существенно.) Поскольку вычислимая последовательность образует одноэлементное эффективно нулевое по Шнорру множество, отсюда следует, что среди эффективно нулевых по Шнорру множеств нет наибольшего (другими словами, объединение нулевых по Шнорру множеств не является нулевым по Шнорру). Тем не менее можно назвать последовательность, не входящую ни в одно нулевое по Шнорру множество, *случайной по Шнорру* (или *типической по Шнорру*).

Полученный класс последовательностей оказывается более широким: как показывает следующая задача (и результаты главы 5), существуют случайные по Шнорру последовательности, не являющиеся случайными по Мартин-Лёфу.

$\boxed{\mathbf{89}}$ Покажите, что существует случайная по Шнорру последовательность $\omega = \omega_0\omega_1\omega_2\ldots$, у которой сложность начальных отрезков растёт логарифмически, то есть $KS(\omega_0\ldots\omega_{n-1}) = O(\log n)$.

[Указание. В задаче 87 мы видели, как строить вычислимую последовательность, не принадлежащую заданному эффективно нулевому по Шнорру множеству. Если в какой-то момент мы захотим ввести в это построение другое эффективно нулевое по Шнорру множество, это вполне возможно, надо только выбрать достаточно малое покрытие (исходя из имеющегося на данный момент запаса). Так можно вводить эффективно нулевые множества одно за другим. Это не даст вычислимой случайной по Шнорру последовательности (которой быть не может), поскольку нам нужна дополнительная информация о том, какие алгоритмы задают эффективно нулевые

по Шнорру множества, а какие нет. Но если вводить новые алгоритмы постепенно, когда построенный кусок последовательности уже длинен, то эта дополнительная информация будет логарифмической по сравнению с длиной последовательности.]

Мы вернёмся к определению случайности по Шнорру в разделе 9.8, где будет дана его переформулировка в терминах вычислимых мартингалов.

90 Покажите, что последовательность ω не является случайной по Шнорру тогда и только тогда, когда существует вычислимая последовательность слов $x_0, x_1, \ldots$, для которой ряд $\sum_i p(x_i)$ вычислимо сходится и среди x_i имеется бесконечное много начал последовательности ω.). [Указание. Это утверждение является аналогом критерия Соловея случайности по Мартин-Лёфу, теоремы 31, и может быть доказано аналогичным образом. В данном случае, впрочем, работает и стандартное доказательство леммы Бореля – Кантелли.]

Ещё один вариант случайности возникает, если в определении эффективно нулевого множества рассматривать конечные покрытия (заданные списком) вместо бесконечных: по рациональному $\varepsilon > 0$ надо указывать список интервалов общей меры меньше ε, покрывающий множество. (Это соответствует продолжению меры по Жордану в курсах математического анализа.) При этом получается меньший класс нулевых множеств (например, нельзя покрыть все рациональные числа).

91 Покажите, что множество является нулевым в этом смысле тогда и только тогда, когда оно содержится в дополнении эффективно открытого множества полной меры.

Последовательности, не содержащиеся в таких множествах (= входящие в любое эффективно открытое множество полной меры), называют *случайными по Курцу* (Kurtz random). Поскольку нулевых множеств в этом определении меньше, чем у Мартин-Лёфа, и даже меньше, чем у Шнорра, то случайных последовательностей будет больше, и даже строго больше, как показывает следующая задача.

92 Покажите, что случайная по Шнорру (относительно равномерной меры) последовательность всегда удовлетворяет усиленному закону больших чисел. Покажите, что случайная по Курцу последовательность может не удовлетворять усиленному закону больших чисел. [Указание. Построенное при доказательстве усиленного закона больших чисел множество имеет вычислимую меру, поскольку ряд экспоненциально сходится. Во втором случае можно рассмотреть генерические последовательности, см. раздел 5.9, с. 203.]

3.5. Дефект случайности

Согласно определению Мартин-Лёфа, для эффективно нулевого множества A по любому $\varepsilon > 0$ можно эффективно указать покрытие A семейством интервалов, суммарная мера которых не превосходит ε. Объединение интервалов этого семейства будет открытым множеством меры не более ε.

Вообще объединения вычислимых последовательностей интервалов называют *эффективно открытыми* множествами. Как и в определении эффективно нулевого множества (с. 70), мы разрешаем не всюду определённые последовательности,

так что пустое множество тоже эффективно открыто. Переформулировка: эффективно открытое множество есть объединение перечислимого семейства интервалов (появляющихся на выходе перечисляющего их алгоритма).

Определение эффективно нулевого множества можно теперь переформулировать:

Теорема 38. *Множество A является эффективно нулевым относительно меры μ тогда и только тогда, когда $A \subset \bigcap_n U_n$, где U_n — равномерно эффективно открытые множества и $\mu(U_n) \leqslant 2^{-n}$. При этом можно дополнительно предполагать, что $U_1 \supset U_2 \supset \ldots$.*

Говоря о равномерности, мы имеем в виду существование алгоритма, который на вход получает n и перечисляет интервалы, объединение которых составляет U_n.

◀ Это определение отличается от данного нами раньше в нескольких отношениях. Во-первых, в качестве ε выбираются числа вида 2^{-n} (которыми, очевидно, можно ограничиться).

Во-вторых, мы говорим здесь о мере эффективно открытых множеств, а о не сумме мер интервалов, объединениями которых они являются. Но это не важно, так как интервалы можно считать непересекающимися: добавляя новый интервал канторовского пространства, мы вычитаем из него все старые, разбивая остаток в сумму непересекающихся интервалов. Тогда сумма мер не отличается от меры суммы.

Наконец, мы требуем вложенности множеств. Этого легко добиться, перейдя к последовательности $U_1, U_1 \cap U_2, U_1 \cap U_2 \cap U_3, \ldots$ — надо только убедиться, что пересечение конечного числа эффективно открытых множества эффективно открыто (и соответствующий алгоритм может быть эффективно указан). В самом деле, пусть даны два алгоритма, перечисляющих интервалы для U_1 и U_2. Тогда в каждый момент текущие приближения для U_1 и U_2 можно пересечь, это будет конечное объединение интервалов, и все эти интервалы можно добавить в перечисление для $U_1 \cap U_2$ (их даже можно сделать непересекающимися, как описано выше).▶

Именно так и выглядело оригинальное определение Мартин-Лёфа [97]; семейства множеств U_n с указанными свойствами он называл *тестами*. Имея такой тест $U_1, U_2, \ldots$, можно определить функцию «дефекта случайности», ему соответствующую: дефект последовательности ω определяется как наибольшее i, при котором $\omega \in U_i$. Дефект бесконечен, когда ω принадлежит всем U_i. При таком определении тест не только отбраковывает последовательности из $\bigcap_n U_n$, но и для остальных последовательностей указывает, насколько они близки к опасной зоне (чем ближе, тем больше дефект).

Наибольшему эффективному множеству соответствует тест, который отбраковывает максимальное число последовательностей. Мартин-Лёф заметил, что можно построить универсальный тест даже и в более сильном смысле:

Теорема 39. *Существует тест случайности, для которого функция дефекта является максимальной с точностью до аддитивной константы.*

◀ Описанное выше построение покрытий для максимального нулевого множества даёт такой тест: мы можем соединить тесты

$$
\begin{array}{l}
U^1\colon\ U_1^1 \supset U_2^1 \supset U_3^1 \supset \ldots \\
U^2\colon\ U_1^2 \supset U_2^2 \supset U_3^2 \supset \ldots \\
U^3\colon\ U_1^3 \supset U_2^3 \supset U_3^3 \supset \ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{array}
$$

в единый тест

$$(U_2^1 \cup U_3^2 \cup \ldots \cup U_{k+1}^k \cup \ldots) \supset (U_3^1 \cup U_4^2 \cup \ldots \cup U_{k+2}^k \cup \ldots) \supset \ldots$$

В этом тесте меры оцениваются как $1/4 + 1/8 + \ldots \leqslant 1/2$ для первого множества, $1/8 + 1/16 + \ldots \leqslant 1/4$ для второго и т. д.; функция дефекта для результирующего теста не меньше $d^i - i$, где d^i — функция дефекта для i-го теста. ▶

Функция дефекта может рассматриваться как компактный способ представления теста: мы рассматриваем функции на Ω, значениями которых являются натуральные числа и бесконечность, при этом множество U_n тех ω, на которых дефект превосходит n, должно быть эффективно открытым (равномерно по n).

Не обязательно ограничиваться функциями с целыми значениями. Будем говорить, что функция u, определённая на Ω и принимающая неотрицательные значения (возможно, бесконечные), является *перечислимой снизу*, если для любого рационального r множество тех последовательностей ω, где $u(\omega) > r$, эффективно открыто равномерно по r. Следующая теорема позволяет дать эквивалентный вариант определения перечислимости снизу. Будем называть *базисной* функцией на канторовском пространстве функцию с рациональными значениями, зависящую только от некоторого начального отрезка последовательности-аргумента. Такая функция задаётся длиной этого отрезка, некоторым числом m, и таблицей из 2^m рациональных значений, соответствующих 2^m вариантам этого отрезка, поэтому базисные функции можно считать конструктивными объектами и говорить о вычислимых последовательностях таких функций.

Теорема 40. *Следующие свойства функции v, определённой на канторовском пространстве и принимающей неотрицательные (возможно, бесконечные) значения, равносильны*:

(а) *v перечислима снизу*;

(б) *v является точной верхней гранью вычислимой последовательности базисных функций*;

(в) *v является поточечным пределом вычислимой неубывающей последовательности базисных функций*;

(г) *v является поточечной суммой вычислимого ряда, составленного из неотрицательных базисных функций.*

◀ Два последних свойства эквивалентны, так как разность и сумма двух базисных функций является базисной функцией.

Чтобы перейти от точной верхней грани к пределу возрастающей последовательности, достаточно заметить, что максимум конечного числа базисных функций является базисной функцией.

Остаётся доказать, что первое свойство равносильно остальным (например, второму). Пусть $v = \sup_i v_i$. Заметим, что $\sup_i v_i(\omega) > r$ тогда и только тогда, когда $v_i(\omega) > r$ при некотором i, так что множество $\{\omega \mid v(\omega) > r\}$ оказывается (равномерно) эффективно открытым.

Наконец, если при каждом r можно алгоритмически перечислять интервалы, на которых $u(\omega) > r$, то функция u есть точная верхняя грань базисных функций, которые равны r внутри таких интервалов и нулю в остальных местах. ►

Отбросив ограничение целочисленными значениями и переходя к экспоненциальной шкале, можно дать такое определение: *ограниченным по вероятности тестом случайности относительно вычислимой меры* μ *на* Ω называется перечислимая снизу функция u, для которой

$$\mu(\{\omega \mid u(\omega) > c\}) < 1/c$$

для любого положительного рационального c. Это определение соответствует интуиции: последовательностей, в которых тест находит много закономерностей (значение теста больше c), не должно быть много (вероятность случайно получить такое большое значение не превосходит $1/c$).

Теорема 41. *Среди ограниченных по вероятности тестов относительно вычислимой меры* μ *существует максимальный (с точностью до постоянного множителя); двоичный логарифм такого теста совпадает с универсальным тестом из теоремы 39 с точностью до* $O(1)$.

◄ Произвольный тест можно заменить на тест со значениями вида 2^n, заменив $u(\omega)$ на максимальную целую степень двойки, меньшую $u(\omega)$; это изменит тест не более чем в константу раз, сохранив перечислимость снизу. Тест с такими значениям соответствует набору эффективно открытых множеств, и условие на тест соответствует условиям на эти множества. Остаётся сослаться на теорему 39. ►

Есть и другой класс тестов, которые называют *ограниченными в среднем* тестами. В этом варианте ограничение на перечислимую снизу неотрицательную функцию u более сильное: $\int u(\omega)\,d\mu(\omega) \leqslant 1$.

Теорема 42. *Среди ограниченных в среднем тестов относительно вычислимой меры* μ *существует максимальный (с точностью до постоянного множителя).*

◄ Семейство всех перечислимых снизу функций можно эффективно перечислять (например, исходя из их представления в виде точной верхней грани перечислимого множества базисных функций). Далее, их можно корректировать, добиваясь, чтобы интеграл был не больше (скажем) 2 и оставляя тесты без изменений. (Здесь важно, что мера μ вычислима.) Получится равномерно перечислимая снизу последовательность $u_1, u_2, \ldots$, составленная из тестов и почти тестов (с точностью до множителя 2). Из неё делаем универсальный тест, сложив u_i с вычислимыми коэффициентами, образующими ряд с суммой $1/2$ или меньше. ►

Логарифмы максимальных тестов (ограниченных в среднем и по вероятности) будем называть *дефектами случайности* относительно вычислимой меры μ (с дополнительным уточнением «ограниченный в среднем» или «ограниченный по вероятности») и обозначать соответственно $\mathbf{d}^E$ и $\mathbf{d}^P$, от слов “expectation” и “probability”. (Вычислимую меру μ мы предполагаем фиксированной и в обозначение не включаем.)

Теорема 43.

$$\mathbf{d}^P(\omega) - 2\log \mathbf{d}^P(\omega) \leqslant \mathbf{d}^E(\omega) \leqslant \mathbf{d}^P(\omega).$$

Оба неравенства понимаются с точностью до $O(1)$-слагаемого (и входящие в них величины определены с такой точностью).

◀ Второе неравенство следует из того, что всякий ограниченный в среднем тест ограничен и по вероятности. Чтобы доказать первое неравенство, убедимся в том, что 2 в степени $\mathbf{d}^P - 2\log \mathbf{d}^P$ имеет ограниченный интеграл (перечислимость снизу следует из того, что функция $x - 2\log x$ монотонно возрастает; это не так для малых значений, но они всё равно не важны, так как дают только конечный вклад в интеграл, и там функцию можно заменить на возрастающую). В самом деле, множество тех последовательностей, где $\mathbf{d}^P(\omega)$ находится между n и $n+1$, имеет меру не больше $1/2^n$, поскольку тест $2^{\mathbf{d}^E}$ там превосходит 2^n, а подынтегральное выражение на этом множестве равно $O(2^n/n^2)$; остаётся заметить, что ряд $1/n^2$ сходится. ▶

$\boxed{\mathbf{93}}$ Покажите, что константу 2 в теореме 43 можно заменить на любое число, большее 1.

Из доказанной теоремы, в частности следует, что любой вид тестов можно использоваться для определения случайности по Мартин-Лёфу (для ограниченных по вероятности тестов мы это уже обсуждали): случайны те и только те последовательности, у которых дефект конечен.

Разница в определениях двух видов тестов напоминает различие между простой колмогоровской сложностью (которую мы рассматривали раньше) и префиксной сложностью (см. следующую главу).

Тесты случайности были определены и изучены ещё в 1960-е и 1970-е годы (ограниченные по вероятности тесты по существу были определены уже в самой первой работе Мартин-Лёфа; ограниченные в среднем тесты изучались в работах Л.А. Левина и П. Гача), но потом они (до недавнего времени) использовались мало. Более подробно об их применениях можно прочесть в [8].

4. Априорная вероятность и префиксная сложность

4.1. Вероятностные машины и полумеры на $\mathbb{N}$

Рассмотрим алгоритм (машину, программу) с датчиком случайных битов. Такой алгоритм содержит операции

$$b := \texttt{random}$$

при которых переменной b присваивается нуль или единица с равными вероятностями. (Встретив такую команду, мы прерываем выполнение алгоритма, бросаем честную монету и результат бросания записываем в переменную b.) Такие алгоритмы называют *вероятностными*.

Результат работы вероятностного алгоритма зависит не только от его входа, но и от результатов бросаний монеты (случайных битов). Таким образом, при данном входе выход (результат работы) алгоритма является случайной величиной.

Более формально вероятность того или иного выхода вероятностного алгоритма A определяется так. На пространстве Ω всех бесконечных последовательностей нулей и единиц рассматривается равномерная бернуллиева мера. При этом мера множества Ω_u всех бесконечных продолжений данного конечного слова u равна $2^{-l(u)}$.

Для данного входа x и последовательности $\omega \in \Omega$ через $A(x, \omega)$ обозначим результат работы A на входе x, если в качестве случайных битов берутся биты из последовательности ω (каждый вызов `random` берёт новый бит). Значение $A(x, \omega)$ — некоторый конструктивный объект (число, слово); оно может быть не определено при некоторых x и ω. Для каждого возможного выхода y алгоритма рассмотрим множество $\{\omega \mid A(x, \omega) = y\}$. Это множество измеримо (и даже открыто в смысле естественной топологии в Ω); оно является объединением интервалов Ω_z для всех слов z, которые являются результатами бросаний монеты к моменту появления выхода y. Мера этого множества и называется *вероятностью выхода y при входе x*.

В этом разделе мы будем рассматривать вероятностные машины без входа, выходами которых являются натуральные числа. Пример: машина бросает монету, пока не появится единица, и выдаёт на выход число нулей перед первой единицей. Для этого алгоритма распределение вероятностей на возможных выходах такое: вероятность p_i появления числа i равна $2^{-(i+1)}$. В самом деле, p_0 есть вероятность того, что первое бросание монеты дало единицу, p_1 — вероятность того, что первое бросание дало нуль, а второе — единицу и так далее.

В данном конкретном случае сумма ряда $\sum p_i$ равна единице: вероятность того, что машина так ничего и не напечатает (так случится, если всё время будут выпадать нули) равна нулю. Но для других алгоритмов эта сумма может быть и меньше единицы.

Итак, каждой вероятностной машине без входа, выдающей натуральные числа на выходе, соответствует последовательность $p_0, p_1, \ldots$, где p_i есть вероятность появления числа i. Какие последовательности действительных чисел $p_0, p_1, \ldots$ могут получиться таким образом? Одно условие очевидно: $\sum p_i \leqslant 1$. Но, конечно, это условие не является достаточным: вероятностных машин счётное число, а возможных последовательностей $p_0, p_1, \ldots$ — континуум.

Начнём с более простого вопроса: какова может быть вероятность остановки вероятностной машины (без входа)? Чтобы сформулировать ответ, введём понятие перечислимого снизу действительного числа.

Говорят, что действительное число α *перечислимо снизу*, если оно есть предел вычислимой неубывающей последовательности рациональных чисел.

$\boxed{\mathbf{94}}$ Докажите, что если действительное число α вычислимо (существует алгоритм, который по любому рациональному $\varepsilon > 0$ указывает приближение к α с погрешностью не более ε), то α перечислимо снизу. [Указание. Последовательность приближений снизу можно переделать в возрастающую.]

$\boxed{\mathbf{95}}$ Докажите, что действительное число α вычислимо тогда и только тогда, когда оба числа α и $-\alpha$ перечислимы снизу.

Вот эквивалентное определение перечислимости снизу: число α перечислимо снизу, если множество всех рациональных чисел, меньших α, перечислимо.

В самом деле, если α есть предел неубывающей вычислимой последовательности $a_0 \leqslant a_1 \leqslant a_2 \leqslant \ldots$, то перечисляя все числа, меньшие какого-либо из a_i, мы перечислим все числа, меньшие α. Напротив, умея перечислять все рациональные числа, меньшие α, мы можем составить из них последовательность. Затем эту последовательность надо сделать неубывающей, выбрасывая из неё члены, меньшие уже встречавшихся.

Понятие перечислимого снизу действительного числа даёт ответ на поставленный нами вопрос:

Теорема 44. (а) *Пусть M — произвольная вероятностная машина* (*без входа*)*. Тогда вероятность её остановки есть перечислимое снизу действительное число.*

(б) *Всякое перечислимое снизу действительное число есть вероятность остановки некоторой вероятностной машины.*

◄ (а) Вероятность p_n остановки машины в течение n (или менее) шагов есть некоторое рациональное число: за n шагов можно сделать не более n бросаний монеты, поэтому каждый исход имеет вероятность, кратную $1/2^n$, и вероятность остановки также кратна $1/2^n$.

Число p_n можно найти, моделируя работу машины при всех вариантах бросаний. С ростом n оно возрастает (точнее, не убывает) и стремится к вероятности остановки машины (без ограничения числа шагов).

(б) Пусть q — произвольное перечислимое снизу число. Это означает, что $q = \lim q_n$ для некоторой вычислимой последовательности

$$q_0 \leqslant q_1 \leqslant q_2 \leqslant \ldots$$

рациональных чисел. Построим машину, вероятность остановки которой равна q. Машина бросает монету и воспринимает полученные биты $b_0, b_1, b_2, \ldots$ как последовательные знаки двоичного числа $\beta = 0{,}b_0 b_1 b_2 \ldots$; параллельно она вычисляет рациональные числа $q_0, q_1, q_2, \ldots$ и останавливается, как только имеющейся у неё информации достаточно, чтобы утверждать, что $\beta < q$. Другими словами, она останавливается, как только $0{,}b_0 b_1 \ldots b_i 111\ldots$ (текущая верхняя оценка для числа β) оказывается меньше q_i (текущей нижней оценки для числа q). Символически этот процесс изображён на рис. 4.1.

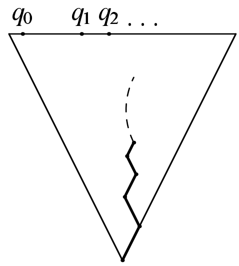


Рис. 4.1. Сравнение $\beta = 0{,}b_0 b_1 \ldots$ и $q = \lim q_i$.

Построенная машина останавливается тогда и только тогда, когда $\beta < q$. Убедимся в этом. Если число β меньше q, то машина остановится. В самом деле, числа q_i стремятся к q, а верхние оценки для числа β стремятся к β, и потому в какой-то момент q_i превысит текущую верхнюю оценку. С другой стороны, в случае остановки $\beta < q$ по построению.

Таким образом, вероятность остановки есть вероятность события $\beta < q$, то есть длина промежутка $[0, q)$, то есть q (число β, составленное из случайных битов, равномерно распределено на отрезке $[0, 1]$). ▶

Вернёмся к вопросу о том, каким может быть распределение вероятностей на выходах вероятностной машины. Нам понадобится следующее определение. Говорят, что последовательность действительных чисел $p_0, p_1, p_2, \ldots$ *перечислима снизу*, если существует вычислимая всюду определённая функция p двух натуральных аргументов с рациональными значениями (разрешается также дополнительное значение $-\infty$), для которой

$$p(i, 0) \leqslant p(i, 1) \leqslant p(i, 2) \leqslant \ldots$$

и

$$p_i = \lim_{n\to\infty} p(i, n)$$

для любого i.

Можно сказать, что в этом определении мы требуем перечислимости снизу чисел p_i «равномерно по i». Можно было бы определить перечислимость снизу и по-другому, как показывает следующая теорема:

Теорема 45. *Последовательность $p_0, p_1, p_2 \ldots$ перечислима снизу тогда и только тогда, когда множество пар $\langle r, i\rangle$, где i — натуральное число, а r — рациональное число, меньшее p_i, перечислимо.*

◀ Перечислимость множества означает, что существует алгоритм, перечисляющий его элементы (в произвольном порядке, с произвольными промежутками; алгоритм может никогда не завершать работу, даже если множество конечно).

Предположим, что последовательность $p_0, p_1, p_2, \ldots$ перечислима снизу и $p_i = \lim_n p(i, n)$. Алгоритм перебирает все пары $\langle r, i\rangle$, возвращаясь к каждой паре бесконечное число раз. При n-м обращении к паре $\langle r, i\rangle$ мы сравниваем r с $p(i, n)$; если оказывается, что $r < p(i, n)$, то пара $\langle r, i\rangle$ включается в перечисление. Ясно, что таким образом мы перечислим все нужные пары и только их ($r < \lim_n p(i, n)$ тогда и только тогда, когда $r < p(i, n)$ для некоторого n).

Напротив, пусть свойство $r < p_i$ перечислимо. Возьмём алгоритм, перечисляющий такие пары. Чтобы вычислить $p(i, n)$, сделаем n шагов перечисления; отберём все пары $\langle r, i\rangle$ с данным i и возьмём в них максимальное r. Положим $p(i, n)$ равным этому r (если пар с нужным значением i не окажется, то $p(i, n) = -\infty$). Легко понять, что с ростом n пар становится больше и значение $p(i, n)$ может только возрасти, а предел $\lim_n p(i, n)$ равен p_i (поскольку в перечислении появляются все рациональные числа, меньшие p_i). ▶

Теперь можно дать обещанную характеризацию распределений вероятностей, соответствующих вероятностным машинам.

Теорема 46. (а) *Пусть M — произвольная вероятностная машина без входа, возможными выходами которой являются натуральные числа. Пусть p_i — вероятность появления числа i на выходе. Тогда последовательность p_i перечислима снизу и $\sum_i p_i \leqslant 1$.*

(б) *Пусть $p_0, p_1, \ldots$ — перечислимая снизу последовательность неотрицательных действительных чисел, причём $\sum_i p_i \leqslant 1$. Тогда существует вероятностная машина M, для которой вероятность появления числа i на выходе равна p_i.*

◀ Первая часть доказательства (что вероятности перечислимы снизу) проходит точно так же, как и раньше, только вместо единственной вероятности остановки мы оцениваем снизу вероятность остановки с ответом i (для каждого i).

Вторая часть на самом деле тоже мало меняется. Раньше у нас выделялась всё бóльшая часть пространства под область остановки. Теперь область остановки будет поделена на счётное число подобластей; для каждого i есть область остановки с ответом i. Возрастающие приближения снизу к p_i составляют требования к размеру соответствующей области. Наша задача — распределять пространство между всеми требованиями. Это можно делать самым простым способом, выделяя каждому новому требованию кусок отрезка (слева направо) в соответствии с увеличением его заказа. Таким образом отрезок (не обязательно весь!) распределяется

между различными возможными выходами. Заметим, что область остановки с выходом i не является связной (она состоит из отдельных промежутков, число которых увеличивается по ходу процесса).

Параллельно с распределением отрезка по областям мы порождаем биты случайного числа β, и как только становится ясно, в чью область попадёт β, соответствующее число выдаётся на выход.

Более формально можно сказать это следующим образом. Пусть $p_i = \lim_n p(i, n)$ в соответствии с определением перечислимости снизу. Без ограничения общности можно считать, что все $p(i, n)$ неотрицательны (заменим отрицательные на нуль), и что для каждого n лишь конечное число значений $p(i, n)$ отличны от нуля (положим $p(i, n) = 0$ при $i > n$). Будем откладывать от нуля слева направо отрезки, помечая каждый натуральным числом. При этом мы хотим, чтобы к n-му шагу суммарная длина всех отрезков, помеченных числом i, равнялась $p(i, n)$. Поэтому при увеличении n на единицу мы смотрим, насколько увеличились значения $p(i, n)$ при разных i, и добавляем отрезки недостающего размера с нужными пометками.

За пределы отрезка $[0, 1]$ мы не выйдем, поскольку $p(i, n) \leqslant p_i$ и $\sum p_i \leqslant 1$.

Вероятностная машина действует следующим образом: мы останавливаемся с выходом i, если полученные к данному моменту случайные биты $b_0 b_1 \ldots b_k$ таковы, что отрезок на действительной прямой, состоящий из чисел, двоичные записи которых начинаются на $b_0 b_1 \ldots b_k$, целиком содержится во внутренности одного из отрезков, помеченного числом i. (Внутренность отрезка $[u, v]$ есть интервал (u, v).) Легко проверить, что выход i появится тогда и только тогда, когда β принадлежит внутренности одного из отрезков, помеченных числом i, а общая мера этого множества (объединения соответствующих интервалов) равна как раз p_i. ►

Будем называть *перечислимой снизу полумерой* на $\mathbb{N}$ любую последовательность чисел p_i, обладающую указанными в теореме свойствами. (Иногда мы будем также писать $p(i)$ вместо p_i.) Таким образом, у нас есть два определения перечислимых снизу полумер: (1) распределения вероятностей на выходах вероятностных машин; (2) перечислимые снизу ряды с неотрицательными членами и суммой не больше 1. Только что доказанная теорема утверждает, что эти два определения эквивалентны.

Слово «полумера» выглядит довольно странно (особенно учитывая обиходное выражение «ограничиться полумерами»), но другого употребительного термина нет. Если не требовать перечислимости, можно назвать *полумерой* на $\mathbb{N}$ любую функцию $i \mapsto p_i$, для которой $\sum_i p_i \leqslant 1$. Такие функции можно рассматривать как распределения вероятностей на множестве $\mathbb{N} \cup \{\perp\}$, где $\perp$ — специальный символ «неопределённости». При этом вероятность числа i есть p_i, а вероятность $\perp$ есть $1 - \sum_i p_i$. Но мы будем рассматривать лишь перечислимые снизу полумеры (если обратное не оговорено явно).

Мы определяли (перечислимые снизу) полумеры на множестве натуральных чисел. Но эти определения без всяких изменений переносятся и на двоичные слова (или любые другие конструктивные объекты). Например, если мы рассмотрим вероятностную машину, выходами которой являются двоичные слова, то она задаёт некоторую перечислимую снизу полумеру на множестве двоичных слов.

Важное замечание: в главе 5 мы будем рассматривать другое понятие полумеры на пространстве конечных и бесконечных двоичных последовательностей. Это понятие будет соответствовать машинам, которые порождают выход бит за битом и не обязаны останавливаться. Но пока что появление на выходе слова x означает, что машина напечатала все биты слова x и остановилась, так что двоичные слова ничем не отличаются в этом смысле от натуральных чисел.

Чтобы подчеркнуть различие между этими определениями, мы будем называть полумеры из этого раздела *дискретными*, в отличие от *непрерывных* полумер, или *полумер на дереве*, которые мы рассматриваем в главе 5.

4.2. Наибольшая полумера

Будем сравнивать полумеры на $\mathbb{N}$ с точностью до мультипликативной константы. Назовём перечислимую снизу полумеру m *максимальной*, или *наибольшей*, если для любой другой перечислимой снизу полумеры m' выполнено равенство $m'(i) \leqslant cm(i)$ для некоторого c и для всех i. (Термин «наибольшая» более точен, поскольку речь идёт именно о наибольших элементах в некотором частично упорядоченном множестве, но чаще говорят о максимальной перечислимой полумере.) Иногда используется термин «универсальная полумера».

Теорема 47. *Существует наибольшая перечислимая снизу полумера на* $\mathbb{N}$.

◀ Мы должны построить вероятностную машину M с таким свойством: для любого другой машины M' вероятность появления произвольного числа i на выходе машины M меньше такой же вероятности для M' не более чем в константу раз.

Этого можно добиться так: пусть машина M сначала случайно выберет вероятностную машину (вероятности выбрать ту или иную машину могут быть любыми, важно только, чтобы всякая вероятностная машина появлялась с положительной вероятностью), а потом моделирует выбранную машину. Если вероятность того, что машина M' будет выбрана, равна p, то вероятность $m(i)$ появления i на выходе машины M не меньше $p \cdot m'(i)$, где $m'(i)$ — вероятность появления i на выходе машины M'. Поэтому можно положить $c = 1/p$.

Случайный выбор можно реализовать, например, так. Перенумеруем все машины каким-либо естественным образом; получится последовательность M_0, M_1, $M_2, \ldots$, включающая все вероятностные машины. Машина M бросает монету, ожидая появления первой единицы. Далее она моделирует работу машины M_i, где i — число нулей перед первой единицей. ▶

Для сравнения проведём параллельно доказательство этой теоремы на языке перечислимых снизу рядов. Мы должны доказать, грубо говоря, что существует «самый плохо сходящийся» перечислимый снизу ряд, члены которого мажорируют (с точностью до умножения на константу) любой другой сходящийся перечислимый снизу ряд. (Точнее говоря, нас интересуют не просто сходящиеся ряды, а ряды с суммой не больше единицы, но это не имеет значения, так как всё равно мы разрешаем умножение на константу.)

Идея доказательства проста: возьмём все перечислимые снизу ряды с неотрицательными членами и суммой не больше 1, и сложим их с коэффициентами, которые образуют сходящийся ряд. Суммарный ряд, очевидно, будет наибольшим (с точностью до умножения на константу). Вопрос только в том, как сделать, чтобы получился перечислимый снизу ряд.

Произвольная перечислимая снизу полумера задаётся вычислимой функцией $p\colon \langle i,n\rangle \mapsto p(i,n)$. Таких функций счётное число (поскольку алгоритмов счётное число). Расположим их в последовательность $p^{(0)}, p^{(1)}, p^{(2)}, \ldots$ и рассмотрим функцию

$$p(i,n) = \sum_{k=0}^{n} \lambda_k p^{(k)}(i,n),$$

где λ_k — вычислимая последовательность рациональных чисел с $\sum_k \lambda_k \leqslant 1$ (например, $\lambda_k = 2^{-k-1}$). Определённая таким образом функция p возрастает с ростом n (при фиксированном i), поскольку сумма включает всё больше членов и сами члены растут. При этом

$$\lim_{n\to\infty} p(i,n) = \sum_k \lambda_k \lim_{n\to\infty} p^{(k)}(i,n)$$

для любого i, то есть построенная полумера есть действительно сумма всех полумер с коэффициентами λ_k.

Это рассуждение, однако, содержит пробел. Проблема в том, что функция $p(i,n)$ должна быть вычислимой. Поэтому недостаточно просто расположить функции, задающие полумеры, в последовательность, нужно ещё, чтобы эта последовательность была вычислимой (как функция трёх аргументов). Мы не можем просто написать подряд все программы и сказать, что $p^{(k)}$ — это функция, вычисляемая k-й программой (для функций двух натуральных аргументов с рациональными значениями). Может случиться, что k-я программа не задаёт перечислимую снизу полумеру (вычисляет не всюду определённую функцию, или эта функция не монотонна по второму аргументу, или сумма пределов больше единицы).

Положение спасает такая

Лемма. *Всякую программу P для функции двух натуральных аргументов с неотрицательными рациональными значениями можно алгоритмически преобразовать в программу P', которая вычисляет всюду определённую функцию того же типа, задающую некоторую перечислимую снизу полумеру. При этом, если сама программа P задавала полумеру, то новая полумера* (*соответствующая P'*) *равна старой.*

Доказательство. Пусть P — данная нам программа (не обязательно всюду определённая). Для начала положим $P'(i,n)$ равным максимальному из неотрицательных чисел, которые получаются за n шагов при вычислении значений $P(i,0), \ldots$ $\ldots, P(i,n)$ (если за n шагов ни одно из вычислений не даёт результата или все результаты отрицательны, то $P'(i,n) = 0$). Это уже гарантирует всюду-определённость P', а также неотрицательность и монотонность по второму аргументу. При этом, если (для данного i) значения $P(i,n)$ определены при всех n, неотрицательны и не убывают по n, то $\lim_n P'(i,n) = \lim_n P(i,n)$.

Нам ещё нужно, чтобы $\sum p'_i \leqslant 1$, где $p'_i = \lim_n P'(i,n)$. Для начала заметим, что можно полагать $P'(i,n)$ равным нулю при всех $n < i$. (Ясно, что такое преобразование не нарушит монотонности и не изменит предела $\lim_n P'(i,n)$.) Сумма $P'(i,n)$ по всем i при фиксированном n теперь будет конечной, и её можно вычислить. Нам нужно, чтобы эта сумма не превосходила единицы. Чтобы этого добиться, принудительно скорректируем P': перестанем увеличивать P' с ростом n, как только такое увеличение сделает сумму слишком большой. (Сперва мы корректируем все значения при $n = 0$, потом при $n = 1$ и так далее.) Лемма доказана.

Используя лемму, мы можем расположить все перечислимые снизу полумеры в вычислимую последовательность и затем сложить их с коэффициентами, получив наибольшую полумеру и тем самым завершив другое доказательство теоремы 47.

Фиксируем некоторую наибольшую перечислимую снизу полумеру, которую будем обозначать m. Значение $m(i)$ этой полумеры на числе i мы будем называть *априорной вероятностью* числа i. Смысл слов «априорная вероятность» такой. Пусть имеется некоторое устройство — чёрный ящик, при включении которого на выходе может появиться (через некоторое время) некоторое натуральное число. Не зная ничего об устройстве ящика, мы хотим до его включения (*a priori*, как сказали бы философы) оценить сверху вероятность появления на выходе числа i. Так вот, если ящик считать вероятностной машиной, а в качестве оценки выбрать число $m(i)$, то мы не сильно занизим оценку (максимум — в константу раз).

Как мы увидим, априорная вероятность числа i тесно связана с его сложностью. Грубо говоря, она тем больше, чем число проще. Это утверждение имеет вполне точный смысл: мы вскоре покажем, что несколько модифицированная сложность (так называемая «префиксная» сложность) числа i равна $-\log m(i)$.

4.3. Префиксные машины

Префиксная сложность отличается от обычной тем, что мы рассматриваем «самоограниченные описания»: декодирующей машине не сообщается, где кончается описание, а она решает это сама. Эту идею можно уточнять разными (и не эквивалентными) способами. Мы обсудим их подробно далее, а сейчас приведём формальные определения.

Пусть f — функция, аргументами которой являются двоичные слова. Будем говорить, что функция f является *префиксно корректной*, если выполнено такое свойство:

$$(f(x) \text{ определено}) \text{ и } (x \text{ — начало } y) \;\Rightarrow\; f(y) \text{ определено и равно } f(x).$$

Теорема 48. *Среди префиксно корректных способов описания имеется оптимальный.*

◀ Напомним, что способом описания мы называем вычислимую функцию, аргументами и значениями которой являются двоичные слова. Теперь мы рассматриваем не все такие функции, как раньше, а лишь префиксно корректные, и сравниваем соответствующие меры сложности (сложность по-прежнему определяется как длина

кратчайшего описания). Теорема утверждает, что в классе префиксно корректных способов описания имеется оптимальный *для этого класса* способ D; оптимальность означает, что для любого *префиксно корректного* способа D' имеет место неравенство $KP_D(x) \leqslant KP_{D'}(x) + O(1)$.

Здесь мы пишем KP вместо KS, чтобы подчеркнуть, что мы рассматриваем только префиксно корректные способы описания, хотя для данного способа D определение $KP_D(x)$ ничем не отличается от $KS_D(x)$.

Раньше (для обычной сложности) оптимальный способ описания строился так:

$$D(\hat{p}y) = p(y),$$

где $\hat{p}$ — какой-либо беспрефиксный код слова p (например, $\hat{p} = \bar{p}01$, где в $\bar{p}$ удвоен каждый бит слова p), а $p(y)$ — результат применения программы p к входу y (слово p понимается как программа в некотором универсальном языке программирования).

Будет ли D префиксно корректным? Легко понять, что нет, поскольку среди программ p есть всякие, в том числе и не префиксно корректные. Если какая-то программа p некорректна и, скажем, $p(0) = a$ и $p(00) = b$ при $a \neq b$, то $D(\hat{p}0) = a$ и $D(\hat{p}00) = b$, что противоречит требованию корректности для D.

Поэтому мы принудительно «скорректируем» программы и вместо $p(y)$ будем рассматривать $[p](y)$, определяемое так:

(1) Параллельно применяем программу p ко всех словам. Как только какое-то вычисление заканчивается, мы выписываем его аргумент и результат. Получаем последовательность пар $\langle y_i, z_i \rangle$, у которых $z_i = p(y_i)$.

(2) Эта последовательность прореживается с учётом требования корректности. Удобно использовать такую терминологию: слова y и y' *совместны*, если одно из них является началом другого (равносильная формулировка: если оба они являются началом некоторого третьего слова). Прореживание состоит в следующем: если какая-то пара $\langle y_i, z_i \rangle$ противоречит одной из предыдущих пар $\langle y_j, z_j \rangle$ при $j < i$, то она выбрасывается. Слово «противоречит» означает, что y_i совместно с y_j, но $z_i \neq z_j$. (Заметим, что можно было бы сравнивать очередную пару только с невыброшенными парами $\langle y_j, z_j \rangle$, но это не важно.)

(3) Все эти построения не зависят от входа y. Вычисляя $[p](y)$, мы ожидаем появления (невыброшенной) пары $\langle y_i, z_i \rangle$, у которой y_i является началом y. Как только такая пара появляется, мы выдаём результат z_i в качестве $[p](y)$.

Легко понять, что какова бы ни была программа p, функция $y \mapsto [p](y)$ является префиксно корректной. В самом деле, пусть $[p](y) = z$. Это значит, что в «прореженной» последовательности встретилась пара $\langle y_i, z \rangle$, где y_i является началом слова y. Пусть теперь y является началом слова y'. При вычислении $[p](y')$ мы также можем воспользоваться парой $\langle y_i, z \rangle$. Более ранние (невыброшенные) пары $\langle y_j, z_j \rangle$ нам либо не подойдут (если y_j не совместно с y_i, то y_j не может быть началом слова y'), либо дадут тот же результат (если y_j совместно с y_i, то $z_i = z_j$).

Если программа p с самого начала была префиксно корректной, то её корректировка ничего не меняет, то есть $[p](y) = p(y)$ при всех y. В самом деле, выброшенных пар не будет, и значение $[p](y)$ совпадёт со значением p на самом слове y или на каком-то его начале (что одно и то же в силу корректности p).

Осталось проверить, что построенный способ описания D является префиксно корректным и оптимальным (в классе префиксно корректных).

В самом деле, мы должны сравнить $D(\hat{p}_1 y_1)$ и $D(\hat{p}_2 y_2)$ в случае, когда $\hat{p}_1 y_1$ является началом $\hat{p}_2 y_2$. В этом случае $\hat{p}_1$ и $\hat{p}_2$ (будучи началами слова $\hat{p}_2 y_2$) совместны, и по свойствам беспрефиксного кодирования $p_1 = p_2$. Следовательно, y_1 есть начало y_2, и можно воспользоваться корректностью $[p_1]$ (или, что то же самое, $[p_2]$).

Корректность D доказана. Оптимальность вытекает из того, что $[p](y) = p(y)$ для корректного p, и потому переход от произвольного корректного D' (с программой p) к D увеличивает сложность не более чем на длину слова $\hat{p}$. ▶

Фиксировав произвольным образом оптимальный префиксно корректный способ описания, мы опускаем индекс и говорим о *префиксной сложности* $KP(x)$ слова x. Как обычно, надо иметь в виду, что замена способа описания приводит к изменению функции сложности (но только на ограниченное слагаемое).

Существует и другой вариант определения префиксной сложности, в котором требование корректности заменяется на другое. Будем называть функцию, аргументами которой являются двоичные слова, *беспрефиксной*, если никакие два слова в её области определения не сравнимы. Если беспрефиксная функция определена на каком-то слове, то она уже не может быть определена ни на его началах, ни на его продолжениях.

Будем рассматривать теперь только беспрефиксные способы описания, то есть беспрефиксные вычислимые функции, аргументами и значениями которых являются двоичные слова. Для них справедлива теорема, аналогичная теореме 48:

Теорема 49. *Среди беспрефиксных способов описания имеется оптимальный.*

◀ Доказательство следует той же схеме, но «корректировать» программы следует иначе. Для каждой программы p построим беспрефиксную функцию $y \mapsto \{p\}(y)$ следующим образом:

(1) Как и раньше, параллельно применяем p ко всем входам, получая последовательность пар $\langle y_i, z_i \rangle$, у которых $z_i = p(y_i)$.

(2) Эта последовательность прореживается: если в какой-то паре $\langle y_i, z_i \rangle$ слово y_i совместно с каким-то из предыдущих y_j (при $j < i$), то эта пара выбрасывается.

(3) Далее, имея вход y, мы ожидаем появления (невыброшенной) пары $\langle y_i, z_i \rangle$ с первым членом $y_i = y$. Второй член этой пары и будет $\{p\}(y)$.

Легко проверить, что функция $y \mapsto \{p\}(y)$ является беспрефиксной и что описанное преобразование оставляет нетронутыми беспрефиксные функции.

Далее рассуждение в точности повторяет доказательство теоремы 48. ▶

Теперь можно фиксировать оптимальный беспрефиксный способ описания и рассмотреть соответствующую функцию сложности (обозначим её KP').

Какая же из мер сложности KP и KP' является «правильной»? Это скорее дело вкуса. Мы увидим вскоре (см. раздел 4.5), что на самом деле эти меры совпадают с точностью до $O(1)$ (и совпадают с минус логарифмом априорной вероятности), так что скорее следует спрашивать не о том, какая мера сложности лучше, а о том, какое определение более естественно. Это тем более вопрос вкуса. Авторам представляется, что более естественно определение с префиксно-корректными

способами описания (не зря мы его привели первым). Вместе с тем в некоторых рассуждениях (например, при доказательстве теоремы о сложности пары в разделе 4.6) использование второго определения позволяет дать простое и короткое доказательство.

(Об истории появления префиксной сложности см. [11]; отметим, что в первых русских публикациях, начиная с диссертации Левина, использовались префиксно-корректные описания, а Чейтин использовал беспрефиксные. В современной англоязычной литературе для префиксной сложности используется обозначение $K(x)$.)

Насколько свойства префиксной сложности (KP, KP') отличаются от свойств обычной? Что из ранее известных свойств остаётся в силе?

- Прежде всего отметим очевидное свойство:

  $$KS(x) \leqslant KP(x) + O(1) \quad \text{и} \quad KS(x) \leqslant KP'(x) + O(1),$$

  поскольку беспрефиксные и префиксно корректные способы описания, используемые при определении префиксной сложности, являются частным случаем способов описания из определения KS.

- Мы видели, что $KS(x) \leqslant l(x) + O(1)$ (достаточно рассмотреть тождественный способ описания). Теперь это рассуждение не проходит, так как тождественный способ описания не является ни беспрефиксным, ни префиксно корректным. Как мы покажем в разделе 4.5, это не случайно: для префиксной сложности аналогичное неравенство неверно.

- Тем не менее можно дать оценки префиксной сложности в терминах длины. (Мы сделаем это для KP', конструкции для KP аналогичны.) Докажем, что $KP'(x) \leqslant 2l(x) + O(1)$. В самом деле, способ описания D, определённый формулой

  $$D(\bar{x}01) = x$$

  ($\bar{x}$ получается удвоением всех битов в x), является беспрефиксным. Выбирая более экономное беспрефиксное кодирование $\hat{x}$ вместо $\bar{x}01$, можно получить и лучшие оценки:

  $$KP'(x) \leqslant l(x) + 2\log l(x) + O(1)$$

  получится, если положить $\hat{x} = \overline{\mathrm{bin}(l(x))}01x$; итерируя эту конструкцию, получаем

  $$KP'(x) \leqslant l(x) + \log l(x) + 2\log\log l(x) + O(1)$$

  и так далее.

- Свойство невозрастания сложности при алгоритмическом преобразовании остаётся верным:

  $$KP'(A(x)) \leqslant KP'(x) + O(1)$$

(константа зависит от алгоритма A, но не от x). В самом деле, легко проверить, что если D — беспрефиксный способ описания, то композиция $x \mapsto A(D(x))$ также является беспрефиксным способом описания. Заменив слово «беспрефиксный» на «префиксно корректный», получаем аналогичное утверждение для KP вместо KP'. Это свойство позволяет говорить о префиксной сложности произвольных конструктивных объектов (пар слов, натуральных чисел, конечных множеств слов и т. п.), не уточняя способа их кодирования.

- Как мы увидим, неравенство для сложности пары слов с префиксной сложностью уже не требует логарифмической добавки:

$$KP(x, y) \leqslant KP(x) + KP(y) + O(1)$$

(теорема 60 в разделе 4.6, с. 113).

- Применим свойство невозрастания сложности при алгоритмическом преобразовании, взяв в качестве A оптимальный способ описания D для обычной (не префиксной) колмогоровской сложности. Если p является кратчайшим описанием x относительно D, то $D(p) = x$ и $l(p) = KS(x)$, поэтому

$$KP(x) = KP(D(p)) \leqslant KP(p) + O(1) \leqslant l(p) + 2\log l(p) + O(1) = \\ = KS(x) + 2\log KS(x) + O(1).$$

Мы воспользовались свойством $KP(p) \leqslant l(p) + 2\log l(p) + O(1)$, доказанным ранее; если использовать более точные оценки, то получим неравенство

$$KP(x) \leqslant KS(x) + \log KS(x) + 2\log\log KS(x) + O(1)$$

и аналогичные ему неравенства.

4.4. Отступление: машины с самоограниченным входом

Этот раздел почти не используется в дальнейшем. Мы попытаемся проанализировать идею «самоограниченного» входа и её возможные уточнения, тем самым мотивировав определения префиксно корректных и беспрефиксных способов описания.

Обычно, подавая на вход машины двоичное слово, мы указываем начало и конец этого слова. Например, при определении вычислимости на машине Тьюринга обычно предполагают, что машина изначально видит первый символ слова, а конец слова отмечен специальным указателем (им может быть, например, символ «пробел», который идёт за последним битом слова).

Говоря о самоограниченном входе, мы имеем в виду несколько другую ситуацию: машина получает биты слова один из другим (слева направо) и в некоторый момент выдаёт ответ.

4.4.1. Беспрефиксные функции

Вот одно из возможных уточнений. Будем считать, что у машины, помимо рабочей ленты, есть *входная лента*, на которой имеется односторонняя читающая головка. Крайняя левая клетка ленты содержит специальный маркер #, справа от которого записана бесконечная последовательность нулей и единиц (рис. 4.2).

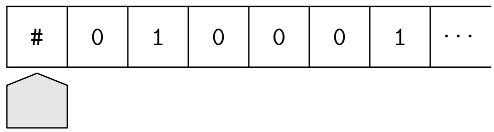


Рис. 4.2. Односторонняя читающая головка на входной ленте.

Изначально читающая головка стоит у левого края ленты (и видит специальный маркер). Поведение машины Тьюринга определяется символом, который видит читающая головка, а также (как всегда) символом, который видит головка на рабочей ленте. В зависимости от этих символов и текущего состояния машина предпринимает то или иное действие. Это действие состоит в изменении внутреннего состояния, записи нового символа на рабочей ленте, а также может включать в себя сдвиг на рабочей ленте и сдвиг вправо читающей головки. Результат работы машины обычным образом записывается на рабочей ленте (изначально пустой).

Пусть M — такая машина. Будем запускать её на всевозможных входных лентах. Как только машина останавливается, мы записываем два слова: ту часть входной ленты, которую она успела прочесть (слово x), и результат работы (y). Множество полученных таким образом пар $\langle x, y\rangle$ обозначим Γ_M. Если две различные пары $\langle x_1, y_1\rangle$ и $\langle x_2, y_2\rangle$ принадлежат Γ_M, то слова x_1 и x_2 несравнимы. В самом деле, если x_1 является началом x_2, то вычисление с x_2 на входе должно протекать так же, как и вычисление с x_1 (поскольку символы на входе те же) и вместе с ним должно закончиться. Поэтому $y_1 = y_2$, а если $x_1 \neq x_2$, то часть слова x_2 останется непрочитанной, что противоречит определению множества Γ_M.

В частности, первые члены всех пар из Γ_M различны, поэтому Γ_M задаёт некоторую функцию γ_M, аргументами и значениями которой являются двоичные слова. Будем говорить, что машина M *беспрефиксно вычисляет* эту функцию. Легко понять, что функция γ_M вычислима в обычном смысле: чтобы вычислить $\gamma_M(x)$, надо запустить M на x и после остановки дополнительно проверить, что головка на входной ленте дошла до конца слова x, но не вышла за его пределы. Ясно также, что функция γ_M является беспрефиксной (любые два слова из её области определения несравнимы). Верно и обратное утверждение:

Теорема 50. *Любая беспрефиксная вычислимая функция беспрефиксно вычислима некоторой машиной.*

◄ Это утверждение вовсе не самоочевидно, поскольку машина F, вычисляющая (в обычном смысле) некоторую беспрефиксную функцию f, знает, где кончается вход, и может использовать эту информацию. Тем не менее оно верно: если

функция f беспрефиксная, мы можем построить другую машину M, для которой $\gamma_M = f$.

Это делается следующим образом. Машина M на рабочей ленте параллельно моделирует вычисления машины F на всех аргументах, время от времени читая очередные символы на входной ленте (когда именно это нужно делать, описано ниже). При появлении новой пары слов x, y, для которых $f(x) = y$, мы сравниваем x с уже прочитанной частью входа (вначале она пуста). Если прочитанная часть входа не является началом x, то мы не делаем ничего (и продолжаем порождать пары слов). Если прочитанная часть совпадает с x, то мы останавливаемся и выдаём на вход y. Если же прочитанная часть является собственным началом слова x, то мы читаем следующий бит входа и повторяем сравнение — до тех пор, пока мы не прочтём на входе либо само слово x (в этом случае M останавливается с выходом y), либо какое-то слово, не являющееся началом слова x. На этом обработка пары $\langle x, y\rangle$ заканчивается, и мы ждём появления следующей пары.

Как выглядит этот процесс вычисления? Вначале на входе не прочитано ничего. Когда появляется первая пара $\langle x, y\rangle$, мы смотрим, пусто ли x. Если да, то (так ничего и не прочтя на входе) печатаем на выходе y и останавливаемся. Если нет, то читаем вход до тех пор, пока либо не прочтём всё x, либо не отклонимся от x (то есть прочтём кусок входа, который не является началом слова x). В первом случае мы печатаем на выходе y и останавливаемся, во втором заканчиваем обработку пары $\langle x, y\rangle$ и ждём появления следующей пары.

Формально говоря, инвариант, который выполнен после обработки нескольких пар, таков (обозначим прочитанную часть входа через r): либо

(1) $f(r)$ определено, машина закончила работу и выдала на выход $f(r)$, либо

(2) r не является началом слова x ни для какой из обнаруженных пар $\langle x, y\rangle$, но всякое собственное начало r' слова r является собственным началом одного из таких слов. (Собственное начало — это начало, не совпадающее со всем словом.)

Используя этот инвариант, легко завершить формальную проверку, но мы вместо этого подчеркнём основную идею: если мы уже прочли некоторое слово r, и выяснилось, что функция f определена на некотором собственном продолжении слова r, то заведомо $f(r)$ не определено, и потому можно прочесть следующий бит входа, не опасаясь прочесть лишнее. ►

Фактически та же модель вычислений может быть описана в более привычных для программистов терминах. Представим себе программу, которая использует оператор

$$b := \texttt{NextBit}$$

При выполнении этого оператора работа программы приостанавливается, на экране появляется надпись «Введите следующий бит». Когда клиент это делает (скажем, нажав клавишу «0» или «1»), соответствующий бит помещается в переменную b, и работа программы продолжается.

Каждой такой программе соответствует функция: её аргумент x бит за битом сообщается программе в ответ на её запросы; значение равно y, если программа напечатала y и остановилась, при этом запросив все биты слова x и только их.

(Если она запросила следующий бит, прочтя всё слово x, или недоузнала биты этого слова, то x не входит в область определения функции.)

Адаптируя приведённые выше рассуждения, легко показать, что функции, соответствующие таким программам — это в точности вычислимые беспрефиксные функции. (Сдвиг головки на входной ленте в точности соответствует запросу следующего бита.)

4.4.2. Префиксно корректные функции

Есть и другая, более привычная схема работы программы, получающей свой вход бит за битом (причём конец входа никак не отмечается). Будем считать, что клиент набирает входное слово, не дожидаясь запросов, нажимая на клавиатуре клавиши «0» и «1» (и никак не обозначая конец входа). Нажатые им клавиши запоминаются и поступают в программу по требованию.

Очередной бит читается командой

$$b := \texttt{NextBit}$$

Помимо неё, есть команда

$$b := \texttt{NextExists}$$

Она позволяет выяснить, имеется ли в очереди нажатых клавиш ещё не прочитанная. Следует уточнить также, что происходит, если при выполнении команды `NextBit` очередь нажатых клавиш пуста. Можно считать, что это ведёт к аварии, а можно считать, что работа всего лишь приостанавливается до появления следующего бита (нажатия следующей клавиши). Какой из этих вариантов выбрать, не имеет значения, поскольку можно дожидаться появления входного бита, прежде чем его читать. Для этого нужно написать что-то вроде

while not `NextExists` **do** {nothing};
$b :=$ `NextBit`

Программисты бы назвали описываемый в этом разделе механизм доступа к входу «неблокирующим» чтением в отличие от ранее рассмотренного «блокирующего». Неблокирующее чтение позволяет программе продолжать внутреннюю работу, одновременно следя за входом и ожидая появления там нового символа.

При этом возникает очевидная проблема: вообще говоря, выход теперь зависит не только от входных битов, но также и от момента, в который они были поданы на вход.

Назовём программу *корректной*, если для неё результат работы не зависит от моментов нажатия клавиш: для данного входного слова x либо работа программы не заканчивается, в какие моменты его ни подавай, либо всегда заканчивается с одним и тем же результатом.

Каждой корректной программе соответствует частичная функция (зависимость выхода от входа).

Теорема 51. (а) *Эта функция вычислима и является префиксно корректной.* (б) *Любая префиксно корректная вычислимая функция соответствует некоторой корректной программе.*

◀ (а) Вычислимость функции очевидна: если есть программа, вычисляющая её в описанном режиме (вход подаётся по частям), то есть и программа, вычисляющая её в обычном смысле. Проверим префиксную корректность. По определению (раздел 4.3) мы должны доказать, что если корректная программа выдаёт выход y на входе x, то она даёт тот же выход и для входа x', являющегося продолжением входа x. В самом деле, проследим за работой программы на входе x. По предположению она напечатает y и остановится. В самый последний момент, когда машина уже всё сделала и собирается остановиться, добавим к входу x недостающий до x' кусок. Это уже не повлияет на работу программы, и она по-прежнему напечатает y при входе x'. В силу корректности программа будет печатать слово y при любом другом выборе моментов нажатия клавиш для букв из x'.

(Замечание в скобках: на самом деле спешка тут нужна лишь для наглядности, ведь мы можем нажать оставшиеся клавиши и после завершения работы программы, определение этого не запрещает.)

(б) Покажем, что для любой префиксно корректной вычислимой функции можно построить соответствующую программу. Пусть f — такая функция.

Корректная программа действует следующим образом. Она параллельно применяет f ко всем словам, а также регулярно проверяет, не поступил ли на вход новый символ (и читает его). Если выяснится, что $f(x) = y$ для некоторых x и y, причём текущее слово на входе равно x или некоторому продолжению слова x, то на выход выдаётся y и работа завершается.

Если $f(x) = y$, то эта программа выдаст y, в какие моменты не подавай биты слова x на вход. В самом деле, если подождать достаточно долго, то слово x будет подано целиком, а моделирование $f(x)$ завершится. В этот момент машина остановится и выдаст y, если она не остановилась раньше. А раньше она могла остановиться, если $f(x')$ оказалось определённым для некоторого начала x' входной последовательности. Но тогда x' является началом x и $f(x')$ должно равняться тому же слову y, поскольку функция f префиксно корректна по предположению. Поэтому на ответ это не повлияет.

С другой стороны, если $f(x)$ не определено, а функция f префиксно корректна, то $f(x')$ не определено и для всех начал слова x, поэтому работа машины не закончится. ▶

Эта теорема показывает, что корректным машинам с неблокирующим чтением соответствуют вычислимые префиксно корректные функции и только они, тем самым мотивируя определение префиксно корректной функции.

96 Докажите, что существует алгоритм, преобразующий любую программу p описанного вида (с командами `NextBit` и `NextExists`) в другую программу p' того же вида, причём p' всегда корректна и вычисляет ту же функцию, что и p, если программа p корректна. [Указание. Воспользуйтесь конструкцией из доказательства теоремы 51 в обе стороны.]

97 (Продолжение.) Докажите, что тем не менее не существует алгоритма, который по заданной программе p проверял бы, является ли она корректной. [Указание. Это делается обычным для теории вычислимых функций способом: сведением проблемы остановки. См., например, [175].]

4.4.3. Непрерывные вычислимые отображения

Существует и другая, более абстрактная мотивировка понятия префиксно корректной функции. Она исходит из общей схемы определения вычислимости для объектов высших типов, но мы изложим её применительно к конкретной ситуации (уже и так этот раздел — предназначенный дотошным читателям! — чрезмерно скучен). Коротко говоря, префиксно корректные функции соответствуют вычислимым непрерывным отображениям битовых последовательностей в натуральные числа с дополнительным элементом «неопределённость». Сейчас мы объясним, что имеется в виду.

Рассмотрим множество Σ конечных и бесконечных двоичных последовательностей: $\Sigma = \Xi \cup \Omega$. Для каждого (конечного) слова x рассмотрим множество Σ_x его конечных и бесконечных продолжений. Введём на Σ частичный порядок, считая, что $x \leqslant y$, если y является продолжением x.

Введём на множестве Σ топологию, считая базовыми открытыми множествами множества Σ_x (это значит, что открытыми множествами считаются всевозможные объединения множеств вида Σ_x). Легко проверить, что это действительно топология (не обладающая свойством отделимости).

Имеет место следующий (почти очевидный) факт:

Теорема 52. *Множество $A \subset \Sigma$ открыто в этой топологии тогда и только тогда, когда оно обладает двумя свойствами*:

(1) *вместе с каждым двоичным словом оно содержит все его* (*конечные или бесконечные*) *продолжения*;

(2) *если бесконечная последовательность принадлежит A, то некоторое её конечное начало также принадлежит A.*

◄ Объединение базовых открытых множеств очевидно обладает свойствами (1) и (2). С другой стороны, если A обладает этими свойствами, то оно является объединением множеств Σ_x по всем (конечным) словам x, принадлежащим A. ►

Теперь введём топологию на множестве $\mathbb{N}_\perp$ натуральных чисел с добавленным к нему элементом $\perp$ (неопределённость). На этом множестве также полезно ввести частичный порядок, считая, что $\perp$ меньше всех остальных элементов, а между собой они несравнимы (рис. 4.3).

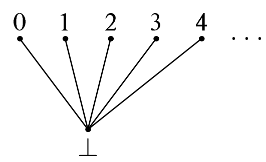


Рис. 4.3. Топологическое пространство $\mathbb{N}_\perp$.

Будем считать открытыми любые множества, не содержащие элемента $\perp$, а также всё пространство $\mathbb{N} \cup \{\perp\}$. (Легко проверить, что аксиомы топологического пространства выполнены; пространство это также не является отделимым.)

Частичные функции из Σ в $\mathbb{N}$ отождествим с отображениями вида $\Sigma \to \mathbb{N}_\perp$, определёнными на всём Σ (значение $\perp$ соответствует неопределённым значениям функции). Легко описать класс непрерывных отображений (напомним, что отображение называется непрерывным, если прообраз любого открытого множества открыт):

Теорема 53. *Отображение $F\colon \Sigma \to \mathbb{N}_\perp$ непрерывно тогда и только тогда, когда выполнены два условия*:

(1) *F монотонно (если $x \leqslant y$, то $F(x) \leqslant F(y)$ в смысле введённых на Σ и $\mathbb{N}_\perp$ порядков)*;

(2) *Если $F(x) \neq \perp$ для бесконечной последовательности x, то $F(x') \neq \perp$ для некоторого конечного начала $x' \leqslant x$.*

◀ Пусть F непрерывно. Проверим условие (1). Пусть $x \leqslant y$. Если $F(x) \nleqslant F(y)$, то $F(x)$ есть натуральное число (а не $\perp$) и $F(x) \neq F(y)$. Тогда прообраз открытого множества $\{F(x)\}$ содержит x, но не содержит y и потому не является открытым.

Проверим условие (2). Если $F(x) \neq \perp$ для бесконечной последовательности x, то открытый прообраз открытого множества $\{F(x)\}$ содержит x и потому должен содержать некоторое конечное начало последовательности x.

Осталось проверить, что если для F выполнены условия (1) и (2), то отображение F непрерывно. Достаточно проверить, что прообраз каждого натурального числа открыт (прообраз всего пространства открыт, а остальные открытые множества состоят из натуральных чисел). Для этого достаточно проверить условия (1) и (2) предыдущей теоремы, которые прямо следуют из предположения. (Заметим, что если x' есть начало x и $F(x') \neq \perp$, то $F(x') = F(x)$ в силу монотонности.) ▶

С каждым непрерывным отображением $F\colon \Sigma \to \mathbb{N}_\perp$ свяжем множество Γ_F всех пар $\langle x, n\rangle \in \Xi \times \mathbb{N}$, для которых $F(x) = n$. Заметим, что множество Γ_F является лишь частью графика отображения F (мы рассматриваем лишь конечные слова x и не разрешаем элементу $\perp$ быть вторым членом пары.)

Теорема 54. *Соответствие $F \mapsto \Gamma_F$ является взаимно однозначным соответствием между непрерывными отображениями $\Sigma \to \mathbb{N}_\perp$ и множествами $A \subset \subset \Xi \times \mathbb{N}$, обладающими двумя свойствами*:

(1) $\langle x, n\rangle \in A$, $x \leqslant y \Rightarrow \langle y, n\rangle \in A$;

(2) $\langle x, n\rangle \in A$, $\langle x, m\rangle \in A \Rightarrow m = n$.

◀ Пусть отображение F непрерывно. Если $F(x) = n \in \mathbb{N}$, то условие (1) предыдущей теоремы гарантирует, что $F(y) = n$ для любого слова y, продолжающего x. Тем самым условие (1) выполнено для множества Γ_F. Выполнено и условие (2), поскольку $F(x)$ не может быть равно двум разным натуральным числам одновременно. Таким образом, для любого непрерывного F множество Γ_F обладает свойствами (1) и (2).

Легко видеть, что Γ_F однозначно определяет F: чтобы найти $F(x)$ для конечного x, мы ищем пару $\langle x, n\rangle \in \Gamma_F$. Если такой пары нет, то $F(x) = \perp$. Для бесконечного x значение $F(x)$ однозначно определяется по непрерывности.

Осталось показать, что любое множество A, обладающее свойствами (1) и (2), равно Γ_F при некотором F. В самом деле, определим $F(x)$ при конечном x как то

единственное n, для которого $\langle x, n\rangle \in A$ (свойство (2) гарантирует единственность) или как $\perp$, если такого n не существует. Свойство (1) гарантирует, что построенная таким образом функция (определённая на конечных словах) монотонна. Осталось доопределить её на бесконечных словах. Для бесконечного $x \in \Sigma$ положим $F(x)$ равным $F(x')$, если найдётся конечное $x' \leqslant x$, для которого $F(x') \neq \perp$; если такого x' не найдётся, то $F(x) = \perp$. Корректность определения (независимость $F(x)$ от выбора $x' \leqslant x$) гарантируется свойством (1). Очевидно, построенная таким образом функция F обладает свойствами (1) и (2) предыдущей теоремы и потому непрерывна. Столь же очевидно, что $\Gamma_F = A$. ►

Легко понять, что свойства (1) и (2) последней теоремы означают, что A представляет собой график префиксно корректной функции. Тем самым мы получаем взаимно однозначное соответствие между непрерывными отображениями $\Sigma \to \mathbb{N}_\perp$ и префиксно корректными функциями.

Назовём непрерывное отображение $F\colon \Sigma \to \mathbb{N}_\perp$ *вычислимым*, если множество Γ_F перечислимо. Легко проверить, что это равносильно вычислимости соответствующей частичной функции из Ξ в $\mathbb{N}$. (Частичная функция из Ξ в $\mathbb{N}$ вычислима тогда и только тогда, когда её график перечислим.) Поэтому вычислимые отображения типа $\Sigma \to \mathbb{N}_\perp$ соответствуют вычислимым префиксно корректным функциям, что и является обещанной дополнительной мотивировкой понятия префиксно корректной функции.

4.5. Основная теорема о префиксной сложности

В этом разделе мы докажем, что три меры сложности — два варианта префиксной сложности (KP с префиксно корректными способами описания и KP' с беспрефиксными), а также минус логарифм априорной вероятности совпадают с точностью до $O(1)$. Для этого мы докажем три неравенства в цепочке

$$-\log m(x) \leqslant KP(x) \leqslant KP'(x) \leqslant -\log m(x)$$

(с точностью до $O(1)$). Начнём с двух простых неравенств.

Теорема 55.
$$KP(x) \leqslant KP'(x) + O(1).$$

◄ Неравенство было бы очевидным, если бы всякий беспрефиксный способ описания был бы префиксно корректным. Однако это не так: если беспрефиксный способ D определён на каком-то слове u, то он не определён ни на каком продолжении слова u (в то время как префиксно корректный способ описания должен быть определён на всех продолжениях и иметь то же значение).

Поэтому требуется чуть более сложная конструкция. Пусть D — беспрефиксный способ описания. Построим новый способ описания D'. Если D был определён на некотором слове y и принимал на нём значение x, то мы положим D' равным x на всех продолжениях слова y. Другими словами, чтобы вычислить $D'(y)$, мы применяем D ко всем началам слова y. Лишь одно из значений $D(y')$ (где y' — начало слова y) может быть определённым (в силу беспрефиксности). Как только

такое y' найдётся, мы полагаем $D'(y)$ равным $D(y')$. Другими словами, $D'(y) = x$ тогда и только тогда, когда $D(y') = x$ для некоторого начала y' слова y.

Из построения видно, что функция D' вычислима, префиксно корректна и является продолжением функции D. Поэтому сложность относительно D' не превосходит сложности относительно D. (На самом деле сложность относительно D' в точности равна сложности относительно D, так как кратчайшие описания остались теми же самыми.) ►

Можно пытаться доказать обратное неравенство аналогичным способом. Действительно, для каждого префиксно корректного отображения D можно построить беспрефиксное отображение D', сузив D на «нижние точки» своей области определения (то есть положив $D'(y) = z$, если $D(y) = z$ и $D(y')$ не определено ни для какого начала y' слова y).

Это преобразование префиксно корректных отображений в беспрефиксные в точности обратно описанному выше преобразованию беспрефиксных отображений в префиксно корректные. Но тут мы сталкиваемся с проблемой, поскольку новое преобразование (в отличие от прежнего) не сохраняет вычислимость.

$\boxed{\mathbf{98}}$ Приведите пример вычислимой префиксно корректной функции, для которой соответствующая беспрефиксная функция не является вычислимой. [Указание. Пусть A — перечислимое множество натуральных чисел, не являющееся разрешимым и потому имеющее неперечислимое дополнение. Положим $f(0^n11x) = 0$ при всех n и всех x и $f(0^n1x) = 0$ при всех $n \in A$ и всех x.]

Эта задача говорит, что в некотором смысле неблокирующее чтение является более гибким средством, чем блокирующее (см. раздел 4.4).

Теорема 56.
$$-\log m(x) \leqslant KP(x) + O(1).$$

◄ Мы должны показать, что $-\log m(x) \leqslant KP(x) + O(1)$ или что $2^{-KP(x)} \leqslant$ $\leqslant cm(x)$ для некоторой константы c. В силу максимальности полумеры m достаточно показать, что функция $x \mapsto 2^{-KP(x)}$ не больше некоторой перечислимой снизу полумеры. (Здесь, говоря о полумерах, мы считаем аргументами двоичные слова, то есть рассматриваем вероятностные машины, выходом которых являются двоичные слова, см. раздел 4.1.)

Укажем эту полумеру, построив соответствующую вероятностную машину. Возьмём оптимальный префиксно корректный способ описания D, использованный в определении KP. Наша вероятностная машина (без входа) будет применять D к случайной последовательности битов. Точнее говоря, D применяется параллельно ко всем началам последовательности случайных битов $b_0, b_1, b_2, \ldots$; как только одно из вычислений
$$D(\Lambda), D(b_0), D(b_0b_1), D(b_0b_1b_2), \ldots$$
закончится, его результат и будет результатом работы машины. Заметим, что не имеет значение, какое именно из закончившихся вычислений мы возьмём (корректность гарантирует, что результат один и тот же).

Для любого слова x рассмотрим кратчайшее описание p слова x (относительно D). Тогда вероятность появления слова x как выхода вероятностной машины не меньше $2^{-l(p)}$. В самом деле, если первые $l(p)$ случайных битов совпали с p, то результат работы машины гарантированно равен x. Тем самым вероятность появления x не меньше $2^{-l(p)} = 2^{-KP(x)}$, что и требовалось. ►

Несколько другое доказательство того же результата обходится без вероятностных машин. В самом деле, мы знаем, что функция сложности перечислима сверху, поэтому функция $x \mapsto 2^{-KP(x)}$ перечислима снизу. Остаётся сослаться на следующую почти очевидную теорему:

Теорема 57.

$$\sum_x 2^{-KP(x)} \leqslant 1.$$

◄ В самом деле, пусть p_x — кратчайшая программа для слова x. Тогда слова p_x и p_y несравнимы (при $x \neq y$). Остаётся воспользоваться такой простой леммой.

Лемма. *Пусть $p_0, p_1, p_2, \ldots$ — попарно несравнимые слова (это значит, что ни одно из них не является началом другого). Тогда $\sum_i 2^{-l(p_i)} \leqslant 1$.*

В самом деле, $2^{-l(p_i)}$ есть (равномерная бернуллиева) мера множества Ω_{p_i} всех (бесконечных) продолжений слова p_i. Поскольку слова p_i несравнимы, множества их продолжений не пересекаются и потому меры этих множеств (вероятности получить p_i в начале случайной последовательности битов) в сумме не превосходят единицы. Лемма (а с ней и теорема 57) доказана. ►

Кстати, мы заодно доказали, что для префиксной сложности не выполняется неравенство $KP(x) \leqslant l(x) + O(1)$. В самом деле, если бы это было так, то сумма

$$\sum_x 2^{-l(x)}$$

была бы конечной. А между тем для каждого n сумма по всем x длины n равна единице (ибо состоит из 2^n членов по 2^{-n} каждый), и потому общая сумма бесконечна.

$\boxed{99}$ Докажите, что более слабая оценка $KP(x) \leqslant l(x) + \log l(x) + O(1)$ также не имеет места (то есть что разность $KP(x) - l(x) - \log l(x)$ не ограничена). [Указание. Гармонический ряд расходится.]

Осталось доказать третье, наиболее сложное неравенство:

Теорема 58.

$$KP'(x) \leqslant -\log m(x) + O(1).$$

◄ Попытаемся вначале объяснить идею доказательства. Имеется перечислимая снизу полумера $m(x)$. Это значит, что мы постепенно узнаём всё бо́льшие и бо́льшие оценки снизу для $m(x)$. Большое значение $m(x)$ означает для нас, что $KP'(x)$ должно быть малым, то есть означает необходимость предусмотреть для x короткое описание p. При этом описания различных объектов должны быть несравнимыми. Несравнимость удобно представлять себе геометрически. Для каждого двоичного слова p мы рассматриваем отрезок, который состоит из чисел, двоичная запись которых начинается на p. Таким образом, $[0, 1]$ разбивается на две половины

I_0 (левую) и I_1 (правую), которые в свою очередь делятся пополам на I_{00}, I_{01} и I_{10}, I_{11} и так далее. В этом представлении несравнимость слов p_1 и p_2 означает, что отрезки I_{p_1} и I_{p_2} не пересекаются. При этом неравенство $l(p) \leqslant -\log_2 m(x)$ можно переписать как $2^{-l(p)} \geqslant m(x)$: длина отрезка I_p должна быть не меньше $m(x)$.

Тем самым нашу задачу можно сформулировать так: каждому объекту x нужно выделить отрезок длины не менее $m(x)$, причём отрезки, выделенные разным объектам, не должны перекрываться.

Сказанное требует уточнений. Во-первых, разрешается выделять отрезки длиной не $m(x)$, а $\varepsilon m(x)$ для некоторого фиксированного ε (что соответствует аддитивной константе в исходном неравенстве). Во-вторых, нужны не произвольные отрезки, а только «регулярные» (половинки, четвертинки и другие отрезки вида I_p). Но с учётом первого замечания это не так важно, поскольку любой отрезок (не выходящий за пределы отрезка $[0,1]$) содержит строго внутри себя регулярный подотрезок длиной не менее четверти исходного.

Таким образом, мы приходим почти к той же задаче, что была рассмотрена в разделе 4.1. К нам приходят клиенты с просьбами выделить им место на отрезке $[0,1]$. По-прежнему сумма всех запросов не превосходит единицы и клиенты сообщают свои запросы постепенно (время от времени меняя их в сторону увеличения). Теперь, однако, клиентам важна не общая длина выделенных для них отрезков, а длина непрерывного участка, что усложняет задачу. В качестве компенсации мы имеем право выделять место с некоторым коэффициентом (урезать все запросы в фиксированное число раз). Но это нам не поможет, если при каждом увеличении запроса отводить новый участок — при такой стратегии никакого коэффициента не хватит. Решение проблемы: рассматривать новый запрос (выделяя новый отрезок) только, когда запрос увеличился вдвое. При этом общая сумма рассмотренных запросов не более чем вдвое превосходит последний запрос (ведь сумма геометрической прогрессии со знаменателем 2 не больше удвоенного последнего члена). С другой стороны, требования выполнены с точностью до мультипликативной константы.

Эту схему можно довести до вполне строгого доказательства. Мы, однако, изложим немного другое доказательство. Оно основано на следующей «лемме о шторах», называемой часто леммой Крафта – Чейтина (см. [23]). (Она является вычислимым аналогом леммы Крафта из теории информации, см. с. 244.)

Лемма. *Пусть $l_0, l_1, l_2, \ldots$ — вычислимая последовательность натуральных чисел, для которой*

$$\sum_i 2^{-l_i} \leqslant 1.$$

Тогда существует вычислимая последовательность попарно несравнимых двоичных слов $x_0, x_1, x_2, \ldots$, для которых $l(x_i) = l_i$.

Заметим, что указанное в лемме неравенство необходимо для существования несравнимых слов x_i с длинами l_i, поскольку отрезки I_{x_i} не перекрываются и имеют длины как раз 2^{-l_i}. Лемма утверждает, что это необходимое условие одновременно является достаточным.

Доказательство. Наглядно можно сформулировать лемму так. Есть окно шириной 1 (отрезок $[0,1]$). Нам приносят одну за другой шторы, которыми мы зана-

вешиваем окно (так, чтобы они не перекрывались). При этом шторы могут иметь ширину 1, 1/2, 1/4, 1/8, ..., и вешать их можно не куда угодно, а с ограничениями. Штору размера 1 можно повесить единственным способом. Штору размера 1/2 можно вешать двумя способами (закрыв левую половину окна или закрыв правую). Для шторы размера 1/4 имеется четыре возможных положения (четыре четверти отрезка), и так далее. Выбор положения для шторы шириной 2^{-l} соответствует выбору слова длины l. Требование, чтобы шторы не перекрывались, означает, что соответствующие им слова несравнимы. Наконец, условие леммы утверждает, что суммарная ширина штор не превосходит ширины окна.

(Более компьютерная метафора: распределение памяти между процессами; каждый процесс требует $1/2^k$ доступной памяти, и требует, чтобы выделенный участок был выровнен [aligned]; память не освобождается; мы доказываем, что если сумма всех запросов не больше общего размера памяти, то их можно удовлетворить в порядке поступления, применяя описанный далее алгоритм.)

Выбор положения для очередной шторы происходит по такому алгоритму: свободная часть окна представляется в виде объединения непересекающихся «виртуальных штор», причём размеры всех виртуальных штор различны. Виртуальная штора обозначает место, которое пока свободно, но на которое потом можно будет повесить реальную штору того же размера. (Свободная память представлена в виде списка правильно выровненных блоков разного размера.)

Вначале окно свободно, и есть одна виртуальная штора ширины 1. Пусть теперь нам приносят штору ширины w. Если среди виртуальных штор имеется штора ширины w, то всё просто: изымаем её из набора и вешаем на её место реальную. Пусть это не так. Заметим, что тогда среди виртуальных штор обязательно есть штора ширины больше w. (В самом деле, иначе в наборе были бы только виртуальные шторы $w/2, w/4, \ldots$, и общая ширина свободного пространства была бы меньше w. Напомним, что все виртуальные шторы разного размера.)

Итак, мы знаем, что есть хотя бы одна виртуальная штора ширины $w' > w$. Выберем из таких штор самую узкую (подобная стратегия при распределении памяти называется best fit). Отрежем от неё слева кусок размера w, а остальное разрежем на части $w, 2w, 4w, \ldots, w'/2$ (ведь $w + w + 2w + 4w + \ldots + (w'/2) = w'$). Эти части и будут новыми виртуальными шторами. Поскольку w' было выбрано минимальным, штор промежуточного размера (от w до $w'/2$) раньше не было, так что наше условие (требующее, чтобы виртуальные шторы были разных размеров) продолжает выполняться. Лемма доказана.

$\boxed{\textbf{100}}$ Докажите, что ту же стратегию можно описать другими словами: каждую следующую штору вешаем на самое левое из возможных мест. [Указание: описанная в лемме конструкция сохраняет такое свойство: длины виртуальных штор возрастают слева направо.]

Следствие. *Пусть l_i — вычислимая последовательность натуральных чисел, причём $\sum_i 2^{-l_i} \leqslant 1$. Тогда $KP'(i) \leqslant l_i + O(1)$.*

В самом деле, лемма позволяет указать последовательность несравнимых слов x_i длины l_i. Определим способ описания D, положив $D(x_i) = i$. Несравнимость слов x_i гарантирует, что этот способ описания будет беспрефиксным. Вычислимость

последовательности x_i гарантирует вычислимость D (имея вход x, мы сравниваем его по очереди со всеми x_i до совпадения, после чего выдаём на выход i).

(Заметим, что мы, как и раньше, свободно переходим от натуральных чисел к словам, говоря об их сложности и априорной вероятности.)

Вернёмся к доказательству теоремы. Мы рассматриваем наибольшую (максимальную) перечислимую снизу полумеру m. Согласно определению перечислимости снизу,

$$m(x) = \lim_{i\to\infty} m(x, i),$$

где $m(x, i)$ — вычислимая и монотонная по i функция двух аргументов с рациональными значениями. Округлим $m(x, i)$ в сторону увеличения до ближайшей степени двойки (то есть до одного из чисел $1, 1/2, 1/4, 1/8, \ldots$), при этом нули оставляем нулями. Назовём результат округления $m'(x, i)$. По-прежнему m' — вычислимая функция, монотонная по второму аргументу. Очевидно, $m'(x, i)$ не меньше $m(x, i)$, но превосходит $m(x, i)$ не более чем вдвое.

Назовём пару $\langle x, i\rangle$ *граничной*, если $m'(x, i)$ больше $m'(x, i-1)$ (или если $i = 0$ и $m'(x, 0) > 0$). Граничные пары отмечают те места, где запросы $m'(x, i)$ увеличиваются с ростом i (при данном x).

Покажем, что сумма всех $m'(x, i)$ по всем граничным парам $\langle x, i\rangle$ не превосходит 4. Для этого достаточно показать, что сумма $m'(x, i)$ по всем граничным i (для данного x) не превосходит $4m(x)$. А это делается так. В указанной сумме (для фиксированного x) каждый следующий член по крайней мере вдвое больше предыдущего, поэтому вся сумма не превосходит удвоенного последнего члена. А он, в свою очередь, не более чем вдвое превосходит $m(x, i)$ для какого-то i. Вспоминая, что $m(x, i) \leqslant m(x)$, заключаем, что интересующая нас сумма не больше $4m(x)$, что и требовалось.

Множество всех граничных пар $\langle x, i\rangle$ разрешимо. В самом деле, по паре можно проверить, является ли она граничной, посмотрев на рациональные числа $m'(x, i)$ и $m'(x, i-1)$.

Перенумеровав все пары в каком-то порядке и отобрав граничные, построим вычислимую последовательность $\langle x_0, i_0\rangle, \langle x_1, i_1\rangle, \ldots$ содержащую каждую граничную пару по одному разу. Рассмотрим вычислимую последовательность чисел l_n, заданных соотношением

$$2^{-l_n} = m'(x_n, i_n)/4.$$

Тогда по доказанному

$$\sum_n 2^{-l_n} = (1/4)\sum_n m'(x_n, i_n) \leqslant 1,$$

и потому $KP'(n) \leqslant l_n + O(1)$. Поскольку x_n алгоритмически получается по n, то (как мы видели в разделе про префиксную сложность)

$$KP'(x_n) \leqslant KP'(n) + O(1) \leqslant l_n + O(1) = -\log m'(x_n, i_n) + O(1).$$

Данное слово x появляется среди x_n столько раз, сколько имеется граничных пар вида $\langle x, i\rangle$. Если $\langle x_n, i_n\rangle$ — граничная пара с данным x и наибольшим i, то $m'(x_n, i_n)$

больше или равно $m(x,i)$ при всех i (поскольку округляем мы в сторону увеличения, и бóльших граничных пар нет). Поэтому в этом случае $m'(x_n, i_n) \geqslant m(x)$. Отсюда следует, что

$$KP'(x) \leqslant -\log m(x) + O(1),$$

что и требовалось доказать. ▶

Итак, три доказанных неравенства позволяют заключить, что KP, KP' и $-\log m$ отличаются не более чем на $O(1)$. Имея это в виду, мы не будем различать KP и KP' (кроме тех случаев, когда в каком-то рассуждении годится только одно из этих двух определений).

Подробное изложение доказательства могло создать впечатление его сложности, но основная идея в сущности проста. В одну сторону: наличие короткого описания p у слова x гарантирует большýю вероятность порождения слова x, так как при случайном выборе программы мы натыкаемся на p с большой вероятностью. В другую сторону немного сложнее: большая вероятность не означает короткого описания, вместо этого может быть много длинных. Однако алгоритм выделения места позволяет их консолидировать: когда у некоторого объекта x накопится описаний на 2^{-k}, надо выделить для x нетронутый отрезок длиной $\Omega(2^{-k})$, это можно делать, например, слева направо (уменьшая отрезок до правильно выровненного) — или использовать «лемму о шторах».

Аналогичные утверждения можно доказать и для релятивизованной (относительно некоторого оракула) префиксной сложности: KP^A совпадает с KP'^A и с $-\log m^A$ с точностью до $O(1)$. Было бы интересно выяснить (авторы этого не знают), что из этих равенств сохраняется для более общих видов оракулов (типа сводимости по перечислению, или точек в конструктивном метрическом пространстве).

Отметим ещё (это нам пригодится при доказательстве критерия случайности в терминах монотонной сложности, раздел 5.6), что фактически мы доказали нечто большее:

Теорема 59. *По любой перечислимой снизу последовательности чисел $p_0, p_1, p_2, \ldots$, у которой $\sum_i p_i \leqslant 1$, можно эффективно указать беспрефиксный способ описания D, для которого $KP'_D(i) \leqslant -\log_2 p_i + 2$.*

(Имеется в виду, что дан алгоритм, перечисляющий множество всех пар $\langle r, i\rangle$, у которых $r < p_i$. Описанная в доказательстве теоремы конструкция алгоритмически преобразует его в беспрефиксный способ описания с нужным свойством.)

4.6. Свойства префиксной сложности

Вернёмся теперь к свойствам префиксной сложности. Для начала покажем, как некоторые уже известные нам свойства могут быть получены с использованием априорной вероятности (универсальной полумеры).

Ряд $\sum 1/n^2$, как известно из курса анализа, сходится. Умножив его на подходящую константу, получаем перечислимую снизу полумеру (с суммой меньше

единицы). Поэтому априорная вероятность числа n не меньше c/n^2 при некотором c, то есть

$$KP(n) \leqslant 2\log n + O(1).$$

Пусть x_n — слово с порядковым номером n в последовательности Λ, 0, 1, 00, 01, 10, 11, 000, ...; тогда

$$KP(x_n) \leqslant KP(n) + O(1) \leqslant 2\log n + O(1) = 2l(x_n) + O(1)$$

(легко проверить, что x_n есть двоичная запись числа $n+1$ без старшей цифры 1 и потому имеет длину $\log n + O(1)$). (Строго говоря, случай $n = 0$ следовало бы разбирать отдельно, поскольку $1/0^2$ и $\log 0$ не определены, но необходимые исправления очевидны.) Получается оценка $KP(x) \leqslant 2l(x) + O(1)$.

Чтобы получить лучшую оценку префиксной сложности (уже встречавшуюся нам на с. 98), рассмотрим более медленно сходящийся ряд

$$\sum \frac{1}{n\log^2 n}$$

(его сходимость можно установить, например, с помощью интегрального признака сходимости и интегрирования по частям). Он даёт оценку $KP(n) \leqslant \log n + 2\log\log n + O(1)$, которая после перехода к словам превращается в

$$KP(x) \leqslant l(x) + 2\log l(x) + O(1).$$

Можно дальше улучшать оценку, воспользовавшись рядами $\sum 1/(n\log n(\log\log n)^2)$, $\sum 1/(n\log n\log\log n(\log\log\log n)^2)$ и так далее.

Докажем теперь обещанное неравенство для энтропии пары.

Теорема 60.

$$KP(x,y) \leqslant KP(x) + KP(y) + O(1).$$

◀ Под $KP(x,y)$ мы понимаем сложность слова $[x,y]$, где $\langle x,y\rangle \mapsto [x,y]$ — какое-либо вычислимое однозначное кодирование пар. (С точностью до ограниченного слагаемого сложность пары не зависит от выбора кодирования, поскольку переход от одного кодирования к другому — вычислимое преобразование.)

Рассмотрим функцию m', определённую соотношением

$$m'([x,y]) = m(x)m(y),$$

где m — априорная вероятность. (Если не всякое слово является кодом пары, полагаем $m'(z) = 0$ для слов z, не являющихся кодами.) Очевидно, m' перечислима снизу (перемножая вычислимые последовательности, сходящиеся снизу к $m(x)$ и $m(y)$, получаем возрастающую вычислимую последовательность с нужным пределом). Кроме того,

$$\sum_z m'(z) = \sum_{x,y} m'([x,y]) = \sum_{x,y} m(x)m(y) = \sum_x m(x)\sum_y m(y) \leqslant 1\cdot 1 = 1.$$

Следовательно, m' — перечислимая снизу полумера. Сравнивая её с априорной вероятностью, находим, что $m'([x,y]) \leqslant cm([x,y])$ для некоторого c, и потому

$$KP([x,y]) \leqslant KP(x) + KP(y) + O(1),$$

что и требовалось доказать. ►

$\boxed{\mathbf{101}}$ Докажите, что $\sum_y m([x,y])$ отличается от $m(x)$ не более чем на мультипликативную константу (в обе стороны). Докажите аналогичное утверждение для $\max_y m([x,y])$.

$\boxed{\mathbf{102}}$ Докажите, что для любой возрастающей вычислимой функции $f\colon \mathbb{N} \to \mathbb{N}$ величина $\sum\{m(k) \mid f(n) \leqslant k < f(n+1)\}$ отличается от $m(n)$ не более чем на мультипликативную константу (в обе стороны): каким вычислимым способом ни разбивай сходящийся ряд $\sum m(n)$ на группы, получается тот же самый (с точностью до ограниченного множителя) ряд!

Попытаемся теперь доказать теорему 60 в терминах описаний. Здесь обнаруживается такой любопытный факт: определение с префиксно корректными способами не помогает, а с беспрефиксными всё получается. Итак, мы доказываем, что $KP'([x,y]) \leqslant KP'(x) + KP'(y) + O(1)$. Пусть D — оптимальный беспрефиксный способ описания, использованный при определении KP'.

Определим новый способ описания формулой

$$D'(pq) = [D(p), D(q)],$$

где pq означает конкатенацию слов p и q. Эта формула имеет смысл, если $D(p)$ и $D(q)$ определены. Требуется, однако, проверить, что это определение корректно, то есть что одно и то же слово x не может быть представлено двумя разными способами $x = pq = p'q'$, если $D(p), D(q), D(p'), D(q')$ все определены. В самом деле, в этом случае слова p и p' согласованы (они являются началами одного и того же слова x, и потому одно из слов p и p' есть начало другого), поэтому $D(p)$ и $D(p')$ не могут быть одновременно определены при $p \neq p'$ (функция D является беспрефиксной). А раз $p = p'$, то и $q = q'$.

Аналогичным образом легко проверить, что функция D' является беспрефиксной: если pq есть начало $p'q'$, то p и p' согласованы. Раз $D(p)$ и $D(p')$ определены, то $p = p'$. Поэтому q является началом q'; раз $D(q)$ и $D(q')$ определены, то $q = q'$.

Функция D' вычислима: чтобы найти $D'(x)$, мы разбиваем x на p и q всеми возможными способами и параллельно вычисляем $D(p)$ и $D(q)$ для всех вариантов. Как мы видели, удачным может быть не более одного разбиения, и оно даёт ответ (если существует).

Остаётся заметить, что

$$KP_{D'}([x,y]) \leqslant KP_D(x) + KP_D(y).$$

В самом деле, если p и q — кратчайшие описания для x и y относительно D, то pq — описание $[x,y]$ относительно D', имеющее длину $KP_D(x) + KP_D(y)$.

Неформально машину D' можно описать так: она читает вход, пока не прочтёт там описание для x. После этого остаток входа читается как описание для y.

103 Докажите теорему 60, используя определение беспрефиксных способов описания в терминах машин с блокирующим чтением.

104 Назовём множество слов *беспрефиксным*, если никакие два слова в нём не сравнимы. Покажите, что если A и B — беспрефиксные множества слов, то и множество
$$AB = \{ab \mid a \in A, b \in B\}$$
является беспрефиксным.

В данном случае рассуждение с априорной вероятностью было, пожалуй, проще. В следующем примере это уже не так.

Теорема 61.
$$KP(x, KP(x)) = KP(x) + O(1).$$

Аналогичное равенство для обычной (не префиксной) сложности составляло содержание задачи 23.

◀ Легко видеть, что $KP(x) \leqslant KP(x, KP(x)) + O(1)$, поскольку x получается алгоритмически по $[x, KP(x)]$. (Квадратные скобки обозначают вычислимое однозначное кодирование, используемое при определении префиксной сложности пары.)

Для доказательства обратного неравенства снова используем KP', то есть беспрефиксные способы описания. Пусть D — оптимальный беспрефиксный способ описания. Рассмотрим новый способ описания
$$D'(p) = [D(p), l(p)],$$
имеющий ту же область определения, что и D, и потому беспрефиксный. Если p — кратчайшее описание слова x относительно D, то $l(p) = KP'(x)$, и потому p является описанием пары $[x, KP'(x)]$ относительно D'. Поэтому $KP_{D'}([x, KP'(x)]) \leqslant \leqslant l(p) = KP'(x)$.

Теорема доказана? На самом деле тут есть тонкость, требующая разъяснения. Мы доказали теорему для функции KP', соответствующей беспрефиксному способу описания. Если заменить её на KP (соответствующую префиксно корректному способу описания), то правая часть изменится не более чем на константу. Вопрос в том, почему это верно и для левой части, поскольку KP стоит внутри в качестве аргумента.

Чтобы ответить на этот вопрос, достаточно показать, что $KP(x, n)$ (сложность пары, в которой второй аргумент мы считаем натуральным числом) меняется не более чем на константу при изменении n на единицу. Это следует из того, что отображения $[x, n] \mapsto [x, n + 1]$ и $[x, n] \mapsto [x, n - 1]$ вычислимы и потому могут увеличивать сложность не более чем на константу. ▶

Для сравнения проведём также и доказательство с помощью априорной вероятности. Пусть $m(x)$ — априорная вероятность слова x. Рассмотрим функцию m', определённую так:
$$m'([x, k]) = \begin{cases} 2^{-k}, & \text{если } 2^{-k} < m(x), \\ 0 & \text{в противном случае.} \end{cases}$$

Эта функция перечислима снизу (для данных x и k мы выдаём нуль, пока не выяснится, что $2^{-k} < m(x)$, а потом выдаём 2^{-k}).

Сумма $m'([x, k])$ по всем k (при данном x) представляет собой геометрическую прогрессию со знаменателем 2, и потому не больше $2m(x)$ (последний член прогрессии меньше $m(x)$). Следовательно, сумма $m'([x, k])$ по всем x и k конечна. Сравнивая m' с априорной вероятностью и переходя к логарифмам, находим, что

$$KP(x, k) \leqslant k + O(1)$$

при $2^{-k} < m(x)$. Теперь положим $k = -\lfloor \log m(x) \rfloor + 1$, тогда $2^{-k} < m(x)$ и получается, что

$$KP(x, -\lfloor \log m(x) \rfloor + 1) \leqslant KP(x) + O(1).$$

Остаётся (как и раньше) заметить, что изменение второго члена пары на единицу меняет сложность не более чем на константу, и второе доказательство теоремы завершено.

105 Приведённое только что доказательство теоремы 61 фактически доказывает чуть больше: $KP(x, u) \leqslant u + O(1)$ при $KP(x) \leqslant u$. Как вывести это из утверждения теоремы 61, не апеллируя к доказательству?

Что можно сказать об алгоритмических свойствах функции $KP(x)$? Как и обычная колмогоровская сложность, она перечислима сверху, но не вычислима и даже не имеет никакой нетривиальной (неограниченной) вычислимой нижней оценки. (Поскольку $KP(x) \leqslant 2\,KS(x) + O(1)$, любая нетривиальная нижняя оценка для KP немедленно давала бы нетривиальную нижнюю оценку для KS.)

Ранее (теорема 8, с. 29) мы видели, что $KS(x)$ можно определить как минимальную перечислимую сверху функцию K, для которой множества $\{x \mid K(x) < n\}$ имеют размер $O(2^n)$. Аналогичное утверждение для префиксной сложности выглядит так:

Теорема 62. *KP есть минимальная (с точностью до $O(1)$) перечислимая сверху функция K, для которой сумма ряда $\sum_x 2^{-K(x)}$ конечна.*

◄ Как мы видели, функция KP перечислима сверху и ряд $\sum_x 2^{-KP(x)}$ сходится. Если K — перечислимая сверху функция с теми же свойствами, то функция $M(x) = c2^{-K(x)}$ при достаточно малом рациональном c является полумерой (она перечислима снизу и сумма меньше 1). Следовательно, $M(x) = O(m(x))$ в силу максимальности априорной вероятности, и потому $\log M(x) \leqslant \log m(x) + O(1)$, то есть $KP(x) \leqslant K(x) + O(1)$. ►

Переформулировка: пусть функция f с натуральными значениями определена на всех словах и перечислима сверху. Тогда свойства «$KP(x) \leqslant f(x) + O(1)$» и $\sum_x 2^{-f(x)} < \infty$ равносильны.

Конечность суммы ряда в только что доказанной теореме является более сильным условием на функцию K, чем использованное в теореме 8 условие. В самом деле, если сумма ряда не больше C, то количество членов ряда, бóльших некоторой границы ε, не превосходит C/ε. Тем самым мы заново получаем, что $KS(x) \leqslant KP(x) + O(1)$.

Поучительно сравнить обычную и префиксную сложности по двум параметрам: средней сложности слов данной длины и числу слов сложности не больше данной. Начнём с первого.

Мы видели, что большинство слов длины n имеют (обычную) сложность $n + O(1)$ (см с. 17, а также задачу 2 на с. 26). Естественно ожидать, что префиксная сложность будет несколько больше.

Теорема 63. (а) $KP(x) \leqslant l(x) + KP(l(x)) + O(1)$.

(б) *Найдётся такая константа c, что для любого n и для любого d доля слов x длины n, у которых $KP(x) < n + KP(n) - d$, не превосходит $c2^{-d}$.*

◀ (а) Пусть $m(x)$ — априорная вероятность слова x. Мы будем также говорить об априорной вероятности $m(n)$ натурального числа n (или, что то же самое, его двоичной записи). Рассмотрим функцию $m'(x)$, равную $2^{-n}m(n)$ для слов x длины n. Значения этой функции на всех словах длины n одинаковы и в сумме дают $m(n)$, поэтому $\sum_x m'(x) \leqslant 1$. Функция m' перечислима снизу. Поскольку полумера m является наибольшей, заключаем, что $m'(x) \leqslant cm(x)$ для некоторого c и для всех x. Переходя к логарифмам, получаем неравенство

$$KP(x) \leqslant n + KP(n) + O(1)$$

для слов длины n (константа в $O(1)$ не зависит от n).

(б) Теперь будем двигаться в обратном направлении и рассмотрим функцию

$$m'(n) = \sum_{l(x)=n} m(x)$$

(сумму априорных вероятностей всех слов данной длины). Поскольку эта функция перечислима снизу и $\sum_n m'(n) \leqslant 1$, то $m'(n) = O(m(n))$ (где $m(n)$ — априорная вероятность натурального числа n). С другой стороны, слово из n нулей имеет (с точностью до константы) ту же априорную вероятность, что и n, так что

$$c_1 m(n) \leqslant \sum_{l(x)=n} m(x) \leqslant c_2 m(n).$$

Другими словами, сумма значений функции $m(x)$ на словах длины n с точностью до константы совпадает с $m(n)$, а среднее значение $m(x)$ на этих словах — с $m(n)/2^n$. Остаётся заметить, что доля тех аргументов, для которых функция в 2^d раз превышает своё среднее, не превосходит 2^{-d} (неравенство Чебышёва). ▶

$\boxed{\textbf{106}}$ Используя эти оценки, докажите, что среднее арифметическое префиксных сложностей всех слов длины n равно $n + KP(n) + O(1)$.

(Для обычной сложности это сделано в задаче 3, с. 26.)

Теперь оценим число слов данной сложности.

Теорема 64. *Логарифм числа слов x, для которых $KP(x) < n$, равен $n - KP(n) + O(1)$.*

◀ Пусть c_n — число слов, для которых $KP(x) < n$. Перепишем основное свойство префиксной сложности, сходимость ряда $\sum 2^{-KP(x)}$, в терминах чисел c_n. Имеется $c_{n+1} - c_n$ слов, сложность которых равна в точности n, поэтому сумму ряда можно переписать как $\sum_n 2^{-n}(c_{n+1} - c_n)$. Перегруппировав члены, убеждаемся, что сумма $\sum_n (2^{-(n-1)} - 2^{-n})c_n = \sum_n 2^{-n} c_n$ конечна. Поскольку c_n перечислимо снизу, то отсюда следует, что $2^{-n} c_n$ не превосходит априорной вероятности $m(n)$ числа n, откуда $c_n \leqslant m(n)2^n$ (с точностью до мультипликативной константы).

С другой стороны, легко построить перечислимую сверху функцию K, аргументами которой являются двоичные слова, а значениями натуральные числа и плюс бесконечность, у которой будет примерно $m(n)2^n$ значений, равных n. Например, можно договориться, что на словах длины n эта функция равна либо n, либо $+\infty$, причём изначально все значения бесконечны, а затем по мере появления всё бóльших оценок для $m(n)$ нужное число бесконечных значений заменяется на n.

Для этой функции сумма $\sum 2^{-K(x)}$ сходится, поэтому $KP(x) \leqslant K(x) + O(1)$, и потому $c_{n+O(1)} \geqslant m(n)2^n$. Остаётся вспомнить, что $m(n)$ и 2^n меняются не более чем в константу раз при изменении n на единицу. ▶

Эти результаты могут создать впечатление, что префиксная и обычная меры сложности отличаются всего лишь выбором шкалы (шкала для префиксной сложности сдвинута относительно обычной, так что префиксная сложность немного больше). Так ли это? Этот вопрос можно естественным образом уточнить: существуют ли две последовательности слов a_n и b_n, для которых $KS(a_n) - KS(b_n) \to +\infty$, но $KP(a_n) - KP(b_n) \to -\infty$? На этот вопрос ответили Ан. А. Мучник и С. Е. Посицельский, доказав существование таких последовательностей в [118]. Другое доказательство дал Миллер (см. [104], где обсуждаются и другие результаты о связи простой и префиксной сложности). Мы ограничимся лишь несколькими простыми замечаниями (см. также раздел 4.7.4, р. 130).

Неравенство $KP(x) \leqslant l(x) + KP(l(x))$ можно итерировать, получая

$$KP(x) \leqslant l(x) + l(l(x)) + KP(l(l(x)) + O(1),$$
$$KP(x) \leqslant l(x) + l(l(x)) + l(l(l(x))) + KP(l(l(l(x))) + O(1)$$

и так далее. Кроме того, можно вспомнить, что если D — оптимальный способ описания для обычной (не префиксной) сложности, то $KP(D(y)) \leqslant KP(y) + O(1)$; сочетая это с неравенством $KP(x) \leqslant l(x) + KP(l(x)) + O(1)$, получаем следующее утверждение:

Теорема 65.

$$KP(x) \leqslant KS(x) + KP(KS(x))) + O(1),$$
$$KP(x) \leqslant KS(x) + KS(KS(x))) + KP(KS(KS(x)) + O(1)$$

и так далее.

Отметим кстати, что все эти неравенства получаются также повторным применением первого из них.

107 Докажите, что

$$KS(x, y) \leqslant KP(x) + KS(y) + O(1)$$

для любых слов x, y. [Указание. Можно подсчитать число пар, где правая часть меньше n, но проще воспользоваться беспрефиксным кодированием.]

Мы говорили, что случайными словами длины n естественно считать те, у которых (простая) колмогоровская сложность максимальна, то есть близка к n. А что будет, если потребовать, чтобы префиксная сложность была максимальна, то есть близка к $n + KP(n)$? Следующая задача показывает, что из этого следует случайность в прежнем смысле.

108 Покажите, что если обычная сложность слова x длины n не превосходит $n - d$, то его префиксная сложность не превосходит $n + KP(n) - d + O(\log d)$. [Указание: соединим беспрефиксные описания для n и d с обычным описанием для x.]

Аналогичное утверждение в обратную сторону неверно, как показал Р. Соловей (см. упомянутую статью Миллера [104] или [117]).

4.7. Условная префиксная сложность и сложность пары

4.7.1. Условная префиксная сложность

Как определить префиксный вариант условной сложности? Каждое из приведённых нами определений можно модифицировать, добавив туда условия. Вот что при этом получится.

Функция D двух аргументов называется *префиксно корректной относительно первого аргумента*, если при любом фиксированном значении второго аргумента получается префиксно корректная функция:

$$D(y, z) \text{ определено и } y \leqslant y' \Rightarrow D(y', z) = D(y, z).$$

Здесь предполагается, что первый аргумент является двоичным словом; $y \leqslant y'$ означает, что y' является продолжением y.

При определении условной сложности в разделе 2.2 мы рассматривали всевозможные вычислимые функции двух аргументов, называя их *способами условного описания.* (В качестве аргументов и значений рассматривались двоичные слова.) Равенство $D(y, z) = x$ читалось так: y является описанием x при известном z. Сложность x при известном z определялась как длина кратчайшего описания. Среди всех способов описания выбирался оптимальный, при котором сложность меньше всего (с точностью до константы).

Ограничимся теперь префиксно корректными по первому аргументу способами описания. Среди них тоже существует оптимальный. Доказательство этого факта аналогично доказательству теоремы 48 на с. 95 (о существовании оптимального префиксно корректного способа безусловного описания), только все приведённые

там рассуждения надо параллельно провести для всех условий z. Именно, надо положить

$$D'(\hat{p}y, z) = [p](y, z),$$

где $[p](y, z)$ означает результат принудительной префиксной коррекции программы p (как функции от y при каждом z в отдельности).

Выбрав оптимальный префиксно корректный способ условного описания, сложность относительно него мы называем условной префиксной сложностью и обозначаем её $KP(x\,|\,z)$ (читается: «условная префиксная сложность слова x при известном слове z).

Это определение можно видоизменить, заменив требование префиксной корректности на требование беспрефиксности по первому аргументу. И для этого класса существует оптимальный способ условного описания; соответствующую сложность будем обозначать $KP'(x\,|\,z)$. Можно доказать, что эти определения различаются не более чем на константу:

$$KP'(x\,|\,z) = KP(x\,|\,z) + O(1),$$

причём константа эта не зависит не только от x, но и от z.

Как и для безусловного случая, прямое доказательство этого равенства затруднительно, и в качестве промежуточного звена используется условная априорная вероятность $m(x\,|\,z)$. Её (как и раньше) можно определить двумя способами. Первый способ использует вероятностные машины с входом. Пусть M — такая машина. Для каждого входного слова z возникает своё распределение вероятностей: выход x (при входе z) имеет некоторую вероятность $p_M(x\,|\,z)$. Функция $\langle x, z\rangle \mapsto p_M(x\,|\,z)$ перечислима снизу (в естественном смысле), и для каждого z сумма $\sum_x p_M(x\,|\,z)$ не превосходит единицы. Эти свойства описывают класс всех функций p_M: если имеется перечислимая снизу неотрицательная функция $\langle x, z\rangle \mapsto p(x\,|\,z)$, для которой $\sum_x p(x\,|\,z) \leqslant 1$ при всех z, то можно построить вероятностную машину M, для которой $p_M = p$.

Среди всех функций p_M существует наибольшая с точностью до умножения на константу. Мы фиксируем одну из таких наибольших функций, называем её *условной априорной вероятностью слова x при известном слове z* и обозначаем $m(x\,|\,z)$.

Далее мы доказываем, что $-\log m(x\,|\,z) \leqslant KP(x\,|\,z) + O(1)$ и что $KP'(x\,|\,z) \leqslant$ $\leqslant -\log m(x\,|\,z) + O(1)$. Неравенство $KP(x\,|\,z) \leqslant KP'(x\,|\,z) + O(1)$, которое доказывается столь же просто, как и в безусловном случае, замыкает круг: все три величины $KP(x\,|\,z)$, $KP'(x\,|\,z)$ и $-\log m(x\,|\,z)$ отличаются не более чем на константу.

Мы не приводим подробно всех доказательств, поскольку они состоят в параллельном проведении «безусловных» доказательств для всех условий z. Можно сказать, что мы имеем дело с «релятивизацией», но несколько особого вида.

Поясним сказанное. В общей теории алгоритмов «релятивизация» означает, что вместо вычислимых функций мы рассматриваем функции, вычислимые относительно некоторого оракула A. (Здесь A — произвольное множество слов. Считается, что алгоритм, отвечающий на вопросы о принадлежности слова x множеству A,

доступен как внешняя процедура с входным параметром x.) Практически все известные теоремы общей теории алгоритмов без труда переносятся и на такие («A-вычислимые») функции.

Кстати, и понятие сложности можно релятивизовать таким образом, определив для любого множества A колмогоровскую сложность $KS^A(x)$ и префиксную колмогоровскую сложность $KP^A(x)$ (см. раздел 6.4). Но сейчас мы делаем не это: в качестве оракула мы рассматриваем доступ к конечному слову z. Конечно, такой оракул не может изменить понятие вычислимости (слово z конечно), но он меняет колмогоровскую сложность, поскольку информация, имеющаяся в z, «выносится за скобки» и не учитывается при подсчёте сложности. В результате вместо обычной колмогоровской сложности $KS(x)$ мы получаем условную сложность $KS(x\,|\,z)$. Другой пример: величину $I(x:y\,|\,z)$ можно считать релятивизованной относительно z версией понятия взаимной информации $I(x:y)$.

Ещё одно важное замечание. Во всех наших рассмотрениях какие бы то ни было требования, связанные с префиксами (началами), относятся только к описаниям. Напротив, и описываемые объекты, и условия для нас лишены какой бы то ни было структуры.

Такой подход не является единственно возможным. Можно учитывать отношение «быть началом» для описываемых объектов (так делают при определении монотонной сложности и сложности разрешения, см. главы 5 и 6). Можно было бы также учитывать отношение «быть началом» и для условий (см. раздел 6.3); такие варианты определений вполне осмысленны, хотя и мало изучены.

Отметим также, что при определении условной префиксной сложности требования, связанные с префиксами, накладываются отдельно при каждом значении условия (второго аргумента функции). Например, мы требуем, чтобы машина сама решала, когда она кончит читать описание — но это решение может зависеть от второго аргумента. Из-за этого утверждение, аналогичное задаче 28 (с. 45), уже не имеет места:

109 Покажите, что $KP(y\,|\,x)$ не превосходит минимальной префиксной сложности программ, переводящих x в y, но обратное неверно (все неравенства понимаются с точностью до $O(1)$). (Оба утверждения верны для любого разумного языка программирования; аддитивная константа может зависеть от выбора языка программирования.) [Указание. Легко видеть, что $KP(y\,|\,l(y)) \leqslant l(y) + O(1)$, так как при известном $l(y)$ каждое слово можно считать своей самоограниченной программой. Если бы обратное утверждение было верно, то отсюда мы получили бы 2^n различных программ префиксной сложности не более n.]

4.7.2. Свойства условной префиксной сложности

Перечислим несколько простых свойств условной префиксной сложности.

- $KP(x\,|\,z) \leqslant KP(x) + O(1)$.

  В самом деле, любой префиксно корректный (или беспрефиксный) способ безусловного описания $y \mapsto D(y)$ можно рассматривать как префиксно кор-

ректный (соответственно беспрефиксный) способ условного описания $\langle y, z\rangle \mapsto \mapsto D(y)$ (второй аргумент фиктивен).

В терминах полумер: всякая вероятностная машина без входа может рассматриваться как вероятностная машина со входом, игнорирующая свой вход. Соответственно любая перечислимая снизу полумера $q(x)$ может рассматриваться как семейство полумер $q'(x\,|\,z) = q(x)$ с фиктивным параметром z.

- $KP(x\,|\,x) = O(1)$.

  Способ описания $D(y, z) = z$ является префиксно корректным (нас интересует префиксная корректность по y, а не по z) и $KP_D(x\,|\,x) = 0$. Беспрефиксный способ описания можно определить так: $D(\Lambda, z) = z$, где Λ — пустое слово; для непустых первых аргументов D не определён. Наконец, перечислимое снизу семейство полумер таково: $q(x\,|\,x) = 1$ и $q(x\,|\,z) = 0$ при $z \neq x$.

- $KP(f(x,z)\,|\,z) \leqslant KP(x\,|\,z) + O(1)$ для любой вычислимой функции f и для любых слов x, z, на которых f определена. (При этом константа в $O(1)$ может зависеть от f, но не от x и z.)

  Если D — оптимальный префиксно корректный [беспрефиксный] способ условного описания, то D': $\langle y, z\rangle \mapsto f(D(y,z),z)$ также будет префиксно корректным [соответственно беспрефиксным] и $KP_{D'}(f(x,z)\,|\,z) \leqslant KP_D(x\,|\,z)$, откуда и следует искомое неравенство.

  В терминах полумер: если $m(x\,|\,z)$ — априорная вероятность x при известном z, рассмотрим полумеру

  $$q(x\,|\,z) = \sum\{m(x'\,|\,z) \mid f(x', z) = x\}$$

  (при каждом z она представляет собой прямой образ полумеры $x \mapsto m(x, z)$ при отображении $x \mapsto f(x, z)$); легко проверить, что функция q перечислима снизу, что $\sum_x q(x\,|\,z) \leqslant 1$ и что $q(f(x,z)\,|\,z) \geqslant m(x\,|\,z)$. Используя оптимальность m, приходим к искомому неравенству (для логарифмов априорных вероятностей).

- $KP(f(x)\,|\,x) = O(1)$ для любой вычислимой функции f и для всех x, при которых $f(x)$ определено.

  (Простое следствие предыдущих.)

- $KP(x\,|\,z) \leqslant KP(x\,|\,f(z)) + O(1)$ для любой вычислимой функции f и любых x, z (если $f(z)$ определено; константа может зависеть от f, но не от x и z).

  (Достаточно рассмотреть способ описания $\langle y, z\rangle \mapsto D(y, f(z))$ или семейство полумер $q(x\,|\,z) = m(x\,|\,f(z))$.)

- $KS(x\,|\,z) \leqslant KP(x\,|\,z) + O(1)$

  В самом деле, префиксно корректные (или беспрефиксные) способы условного описания входят в число способов условного описания, рассматриваемых при определении $KS(x\,|\,z)$.

- $KP(x\,|\,z) \leqslant KS(x\,|\,z) + 2\log KS(x\,|\,z) + O(1)$

  Это утверждение можно вывести из предыдущих. Пусть D — оптимальный способ условного описания (не обязательно префиксно корректный). Тогда

  $$KP(D(y,z)\,|\,z) \leqslant KP(y\,|\,z) + O(1) \leqslant KP(y) + O(1) \leqslant l(y) + 2\log l(y) + O(1).$$

Если в качестве y взято кратчайшее описание x при известном z, то $l(y) = KS(x\,|\,z)$.

Аналогичным образом можно доказать и более сильное неравенство

$$KP(x\,|\,z) \leqslant KS(x\,|\,z) + \log KS(x\,|\,z) + 2\log\log KS(x\,|\,z) + O(1)$$

и так далее.

4.7.3. Префиксная сложность пары

Мы видели (теорема 60, с. 113), что $KP(x,y) \leqslant KP(x) + KP(y) + O(1)$. Это неравенство можно усилить:

Теорема 66.

$$KP(x,y) \leqslant KP(x) + KP(y\,|\,x) + O(1).$$

◀ Есть два варианта доказательства: с беспрефиксными способами описания и с полумерами. С беспрефиксными способами описания мы рассуждаем так. Пусть D — оптимальный беспрефиксный способ безусловного описания, а D_c — оптимальный беспрефиксный способ условного описания. Рассмотрим способ описания D', определяемый формулой

$$D'(uv) = [D(u), D_c(v, D(u))]$$

(для тех u и v, для которых правая часть определена). Как и при доказательстве теоремы 60, легко проверить, что D' корректно определён и является беспрефиксным способом описания; при этом конкатенация кратчайших описаний для x и для y (при известном x) даст описание для $[x,y]$.

(Заметим, что в этом рассуждении существен порядок u и v: если бы мы взяли vu (вместо uv), то сначала надо было бы выделять слово v, и возникла бы трудность: функция D_c является беспрефиксной при фиксированном втором аргументе, а $D(u)$ пока ещё не известно.)

С полумерами: пусть $m(x)$ — априорная вероятность x, а $m(y\,|\,x)$ — априорная вероятность y при известном x. Рассмотрим функцию m', определённую соотношением

$$m'([x,y]) = m(x)m(y\,|\,x)$$

(мы считаем, что $m'(z) = 0$ для слов z, не являющихся кодами пар). Тогда m' перечислима снизу (как произведение двух перечислимых снизу функций), и

$$\sum_z m'(z) = \sum_{x,y} m(x)m(y\,|\,x) = \sum_x \Big[m(x)\sum_y m(y\,|\,x)\Big] \leqslant \sum_x m(x) \leqslant 1.$$

Поэтому $m([x,y]) \geqslant \varepsilon m'([x,y]) = \varepsilon m(x)m(y\,|\,x)$, что и требовалось доказать. ▶

110 Докажите, что $KS(x,y) \leqslant KP(x) + KS(y\,|\,x) + O(1)$.

[Указание. Можно использовать беспрефиксные способы описания, дописав к беспрефиксному описанию для x обычное описание для y при известном x. А можно подсчитать число пар, при которых $KP(x) + KS(y\,|\,x) \leqslant n$. Если первое слагаемое равно k, то пар не больше $2^k \cdot m(k) \cdot 2^{n-k} = 2^n m(k)$, и после суммирования по k получаем $2^n \cdot O(1)$.]

Только что доказанное неравенство можно несколько усилить. Для начала заметим, что мы можем рассматривать в качестве условий не только слова, но и пары слов (поскольку замена одного способа кодирования пар на другой меняет сложность не более чем на константу). Кроме того, можно говорить о сложности троек слов. После этих замечаний можно записать такую цепочку неравенств (для краткости члены $O(1)$ опускаем):

$$KP(x,y) \leqslant KP(x, KP(x), y) \leqslant KP(x, KP(x)) + KP(y \mid x, KP(x)) = \\ = KP(x) + KP(y \mid x, KP(x)).$$

При этом мы использовали равенство $KP(x, KP(x)) = KP(x)$ (теорема 61), а также только что доказанную теорему про энтропию пары. Полученное неравенство усиливает эту теорему, поскольку $KP(y \mid x, KP(x)) \leqslant KP(y \mid x)$ (x получается алгоритмически по $[x, KP(x)]$). Как обнаружили Л. А. Левин (см. [43]) и Г. Чейтин [23], такое уточнение замечательным образом превращает неравенство в равенство:

Теорема 67.

$$KP(x,y) = KP(x) + KP(y \mid x, KP(x)) + O(1).$$

◀ В одну сторону неравенство нами уже доказано (см. обсуждение перед формулировкой теоремы). Можно привести и прямое доказательство: чтобы получить беспрефиксный код пары $\langle x, y\rangle$, достаточно к кратчайшему беспрефиксному коду x приписать справа беспрефиксный код y при известных x и $KP(x)$ (причём в качестве $KP(x)$ надо взять как раз длину написанного беспрефиксного кода для x). Заметим, что после выделения первой части мы знаем не только x, но и $KP(x)$, так что есть всё необходимое для восстановления слова y.

В терминах полумер это доказательство можно изложить так. Рассмотрим функцию m', для которой

$$m'([x,y]) = \sum_{\{k \mid 2^{-k} < 2m(x)\}} 2^{-k} m(y \mid x, k).$$

Эта функция перечислима снизу, сумма по всем x, y конечна (сумма $m(y \mid x, k)$ по y не больше 1; сумма 2^{-k} по всем k, для которых $2^{-k} < 2m(x)$, не больше $4m(x)$, и суммирование по всем x даёт не больше 4). Остаётся сравнить её с априорной вероятностью m и заметить, что при $k = -\lfloor \log_2 m(x) \rfloor$ получаем как раз нужный член суммы.

Перейдём теперь к доказательству обратного неравенства:

$$KP(x) + KP(y \mid x, KP(x)) \leqslant KP(x,y) + O(1).$$

Для начала приведём (неправильное, но более простое) доказательство (неправильного, но более сильного) утверждения

$$KP(x) + KP(y \mid x) \leqslant KP(x,y) + O(1).$$

Переходя к полумерам, можно переписать это неравенство так:

$$m(x)m(y\,|\,x) \geqslant \varepsilon\, m([x,y])$$

(для некоторого ε и для всех x, y). Здесь буквой m обозначены априорные вероятности (условная и безусловная). Перепишем это неравенство как

$$m(y\,|\,x) \geqslant \varepsilon\, \frac{m([x,y])}{m(x)}.$$

Достаточно доказать, что функция

$$m'(y\,|\,x) = \varepsilon\, \frac{m([x,y])}{m(x)}$$

при любом фиксированном x является полумерой (при некотором ε), после чего сослаться на максимальность функции $m(y\,|\,x)$. Просуммируем $m'(y\,|\,x)$ по y. Нам надо убедиться, что сумма

$$\sum_y m'(y\,|\,x) = \varepsilon\, \frac{\sum_y m([x,y])}{m(x)}$$

не превосходит единицы.

Почему это так? В самом деле, функция $x \mapsto \sum_y m([x,y])$ является полумерой (её сумма по всем x есть $\sum_{x,y} m([x,y]) \leqslant 1$) и потому эта функция не больше $m(x)/\varepsilon$ при некотором ε.

Где ошибка в этом рассуждении? Мы забыли, что полумера должна быть перечислима снизу. В одном из двух случаев это действительно так: функция $\sum_y m([x,y])$, как легко понять, будет перечислимой снизу, поскольку функция m была таковой. Но вот про $m([x,y])/m(x)$ этого сказать уже нельзя, поскольку $m(x)$ стоит в знаменателе, а при возрастании знаменателя дробь не увеличивается (как нам бы хотелось), а уменьшается.

Правильное доказательство более слабого неравенства следует той же схеме, но обходит указанную только что трудность. Мы должны доказать, что при $z = KP(x)$ имеет место неравенство

$$m(y\,|\,x,z) \geqslant \varepsilon\, \frac{m([x,y])}{m(x)}.$$

Наша проблема, напомним, в том, что правая часть не является перечислимой снизу. Но при $z = KP(x)$ можно заменить $m(x) \approx 2^{-KP(x)}$ на 2^{-z} и рассмотреть функцию

$$m'(y\,|\,x,z) = m([x,y])2^z.$$

Эта функция перечислима снизу. Зато она может не быть полумерой, поскольку сумма $\sum_y m'(y\,|\,x,z)$ не всегда будет ограничена единицей. Это будет так, лишь если

$$\sum_y m([x,y]) \leqslant 2^{-z}.$$

Мы видели, что $\sum_y m([x,y]) = O(m(x)) = O(2^{-KP(x)})$, поэтому для некоторой константы c выполнено такое свойство:

$$z \leqslant KP(x) - c \;\Rightarrow\; \sum_y m'(y\,|\,x,z) \leqslant 1.$$

Это хорошо, но мало: нам нужно семейство полумер, в котором такое неравенство выполняется при всех x и z, а не только при некоторых. Поэтому мы исправим функцию m', получив функцию m'' с такими свойствами:

- функция $\langle y,x,z\rangle \mapsto m''(y\,|\,x,z)$ перечислима снизу;
- неравенство
$$\sum_y m''(y\,|\,x,z) \leqslant 1$$
выполнено при всех x и z;
- для некоторой константы c
$$z \leqslant KP(x) - c \;\Rightarrow\; m''(y\,|\,x,z) = m'(y\,|\,x,z).$$

Технология такого рода исправления уже обсуждалась в разделе 4.2: строя всё бо́льшие приближения снизу, мы проверяем, не нарушают ли они границу для суммы, и если нарушают, то отбрасываем.

Сравнивая m'' с априорной условной вероятностью и переходя к логарифмам, мы видим, что

$$z \leqslant KP(x) - c \;\Rightarrow\; KP(y\,|\,x,z) \leqslant KP(x,y) - z + c'$$

для некоторых констант c и c' и для всех x, y, z.

Чтобы вывести отсюда утверждение теоремы, подставим $z = KP(x) - c$ в полученное неравенство и заметим, что при изменении z на единицу значение $KP(y\,|\,x,z)$ меняется не более чем на константу (функция прибавления единицы ко второму члену пары вычислима, аналогично для вычитания), поэтому

$$KP(y\,|\,x, KP(x) - c) = KP(y\,|\,x, KP(x)) + O(1). \qquad \blacktriangleright$$

Заметим, что теорема 22 (с. 49), утверждающая, что

$$KS(x,y) = KS(x) + KS(y\,|\,x) + O(\log n)$$

для слов сложности не больше n, является следствием только что доказанной.

В самом деле, замена KP на KS меняет все три члена в равенстве на более чем на $O(\log n)$. Остаётся заметить, что $KS(y\,|\,x, KP(x))$ отличается от $KS(y\,|\,x)$ не более чем на $O(\log n)$. Тем самым мы получили новое доказательство теоремы 22, в которой комбинаторные подсчёты заменены оценками вероятностей.

Вспоминая, что $m(x) \approx \sum_y m([x,y])$ с точностью до ограниченного множителя (задача 101, с. 114), можно переписать утверждение теоремы 67 так:

$$m(y\,|\,x, KP(x)) \approx \frac{m([x,y])}{\sum_y m([x,y])}.$$

Правую часть этого равенства можно интерпретировать как условную вероятность того, что второй член пары равен y, при условии «первый член пары равен x».

111 Докажите, что

$$KP(x\,|\,z) \leqslant KP(x\,|\,y) + KP(y\,|\,z) + O(1)$$

для любых слов x, y, z. (Это неравенство можно ещё усилить, заменить $KP(x\,|\,y)$ на меньшую величину $KP(x\,|\,y, z)$.)

112 Докажите «релятивизованный» вариант теоремы 67:

$$KP(x, y\,|\,z) = KP(x\,|\,z) + KP(y\,|\,x, KP(x\,|\,z), z) + O(1).$$

Дважды применяя теорему 67, можно получить формулу для префиксной сложности тройки слов. Эту тройку можно рассматривать как пару, первый член которой есть пара $\langle x, y\rangle$, а второй — z. Получаем, что

$$KP(x, y, z) = KP(z\,|\,x, y, KP(x, y)) + KP(x, y) + O(1).$$

Ещё раз применяя теорему 67, получаем ответ:

Теорема 68.

$$KP(x, y, z) = KP(z\,|\,x, y, KP(x, y)) + KP(y\,|\,x, KP(x)) + KP(x) + O(1).$$

Можно изменить порядок действий (применяя на втором шаге «релятивизованный» относительно z вариант теоремы 67):

$$\begin{aligned}KP(x, y, z) &= KP(y, z\,|\,x, KP(x)) + KP(x) = \\ &= KP(z\,|\,y, KP(y\,|\,x, KP(x)), x, KP(x)) + KP(y\,|\,x, KP(x)) + KP(x)\end{aligned}$$

(для краткости мы опускаем слагаемые $O(1)$).

Получилась немного другая формула по сравнению с утверждением теоремы 68: два последних слагаемых в ней те же, но первое слагаемое изменилось. Как и раньше, там стоит условная сложность слова z, но вместо условия $KP(x, y)$ появилось два условия $KP(x)$ и $KP(y\,|\,x, KP(x))$ — две сложности, которые в сумме дают как раз $KP(x, y)$ по теореме 67. Поэтому пара этих сложностей содержит не меньше информации, чем $KP(x, y)$. Более удивительно, что (при известных x и y) верно и обратное. В самом деле, подставим в качестве z пару $\langle KP(x), KP(y\,|\,x, KP(x))\rangle$; во второй формуле первое слагаемое обратится в нуль (точнее, в $O(1)$). Приравняв правые части формул, получаем такое следствие:

Теорема 69.

$$\begin{gathered}KP(KP(x)\,|\,x, y, KP(x, y)) = O(1),\\ KP(KP(y\,|\,x, KP(x))\,|\,x, y, KP(x, y)) = O(1).\end{gathered}$$

(Разумеется, симметричные утверждения верны для $KP(y)$ и $KP(x\,|\,y, KP(y))$.)

113 Дайте прямое доказательство теоремы 69. [Указание. Зная x, y и $KP(x,y)$, мы можем искать верхнюю оценку d для $KP(x)$, для которой $KP(y\,|\,x,d)+d$ сравняется с $KP(x,y)$. Совпадение с точностью до $O(1)$ будет означать, что $d = KP(x)+O(1)$: если d превышает $KP(x)$ на некоторое число m, то $KP(y\,|\,x,d)$ может упасть за счёт этого не более чем на $O(\log m)$, поэтому в сумме будет проигрыш.]

С помощью теоремы 67 легко показать, что базисное неравенство (теорема 24, с. 58) для префиксной сложности выполняется с точностью $O(1)$ (для обычной сложности была лишь логарифмическая оценка):

Теорема 70.

$$KP(x,y,z)+KP(x)\leqslant KP(x,y)+KP(x,z)+O(1)$$

для любых трёх слов x, y, z.

◀ В самом деле, правую часть можно переписать как

$$KP(x)+KP(y\,|\,x,KP(x))+KP(x)+KP(z\,|\,x,KP(x)),$$

а левую — как

$$KP(x)+KP(y,z\,|\,x,KP(x))+KP(x),$$

и остаётся доказать, что

$$KP(y,z\,|\,x,KP(x))\leqslant KP(y\,|\,x,KP(x))+KP(z\,|\,x,KP(x)),$$

что есть релятивизованный вариант теоремы 60 (с. 113). ▶

Приведём также прямое доказательство теоремы 70 с помощью полумер. Нам нужно доказать, что (с точностью до ограниченных множителей)

$$m(x,y,z)m(x)\geqslant m(x,y)m(x,z),$$

где m — наибольшая перечислимая снизу полумера. Поделив на $m(x)$, получим неравенство

$$\frac{m(x,y)m(x,z)}{m(x)}\leqslant m(x,y,z).$$

Проверим, что его левая часть имеет конечную сумму (по всем тройкам слов x, y, z). Это следует из того, что

$$\sum_{y,z}\frac{m(x,y)m(x,z)}{m(x)}\leqslant m(x)$$

(ведь $\sum_y m(x,y)\leqslant m(x)$ и $\sum_z m(x,z)\leqslant m(x)$). (Для краткости мы опускаем множители $O(1)$, как если бы они были равны единице.)

Этого, однако, недостаточно: из-за $m(x)$ в знаменателе дробь

$$\frac{m(x,y)m(x,z)}{m(x)}$$

может не быть перечислимой снизу, и потому мы не можем воспользоваться свойством максимальности. Применим такой обходной манёвр (уже использованный нами при доказательстве теоремы 67): построим перечислимую снизу верхнюю оценку для этой дроби.

Это делается так: для каждого натурального n через $m_n(x,y)$ обозначим функцию $m(x,y)$, у которой для каждого x сумма $\sum_y m(x,y)$ принудительно ограничена сверху числом 2^{-n}. Заметим, что $\sum_y m(x,y)$ равно $m(x)$ (как и раньше, ограниченные множители мы опускаем), поэтому $m_n(x,y) = m(x,y)$ при $n = KP(x)$. Теперь рассмотрим функцию

$$\langle x,y,z\rangle \mapsto \sum_{n \geqslant KP(x)} \frac{m_n(x,y)m_n(x,z)}{2^{-n}}.$$

Среди всех слагаемых есть и член с $n = KP(x)$. С другой стороны,

$$\sum_{x,y,z} \sum_{n \geqslant KP(x)} \frac{m_n(x,y)m_n(x,z)}{2^{-n}} \leqslant \sum_x \sum_{n \geqslant KP(x)} \frac{\sum_y m_n(x,y)\sum_z m_n(x,z)}{2^{-n}} \leqslant$$
$$\leqslant \sum_x \sum_{n \geqslant KP(x)} 2^{-n} \leqslant \sum_x 2m(x) \leqslant 2.$$

(Как и раньше, мы опускаем константы — их учёт даст ограниченный множитель в окончательной оценке.)

$\boxed{\mathbf{114}}$ Покажите, что неравенство из теоремы 26 (с. 59) для префиксной сложности выполняется с точностью $O(1)$:

$$2\,KP(x,y,z) \leqslant KP(x,y) + KP(x,z) + KP(y,z) + O(1)$$

для любых слов x,y,z. [Указание: сложите базисное неравенство $KP(x,y,z) + KP(z) \leqslant KP(x,z) + KP(y,z)$ с неравенством $KP(x,y,z) \leqslant KP(x,y) + KP(z)$.]

$\boxed{\mathbf{115}}$ Докажите, что при некотором c для любого слова x и любого числа n найдётся такое слово y длины n, что

$$KP(x,y) \geqslant KP(x) + n - c.$$

[Указание: для любых z и n существует слово y длины n, для которого $KP(y\,|\,z) \geqslant n$.]

Сходное утверждение можно сформулировать не для пар, а для продолжений данного слова x на n битов (аналогичное утверждение для простой колмогоровской сложности составляло содержание задачи 46 на с. 52):

Теорема 71.

$$\max\{KP(xy)\,|\,l(y) = n\} \geqslant KP(x\,|\,n) + n - O(1).$$

Другими словами, при некотором c для всех x и n можно так продлить слово x на n битов, чтобы его сложность увеличилась на $n - c$, правда не по сравнению с $KP(x)$, а лишь по сравнению с $KP(x\,|\,n)$.

◀ Переходя к априорным вероятностям, перепишем искомое неравенство как

$$2^n \min\{m(xy) \mid l(y) = n\} \leqslant m(x \mid n) \cdot O(1).$$

Левая часть не превосходит $\sum\{m(xy) \mid l(y) = n\}$ (если заменить все слагаемые на наименьшее из них, сумма лишь уменьшится). А эта величина представляет собой при каждом n (как функция от x) перечислимую снизу полумеру, так что остаётся воспользоваться максимальностью. ▶

116 Покажите, что чуть более слабое утверждение с $KP(x) - KP(n)$ вместо $KP(x \mid n)$ в правой части можно вывести из задачи 115.

4.7.4. Обычная и префиксная сложности

Мы уже видели оценки для префиксной сложности в терминах обычной (теорема 65). Есть и другие связи. Например, следующее замечание Левина позволяет определить обычную сложность в терминах префиксной:

Теорема 72. *Простую сложность $KS(x)$ можно определить как значение i, при котором $KP(x \mid i) = i$. Более точно, $KP(x \mid KS(x)) = KS(x) + O(1)$; с другой стороны, если $KP(x \mid i) = i + \delta$, то $KS(x) = i + O(\delta)$.*

◀ Мы уже видели в задаче 44 (с. 51), что $KS(x) \leqslant KS(x \mid KS(x))$ с точностью $O(1)$. С другой стороны, $KP(x \mid KS(x)) \leqslant KS(x)$ с той же точностью. В самом деле, при известной длине описания это описание можно считать беспрефиксным (ограничившись описаниями только этой длины). Тем самым первое утверждение теоремы доказано. Остаётся заметить, что функция $i \mapsto KP(x \mid i)$ меняется медленно: изменение i на d приводит к изменению её значения на $O(\log d)$. Поэтому отклонение i от значения $KS(x)$, при котором достигается равенство, приводит почти к такому же по величине отклонению i от $KP(x \mid i)$. ▶

Как недавно заметил Бауэнс, из этого утверждения можно получить равенства, связывающие обычную и префиксную сложности. Возьмём частный случай теоремы о сложности пары:

$$KP(x) = KP(x, KP(x)) = KP(KP(x)) + KP(x \mid KP(x), KP(KP(x)))$$

с точностью $O(1)$. Если нас устраивает точность $O(KP(KP(KP(x))))$, то пару $\langle KP(x), KP(KP(x))\rangle$ в условии можно заменить на разность $KP(x) - KP(KP(x))$, и получится

$$KP(x) - KP(KP(x)) = KP(x \mid KP(x) - KP(KP(x))) + O(KP^{(3)}(x))$$

(используем сокращение $KP^{(i)}(x)$ для i итераций функции KP). Остаётся воспользоваться предыдущей теоремой, и получится такой результат Соловея (из [161]):

Теорема 73.

$$KS(x) = KP(x) - KP(KP(x)) + O(KP^{(3)}(x)).$$

Другими словами, $KP(x) - KS(x)$ равно $KP(KP(x)) + O(KP^{(3)}(x))$.

Соловей заметил также, что в правой части можно заменить KP на KS, то есть что $KP(x) - KS(x) = KS(KS(x)) + O(KS^{(3)}(x))$. На самом деле $O(KP^{(3)}(x))$ и $O(KS^{(3)}(x))$ — это одна и та же точность, и $KP(KP(x))$ и $KS(KS(x))$ совпадают с этой точностью. Чтобы убедиться в этом, заметим, что уже известная формула для $KP(x) - KS(x)$ позволяет утверждать, что

$$|KP(KP(x)) - KP(KS(x))| \leqslant KP^{(3)}(x) + O(\log KP^{(3)}(x))$$

и

$$|KS(KP(x)) - KS(KS(x))| \leqslant KP^{(3)}(x) + O(\log KP^{(3)}(x)),$$

поскольку разница сложностей двух чисел не превосходит префиксной сложности их разности. С другой стороны, в формулу для разности $KP(y) - KS(y)$ можно подставить $y = KP(x)$ и заключить, что

$$KP(KP(x)) - KS(KP(x)) = O(KP^{(3)}(x)).$$

Таким образом, все четыре варианта сложности сложности x (комбинации простой и префиксной) отличаются друг от друга на $O(KP^{(3)}(x))$. В частности,

$$KP(KP(x)) - KS(KS(x)) = O(KP^{(3)}(x)),$$

и остаётся заметить, что из $|u - v| = O(KP(u))$ вытекает $|KP(u) - KP(v)| = O(\log KP(u))$. Таким образом, $KP(KP(KP(x)))$ и $KP(KS(KS(x)))$ «логарифмически близки», если под логарифмической близостью a и b понимать $|a - b| = O(\log a)$. Несложно проверить, что это отношение (логарифмической близости) симметрично и транзитивно (с точностью до увеличения константы в O-обозначении). Остаётся заметить, что $KS(v)$ и $KP(v)$ тоже логарифмически близки (при любом v, в том числе при $v = KS(KS(x))$), так что по транзитивности $KP(KP(KP(x)))$ и $KS(KS(KS(x)))$ также логарифмически близки и уж заведомо $O(KP^{(3)}(x))$ и $O(KS^{(3)}(x))$ — одна и та же точность.

Таким образом, мы получили и другое равенство Соловея из [161]:

Теорема 74.

$$KP(x) = KS(x) + KS(KS(x)) + O(KS^{(3)}(x)).$$

Другими словами, неравенство теоремы 65 (с. 118) оказывается близким к равенству.

117 Дайте прямое доказательство оценки

$$KS(x) \leqslant KP(x) - KP(KP(x)) + KP^{(3)}(x) + O(1),$$

оценив число слов, при которых правая часть мала.

[Указание. Мы видели (теорема 64, с. 117), что логарифм числа слов, для которых $KP(x) \leqslant n$, не превосходит $n - KP(n) + c$ при некотором c и при всех n. Кроме того, такие слова можно перечислять (при известном n), так что каждое такое слово можно восстановить, зная n и порядковый номер слова в перечислении.

Этот порядковый номер мы будем записывать в виде двоичного слова ровно из $n - KP(n) + c$ битов (добавив слева нули, если нужно). Имея такое слово, мы уже знаем $n - KP(n)$ (константу c считаем фиксированной), так что для задания n достаточно задать ещё $KP(n)$, что можно сделать с помощью самоограниченного кода длиной $KP(KP(n))$. Приписав к этому коду упомянутый выше порядковый номер, мы получим (для любого слова x с $KP(x) \leqslant n$, в том числе для всех слов префиксной сложности n) слово длиной $KP(KP(n)) + n - KP(n) + c$, по которому мы можем эффективно восстановить слово x.]

5. Монотонная и априорная сложности и случайность

5.1. Вероятностные машины и полумеры на дереве

Определяя априорную вероятность в главе 4, мы рассматривали вероятностные алгоритмы (машины), которые на выходе печатали натуральное число и останавливались. Теперь мы рассмотрим другой класс вероятностных алгоритмов, которые выдают на выход бит за битом и не обязаны останавливаться. Алгоритм такого вида задаёт случайную величину, значениями которой являются элементы Σ (конечные и бесконечные последовательности нулей и единиц).

В качестве примера рассмотрим алгоритм, выдающий на выходе случайные биты, получаемые от датчика:

```
while true do
   b := random;
   OutputBit(b);
od
```

Соответствующая случайная величина сосредоточена на бесконечных последовательностях (любое множество конечных последовательностей имеет вероятность нуль) и имеет там равномерное распределение. Но в общем случае конечные последовательности могут появляться с положительной вероятностью (это соответствует ситуации, когда алгоритм с некоторого момента новых битов не выдаёт).

Для каждого алгоритма A описанного типа рассмотрим функцию, определённую на двоичных словах и принимающую вещественные значения:

$$a(x) = \Pr[\text{выход } A \text{ начинается на } x].$$

Формально говоря, эту функцию надо определять так. Каждому вероятностному алгоритму A соответствует некоторое отображение $\bar{A}$ множества Ω (бесконечные последовательности нулей и единиц) в множество Σ. Именно, $\bar{A}(\omega)$ есть последовательность битов, которая получается на выходе, если использовать в качестве случайных битов члены последовательности ω (очередной вызов $b := \texttt{random}$ берёт следующий бит последовательности ω). Для приведённой выше программы, очевидно, $\bar{A}(\omega) = \omega$.

После этого $a(x)$ определяется как равномерная мера прообраза множества Σ_x при отображении $\bar{A}$ (где Σ_x есть множество всех конечных и бесконечных последовательностей, начинающихся на слово x).

118 Укажите отображение $\bar{A}$ и функцию a для алгоритма A, который печатает на выходе бесконечную последовательность нулей (и не использует случайных битов).

В этом разделе мы опишем функции, соответствующие вероятностным алгоритмам указанного вида.

Теорема 75. *Пусть A — вероятностный алгоритм описанного вида, a — соответствующая ему функция. Тогда*:

(а) $a(x) \geqslant 0$ *при всех* x;

(б) $a(\Lambda) = 1$ (*здесь* Λ *обозначает пустое слово*);

(в) $a(x) \geqslant a(x0) + a(x1)$ *для любого слова* x;

(г) *функция* a *перечислима снизу.*

Перечислимость снизу определялась в разделе 4.1 (с. 88) для последовательностей действительных чисел. Для функций на словах определение аналогично: $a(x) = \lim_i a(x, i)$, где a — вычислимая функция двух аргументов, $a(x, i)$ определено всех слов x и для всех натуральных i, является рациональным числом (или специальным символом $-\infty$) и не убывает по i.

◀ Первые три утверждения очевидны:

(а) Вероятность всегда неотрицательна.

(б) $a(\Lambda) = 1$, поскольку пустое слово является началом любого выхода.

(в) $a(x) \geqslant a(x0) + a(x1)$, поскольку события «выход начинается на $x0$» и «выход начинается на $x1$» несовместны и являются подмножествами события «выход начинается на x».

Заметим, что неравенство пункта (в) не обязано быть равенством: разность

$$a(x) - a(x0) - a(x1)$$

есть вероятность того, что на выходе появится слово x и больше никаких битов не появится.

(г) Чтобы доказать перечислимость снизу функции a, нам надо уметь получать приближения снизу к $a(x)$ для данного слова x. Будем моделировать поведение алгоритма A для всех возможных значений случайных битов. Время от времени будут обнаруживаться значения случайных битов, которые гарантируют появление на выходе слова x, то есть интервалы в пространстве Ω, на которых функция $\bar{A}$ равна x или какому-нибудь продолжению слова x. Вероятность $a(x)$ равна суммарной мере всех таких интервалов, и в качестве искомого приближения $a(x, i)$ мы берём меру интервалов, обнаруженных к шагу i. ▶

Функции a, определённые на двоичных словах, принимающие действительные значения и удовлетворяющие условиям (а) – (г) теоремы 75, называют *перечислимыми снизу полумерами на дереве* (двоичных слов). Мы будем называть *полумерами на дереве* функции, удовлетворяющие условиям (а) – (в); условие (г) выделяет из них перечислимые снизу.

Важно не смешивать полумеры на дереве с ранее рассматривавшимися полумерами на натуральных числах (или на словах как изолированных конструктивных объектах): прежние полумеры соответствовали алгоритмам, которые выдавали на

выходе какое-либо число (или слово) и останавливались. Чтобы подчеркнуть различие, ранее рассматривавшиеся полумеры можно называть *дискретными*, а полумеры на дереве называть *непрерывными*.

119 Покажите, что полумеры на дереве (функции со свойствами (а) – (в)) соответствуют мерам (в смысле теории меры) на пространстве Σ конечных и бесконечных последовательностей. Найдите меру множества всех бесконечных продолжений слова x при таком соответствии. [Ответ: если a — полумера на дереве, то мера этого множества равна пределу убывающей по n последовательности

$$\alpha_n = \sum\{a(y) \mid y \text{ есть продолжение } x \text{ длины } n\}.$$

Здесь α_n определено при $n \geqslant l(x)$ и равно $a(x)$ при $n = l(x)$.]

120 Покажите, что для перечислимой снизу полумеры на дереве сумма $\sum_x a(x)$ может быть бесконечной. [Указание. Рассмотрите алгоритм, копирующий случайные биты на выход.]

Верно и обратное к теореме 75 утверждение:

Теорема 76. *Всякая перечислимая снизу полумера на дереве соответствует некоторому вероятностному алгоритму.*

◄ Идею доказательства легко объяснить в терминах выделения места, по аналогии с доказательством теоремы 46 (с. 91). Изменение состоит в том, что теперь запросы носят иерархический характер. Две организации (которые мы будем условно называть 0 и 1) требуют выделить им непересекающиеся подмножества в пространстве Ω (которое можно отождествить с отрезком $[0,1]$), при этом размеры запрашиваемых ими областей увеличиваются со временем, но в сумме ни в какой момент не превышают единицы.

В каждой из организаций есть по два подразделения (в организации 0 есть подразделения 00 и 01, в организации 1 есть два подразделения 10 и 11), которые требуют выделения им места внутри области, выделенной для всей организации. При этом их требования тоже растут со временем, но для каждого момента времени в сумме не превышают требований всей организации (в тот же момент). Аналогично для ещё более мелких подразделений 000, 001 и так далее: в пределе каждое подразделение x требует выделить (не обязательно сплошной) участок, имеющий размер $a(x)$, где a — заданная полумера. При этом однажды выделенная любому подразделению область навсегда остаётся за ним и перераспределена быть не может.

Соответствие между этой схемой выделения места и вероятностным алгоритмом таково: если последовательность случайных битов попадает в область, выделенную подразделению x, это означает, что выход алгоритма начинается с x (для данной последовательности случайных битов).

Представив себе описанную ситуацию, легко понять, что такое распределение осуществимо. Тем не менее мы приведём более формальное рассуждение (одновременно объясняя смысл его этапов с точки зрения описанной метафоры).

Лемма 1. *Пусть a — произвольная перечислимая снизу полумера на дереве. Тогда существует вычислимая всюду определённая неубывающая по i функция $\langle x, i\rangle \mapsto a(x, i)$, аргументами которой являются двоичное слово x и натуральное*

число i, значения которой — неотрицательные двоично-рациональные числа, $\lim_i a(x,i) = a(x)$ и для каждого i функция $x \mapsto a(x,i)$ является полумерой, у которой лишь конечное число значений отлично от нуля.

Другими словами, можно считать, что на каждом шаге запросы на выделение места подчиняются дополнительным ограничениям:

- запрашиваемые размеры двоично-рациональны (имеют вид $k/2^n$ для некоторых целых k и n);
- лишь конечное число подразделений имеет ненулевые запросы;
- запросы согласованы (запрос любого подразделения не меньше суммы запросов его частей).

Доказательство. Будем постепенно корректировать функцию a из определения перечислимой снизу полумеры (не меняя само́й предельной полумеры). Сначала добьёмся, чтобы все значения были двоично-рациональными. Это можно сделать, заменив $a(x,i)$ на ближайшее снизу двоично-рациональное число со знаменателем 2^i (отрицательные числа заменяем на нули).

Затем можно добиться выполнения второго условия, положив $a(x,i) = 0$ для слов x длины больше i.

Наконец, можно добиться выполнения третьего условия, выполнив замену

$$a(x,i) := \max(a(x,i), a(x0,i) + a(x1,i))$$

последовательно в порядке убывания длин слов. Поскольку функция $a(x)$ является полумерой по предположению, такие замены сохранят неравенство $a(x,i) \leqslant a(x)$. Легко проверить, что значения полумеры (пределы $\lim_i a(x,i)$ при каждом x) при нашей коррекции не изменятся. Лемма 1 доказана.

Следующую лемму удобно формулировать, введя несколько вспомогательных определений. Будем называть полумеру с конечным числом ненулевых двоично-рациональных значений «простой» полумерой.

Будем называть «простым множеством» объединение конечного числа интервалов в Ω. (Напомним, что интервал в Ω есть множество вида Ω_z, состоящее из всех бесконечных продолжений данного слова z. Таким образом, простыми являются множества, принадлежность к которым определяется конечным числом битов последовательности.)

Будем называть «простым семейством» семейство простых множеств A_x, индексированное двоичными словами x, в котором все множества, кроме конечного числа, пусты, и для каждого слова x множества A_{x0} и A_{x1} являются непересекающимися подмножествами множества A_x. Для такого семейства функция $x \mapsto \mu(A_x)$, где μ — равномерная мера на Ω, является простой полумерой.

Лемма 2. *Для всякой простой полумеры существует простое семейство, которому она соответствует.*

Доказательство. Начнём строить такое семейство, начиная с пустого слова и постепенно увеличивая длину слова-индекса. На каждом шаге нам нужно будет внутри простого множества A_x найти два непересекающихся простых подмножества A_{x0} и A_{x1} заданных мер (при этом сумма этих мер не превосходит меры объемлющего множества). Ясно, что это возможно. Лемма 2 доказана.

Лемма 3. *Пусть дана простая полумера $b(x)$ и соответствующее ей простое семейство B_x. Пусть дана также простая полумера c, причём $c(x) \geqslant b(x)$ для всех x. Тогда можно построить простое семейство C_x, соответствующее полумере c, для которого $C_x \supset B_x$ при всех x.*

Доказательство. Повторим рассуждение из доказательства леммы 2, только теперь внутри простого множества есть два непересекающихся простых подмножества, и надо увеличить их меры до заданных значений, оставив подмножества непересекающимися и не выйдя за пределы множества. Ясно, что это возможно. Лемма 3 доказана.

Доказательства лемм 2 и 3 эффективны в том смысле, что по таблицам значений простых полумер можно алгоритмически построить соответствующие простые семейства.

Будем теперь применять лемму 3 по очереди к простым полумерам, получающимся фиксацией $i = 0, 1, 2, \ldots$ в $a(x, i)$. Получим двупараметрическое семейство простых множеств $U(x, i)$, при этом

- описание множества $U(x, i)$ (список входящих в него интервалов) строится алгоритмически по x и i;
- мера множества $U(x, i)$ равна $a(x, i)$ (и стремится к $a(x)$ при $i \to \infty$);
- при любых x и i множества $U(x0, i)$ и $U(x1, i)$ являются непересекающимися подмножествами множества $U(x, i)$;
- $U(x, i) \subset U(x, i+1)$ при любых x и i.

Теперь искомый вероятностный алгоритм, порождающий полумеру a, можно построить так: строим множества $U(x, i)$ при всех x и i и параллельно запрашиваем случайные биты, записывая их в последовательность ω. Как только обнаруживается, что $\omega \in U(x, i)$ для каких-то x и i, выдаём на выход ещё не выданные биты слова x.

Заметим, что если последовательность ω оказалась в $U(x, i)$, то она автоматически оказывается в $U(y, i)$ для любого начала y слова x. Заметим также, что ω не может оказаться одновременно в $U(x, i)$ и $U(x', i)$ для несовместных слов x и x' (слов, не являющихся началами друг друга). Поэтому уже выданные на выход биты не потребуется «отзывать».

Слово x или его продолжение появится на выходе такого алгоритма тогда и только тогда, когда последовательность ω принадлежит объединению возрастающей последовательности множеств $U(x, i)$ при $i = 0, 1, 2, \ldots$; вероятность этого события есть предел мер множеств $U(x, i)$, а этот предел совпадает с $a(x)$, что и требовалось. ▶

Теоремы 75 и 76 показывают, что перечислимые снизу полумеры на дереве можно эквивалентно определить как функции, соответствующие распределениям на выходе вероятностных алгоритмов.

Среди всех вероятностных алгоритмов описанного вида выделяются алгоритмы, которые почти наверно выдают бесконечную последовательность (вероятность получить конечную последовательность равна нулю). Им соответствуют полумеры, являющиеся вычислимыми мерами. Точнее говоря, верна следующая теорема:

Теорема 77. (а) *Пусть μ — вычислимая мера на пространстве Ω. Тогда функция p, определяемая формулой $p(x) = \mu(\Omega_x)$, является перечислимой снизу полумерой, причём $p(x) = p(x0) + p(x1)$ для всех x.*

(б) *Если для перечислимой снизу полумеры p выполнено равенство $p(x) = p(x0) + p(x1)$ при всех x, то она определяет некоторую вычислимую меру на Ω.*

Напомним, что мы включаем в определение полумеры условие $p(\Lambda) = 1$; без этого, или хотя бы без вычислимости $p(\Lambda)$, второе утверждение теоремы перестаёт быть верным.

◀ (а) Если действительное число α вычислимо, и a_n — рациональное приближение к нему с ошибкой $1/n$, то $b_n = a_n - 1/n$ будет приближением снизу с ошибкой $2/n$. Вообще говоря, последовательность b_n может не быть монотонной (неубывающей), но последовательность

$$c_n = \max(b_0, b_1, \ldots, b_n)$$

заведомо будет вычислимой неубывающей последовательностью рациональных чисел, стремящейся к α. Таким образом, каждое вычислимое действительное число перечислимо снизу. Проводя эту конструкцию параллельно для всех x, мы убеждаемся, что любая вычислимая мера задаёт перечислимую снизу полумеру. Поскольку Ω_x есть объединение непересекающихся подмножеств Ω_{x0} и Ω_{x1}, то $p(x) = p(x0) + p(x1)$ (аддитивность меры).

(б) Пусть p — перечислимая снизу полумера, для которой $p(x) = p(x0) + p(x1)$ при всех x. Покажем индукцией по длине x, как найти приближение к $p(x)$ с любой заданной точностью. Для пустого слова значение $p(\Lambda)$ равно 1 по определению. Если мы уже умеем находить приближения к $p(x)$ сверху и снизу с любой точностью, а хотим сделать это для $p(x0)$ и $p(x1)$, то надо ждать, пока сумма постепенно растущих нижних оценок для $p(x0)$ и $p(x1)$ не приблизится к (убывающей) верхней оценке для $p(x)$. Другими словами, верхние оценки для $p(x1)$ можно получать, вычитая из верхних оценок для $p(x)$ (которые возможны по предположению индукции) нижние оценки для $p(x0)$. ▶

Эту теорему можно интерпретировать следующим образом. Пусть нам нужен датчик случайных чисел, который выдаёт последовательность нулей и единиц, распределённую по некоторой вычислимой мере p (вероятность того, что выдаваемая датчиком последовательность начинается на x, должна быть равна $p(x)$). Тогда теоремы 76 и 77 говорят, что такой датчик можно построить в виде вероятностного алгоритма, который внутри себя имеет равномерный датчик, а на выходе имеет нужное распределение.

Заметим, что в частном случае вычислимых мер можно воспользоваться более простой и явной конструкцией, чем дана в доказательстве теоремы 76. А именно, разобьём отрезок $[0, 1]$ на две части длиной $p(0)$ и $p(1)$. Первую из них разобьём ещё на две части длиной $p(00)$ и $p(01)$, вторую — на части длиной $p(10)$ и $p(11)$ и так далее. Для каждого двоичного слова z получится некоторый отрезок π_z длиной $p(z)$; отрезки π_z для всех слов z данной длины образуют разбиение отрезка $[0, 1]$.

Теперь рассмотрим вероятностный алгоритм, который бросает честную монету и получает последовательность α случайных битов. Эта последовательность рассматривается как двоичное разложение некоторого числа (которое мы тоже будем обозначать α). Параллельно с получением α алгоритм ищет такие z, при которых α содержится строго внутри π_z (и это можно сказать на основе уже имеющейся информации о α и о границах отрезков π_z, которые параллельно вычисляются с возрастающей точностью).

Возникающие при этом слова z являются началами друг друга (чем больше битов числа α известно, тем длиннее может быть z). Они являются началами последовательности битов, которая выдаётся на выход.

Вообще говоря, алгоритм этот может дать конечную последовательность, если число α совпадёт с концом одного из отрезков π_z, но таких концов счётное множество, и вероятность этого события равна 0. Выход алгоритма начинается на z тогда и только тогда, когда число α содержится строго внутри π_z, так что вероятности правильные.

Говоря более формально, описанное преобразование T последовательности битов α в выходную последовательность битов $\beta = T(\alpha)$ переводит равномерную меру в меру p.

(Ничего неожиданного в этом алгоритме нет. Если у вас есть датчик случайных равномерно распределённых на $[0,1]$ чисел, а нужно имитировать несимметричную монету, где 0 и 1 имеют вероятности, скажем, $2/3$ и $1/3$, то надо сравнить случайную точку с границей $2/3$. Для получения второго бита с тем же распределением надо каждый из отрезков $[0,2/3]$ и $[2/3,1]$ снова разделить на неравные части в пропорции $2:1$. Ровно это и делается в описанном алгоритме.)

Теорема 76 показывает, что достаточно иметь «честную монету», то есть генератор независимых случайных равномерно распределённых битов, чтобы (с помощью алгоритмического преобразования) порождать последовательности битов с произвольным вычислимым распределением и даже по произвольной перечислимой снизу полумере. Можно ещё отметить, что выбор равномерного распределения (честной монеты) в качестве источника случайности не принципиален:

121 Покажите, что в теореме 75 можно заменить равномерную меру на любую вычислимую меру и даже полумеру. [Указание: композиция двух алгоритмических преобразований последовательностей также является алгоритмическим преобразованием.]

122 Покажите, что в теореме 76 можно использовать вместо равномерной меры любую вычислимую меру, в которой каждая отдельная последовательность имеет нулевую вероятность. [Указание. Вычислимую меру P можно преобразовать в равномерную, если при получении z от P-датчика выдавать на выход биты слова x, для которого отрезок π_z целиком содержится внутри отрезка I_x (чисел, двоичные разложения которых начинаются на x).]

Условие в этой задаче необходимо: если некоторая бесконечная последовательность имеет положительную меру, то значение отображения $\bar{A}$ на этой последовательности имеет положительную вероятность, и равномерной меры не получится.

Но если этого препятствия нет, то датчик с любым известным заранее вычислимым распределением может заменить равномерный.

Вопрос о том, что можно сделать, если заранее распределение неизвестно, рассматривается при построении экстракторов случайности; разумные постановки его возможны и в классической теории вероятностей, и в терминах колмогоровской сложности. См. краткий обзор А. Вигдерсона [190] и приводимые там ссылки, а также обзор М. Зиманда о сложностных постановках задачи [195]. Мы не затрагиваем этих вопросов в нашей книге.

5.2. Наибольшая перечислимая полумера на дереве

Теорема 78. *Среди всех перечислимых снизу полумер на дереве существует наибольшая (с точностью до константы) полумера a: для любой перечислимой снизу полумеры a' на дереве выполнено неравенство $a'(x) \leqslant ca(x)$ для некоторой константы c и для всех x.*

◀ Как и для полумер на $\mathbb{N}$ (теорема 47, с. 93), рассмотрим вероятностный алгоритм A, который сначала случайно выбирает некоторый вероятностный алгоритм, а затем моделирует выбранный алгоритм. Если полумера a' соответствует вероятностному алгоритму A', то $a'(x) \leqslant (1/\varepsilon)a(x)$, где ε — вероятность того, что будет выбран именно алгоритм A'. ▶

Другой способ доказательства состоит в том, что мы располагаем все перечислимые снизу полумеры в вычислимую последовательность $a_0, a_1, \ldots$, а затем рассматриваем полумеру $a = \sum_i \lambda_i a_i$, где λ_i — ряд с вычислимыми членами и суммой 1 (например, $\lambda_i = 2^{-i-1}$).

Тонкость здесь в том, что надо построить вычислимую последовательность всех полумер, то есть перечислимую снизу функцию $u(i, x)$, для которой: (1) при каждом фиксированном i функция $u_i : x \mapsto u(i, x)$ является полумерой; (2) среди u_i содержатся все перечислимые снизу полумеры.

Это можно сделать, либо перечисляя все вероятностные алгоритмы (что соответствует ранее приведённому доказательству), либо принудительно корректируя программы так, чтобы они перечисляли полумеры. Это рассуждение аналогично случаю дискретных полумер (раздел 4.2, страница 93), и мы не будем приводить детали. (Скажем лишь, что в случае нарушения условия $p(x) \geqslant p(x0) + p(x1)$ надо увеличивать $p(x)$, если только это не приведёт к тому, что $p(\Lambda)$ станет больше единицы.)

$\boxed{123}$ Проведите это рассуждение подробно.

(Замечание в скобках: доказательство теоремы 78 даёт даже немного больше, чем утверждается. А именно, не только вероятность появления слова *с началом* x (то есть $p(x)$), но и вероятность появления на выходе *в точности слова* x, то есть разность $p(x) - p(x0) - p(x1)$, для построенной «универсальной» машины больше, чем для любой другой.)

$\boxed{124}$ Покажите, что всё сказанное выше естественно переносится на вероятностные алгоритмы, которые выдают на выход (по одному) не биты, а натуральные

числа (каждое число выдаётся как единое целое). Такие алгоритмы соответствуют перечислимым снизу мерам на пространстве конечных и бесконечных последовательностей натуральных чисел.

$\boxed{\textbf{125}}$ (Продолжение.) Пусть m — наибольшая перечислимая снизу полумера на пространстве конечных и бесконечных последовательностей натуральных чисел. Покажите, что её ограничение на последовательности длины 1 даёт априорную вероятность на натуральных числах в смысле главы 4, а её ограничение на двоичные слова даёт наибольшую полумеру на двоичном дереве (в смысле теоремы 78).

$\boxed{\textbf{126}}$ Покажите, что $a(0^n1)$ и $m(n)$ отличаются не более чем на ограниченный множитель, если a — максимальная полумера из теоремы 78, а $m(n)$ — априорная вероятность натурального числа n в смысле главы 4. (Здесь вместо 0^n1 можно взять любое беспрефиксное кодирование натуральных чисел.) [Указание: см. далее теорему 79.]

Фиксируем некоторую наибольшую перечислимую снизу полумеру на дереве и обозначим её $a(x)$. Её также называют *универсальной полумерой* на дереве. Величину $a(x)$ можно было бы назвать *априорной вероятностью слова* x (как элемента дерева), но важно не путать её с априорной вероятностью в смысле главы 4. Чтобы различать эти два понятия, можно говорить о *дискретной* и *непрерывной* априорных вероятностях. Величину

$$KA(x) = -\log a(x)$$

будем называть *априорной сложностью* слова x. (Здесь можно не опасаться путаницы, так как в главе 4 аналогичное определение приводило к префиксной сложности и специальное наименование излишне.) Различные варианты выбора наибольшей полумеры приводят к разным функциям априорной сложности, но они отличаются друг от друга не более чем на аддитивную константу (поскольку различные наибольшие полумеры на дереве отличаются не более чем в константу раз).

В англоязычной литературе для априорной сложности иногда используется обозначение $KM(x)$ (а монотонная сложность, которую мы так обозначаем, обозначается $Km(x)$ или $K_m(x)$).

В следующем разделе мы изучим свойства априорной сложности. Заметим сразу же, что формально по нашему определению априорная сложность может не быть целым (и даже рациональным) числом. Но поскольку большинство интересующих нас утверждений всё равно формулируются «с точностью до $O(1)$», то можно было бы заменить $-\log a(x)$ на минимальное число n, при котором $2^{-n} < a(x)$. Тут есть небольшая тонкость: мы используем строгое неравенство, чтобы получаемая функция была перечислима сверху. В дальнейшем мы специально оговариваем те редкие случаи, когда это округление (или его отсутствие) существенно.

5.3. Свойства априорной сложности

Теорема 79. (а) *Априорная сложность монотонна: если x есть начало y, то $KA(x) \leqslant KA(y)$;*

(б) *$KA(x) \leqslant l(x) + O(1)$ для любого x;*

(в) *$KA(x) \leqslant KP(x) + O(1)$ для любого x;*

(г) *если $x_0, x_1, \ldots$ — вычислимая последовательность несравнимых слов (ни одно из них не является началом другого), то $KA(x_i) = KP(x_i) + O(1) = KP(i) + O(1)$;*

(д) *$KP(x) \leqslant KA(x) + 2\log l(x) + O(1)$;*

(е) *более того, $KP(x) \leqslant KA(x) + KP(l(x)) + O(1)$;*

(ё) *и более того, $KP(x \,|\, l(x)) \leqslant KA(x) + O(1)$;*

(ж) *бесконечная последовательность нулей и единиц вычислима тогда и только тогда, когда априорная сложность её начальных отрезков ограничена;*

(з) *если $f\colon \Sigma \to \mathbb{N}_\perp$ — вычислимое непрерывное отображение, то $KP(f(x)) \leqslant KA(x) + O(1)$ для любого слова x, на котором $f(x)$ определено (не равно $\perp$).*

◀ (а) Мера подмножества не превосходит меры множества.

(б) Функция $p(x) = 2^{-l(x)}$ является перечислимой снизу полумерой на дереве, и потому $p(x) \leqslant ca(x)$ для некоторого c.

(в) Машины, которые выдают двоичное слово (как единое целое) и останавливаются, можно рассматривать как частный случай машин, выдающих последовательности битов. Поэтому $m(x) \leqslant ca(x)$, где m — априорная вероятность в смысле главы 4, откуда и следует требуемое.

Поучительно провести это рассуждение в терминах полумер. Положим $m'(x)$ равным сумме $m(y)$ по всем продолжениям y слова x, где m — дискретная априорная вероятность. (Ещё положим в порядке исключения $m'(\Lambda) = 1$.) Тогда m' будет полумерой на дереве и потому $m(x) \leqslant m'(x) = O(a(x))$.

(г) Если x_i — вычислимая последовательность несравнимых двоичных слов, а a — априорная вероятность на дереве (универсальная полумера на дереве), то функция $i \mapsto a(x_i)$ является полумерой на $\mathbb{N}$. В самом деле, она перечислима снизу, а события «x_i появляется на выходе вероятностного алгоритма» несовместны, и потому сумма их вероятностей не превосходит единицы. Поэтому $KP(i) \leqslant KA(x_i) + O(1)$.

С другой стороны, $KP(x_i) = KP(i) + O(1)$, поскольку из i можно алгоритмически получить x_i и наоборот; наконец, $KA(x_i) \leqslant KP(x_i) + O(1)$ согласно (в).

(д) Пусть a — априорная вероятность на дереве. Рассмотрим перечислимую снизу функцию $u(x) = a(x)/l(x)^2$. Сумма значений $a(x)$ по всем словам длины n не превосходит 1, поскольку эти слова попарно несравнимы, поэтому

$$\sum_x u(x) = \sum_n \sum_{l(x)=n} \frac{a(x)}{n^2} \leqslant \sum_n \frac{1}{n^2} = O(1),$$

откуда и следует требуемое.

(е) доказывается аналогично, только надо положить $u(x) = a(x)m(l(x))$, где m — априорная вероятность на $\mathbb{N}$.

(ё) Рассмотрим функцию

$$u(x, n) = \begin{cases} a(x), & \text{если } l(x) = n,\\ 0, & \text{если } l(x) \neq n. \end{cases}$$

При каждом n функция $x \mapsto u(x,n)$ является полумерой в смысле предыдущей главы (сумма не превосходит 1), откуда и следует требуемое.

(ж) Для данной бесконечной вычислимой последовательности ω нулей и единиц рассмотрим вероятностный (по форме) алгоритм, который не использует датчика случайных чисел и печатает один за другим биты последовательности ω. Соответствующая полумера равна 1 на любом начале последовательности, и потому априорные вероятности (мы рассматриваем априорную вероятность на дереве) всех этих начал отделены от нуля.

Обратное рассуждение несколько сложнее. Пусть априорные вероятности всех начал последовательности ω больше некоторого рационального $\varepsilon > 0$. Рассмотрим множество B всех двоичных слов, для которых $a(x) > \varepsilon$. Множество B содержит все начала последовательности ω. Оно является деревом (вместе с любым словом содержит все его начала). Кроме того, любое его подмножество, состоящее из попарно несравнимых слов, содержит не более $1/\varepsilon$ элементов (поскольку соответствующие события не пересекаются и их суммарная вероятность не больше 1). Наконец, B перечислимо (строя приближения снизу к $a(x)$, мы рано или поздно обнаружим любое x, для которого $a(x) > \varepsilon$).

Этих свойств уже достаточно, чтобы заключить, что последовательность ω вычислима. В самом деле, рассмотрим максимально возможное число попарно несравнимых слов $x_1, \ldots, x_N$ из B. Для каждого из слов x_i рассмотрим все его продолжения, лежащие в B. Все они сравнимы (иначе можно было бы заменить x_i на два несравнимых продолжения и увеличить N). Поэтому в дереве B из каждого x_i выходит конечная или бесконечная ветвь без ветвлений, и эта ветвь вычислима (поскольку можно перечислять множество B). Последовательность ω является одной из таких ветвей (если бы ω не проходила ни через одно из слов x_i, то её достаточно длинное начало можно было бы добавить к списку несравнимых слов и увеличить N).

(з) Построим вероятностную машину в смысле главы 4, применяя f к выходу вероятностной машины, соответствующей наибольшей перечислимой полумере на дереве, и сравним результат с наибольшей перечислимой полумерой на $\mathbb{N}$ (логарифм которой равен $KP + O(1)$). ▶

Отметим, что априорная сложность отличается по своим свойствам от уже знакомых нам (обычной и префиксной) сложностей. Прежде всего, её определение использует структуру начал на множестве двоичных слов, и потому алгоритмические преобразования, не сохраняющие этой структуры, могут увеличивать сложность более чем на константу.

127 Покажите, что априорная сложность слова x может быть ограниченной, а сложность слова x^R (те же биты в обратном порядке) — сколь угодно большой. (Формально: существует такое c, что для любого n найдётся слово x с $KA(x) < c$ и $KA(x^R) > n$.) [Указание: слово x может начинаться с единицы, за которой идут нули.]

Тем самым нет смысла (в отличие от префиксной и обычной сложностей) говорить об априорной сложности, скажем, натурального числа или графа.

Априорная сложность слова длины n отличается от уже известных нам вариантов сложностей не более чем на $O(\log n)$, но здесь важно, что под логарифмом стоит именно длина слова, а не его сложность, поскольку, скажем, для слов из одних нулей априорная сложность ограничена, а обычная и префиксная — нет.

128 Докажите, что для любого слова x хотя бы одно из чисел $KA(x0)$ и $KA(x1)$ не меньше $KA(x)+1$. (Здесь существенно, что $KA(x)$ определено как $-\log a(x)$ без округления.) Выведите отсюда, что для любого слова x и числа $n \in \mathbb{N}$ можно найти слово y длины n, для которого $KA(xy) \geqslant KA(x)+n$. Покажите, что существует бесконечная двоичная последовательность ω, у которой априорная сложность любого начала не меньше его длины.

(Ср. теорему 71 на с. 129 и задачу 46 на с. 52; заметим, что здесь нет условия n и даже константы $O(1)$.)

129 Покажите, что разности $KS(x)-KA(x)$ и $KA(x)-KS(x)$ могут быть порядка $\log n$ для некоторых слов длины n (и для сколь угодно больших n). [Указание. $KS(x)$ будет больше $KA(x)$, если взять слово из одних нулей. Обратное соотношение имеет место для начал последовательности из предыдущей задачи, у которых $KA(x)=l(x)+O(1)$, а сложность $KS(x)$ бывает меньше длины примерно на логарифм длины, см. задачу 54.]

130 Докажите, что

$$KA(xy) \leqslant KP(x)+KA(y)+O(1),$$

где xy — соединение слов x и y. Здесь существенно, что речь идёт о соединении именно в таком порядке: покажите, что для другого порядка это неверно. [Указание. Пусть U — вероятностный алгоритм в смысле главы 4, порождающий наибольшую перечислимую снизу полумеру (дискретную), а V — вероятностный алгоритм в смысле этой главы, порождающий наибольшую полумеру на дереве. Рассмотрите алгоритм, который вначале действует как U, а после остановки продолжает работу как V (используя свежие случайные биты и добавляя биты к уже выданным). Чтобы показать, что для $KA(yx)$ аналогичная оценка места не имеет, положите $y=0^n$ и $x=1$.]

Ещё одно свойство априорной сложности непосредственно следует из определения. Пусть дана произвольная вычислимая мера μ на пространстве Ω. Тогда для некоторой константы c и для всех слов x выполняется неравенство

$$KA(x) \leqslant -\log \mu(\Omega_x)+c.$$

В самом деле, априорная вероятность на дереве не меньше меры μ (и даже любой другой перечислимой снизу полумеры) с точностью до ограниченного множителя, что после логарифмирования и даёт требуемое неравенство.

Мы обращаем внимание на это (простое) свойство, поскольку на нём основан критерий случайности по Мартин-Лёфу: последовательность ω случайна по вычислимой мере μ тогда и только тогда, когда для её начальных отрезков x это неравенство превращается в равенство, то есть когда разница $-\log \mu(\Omega_x)-KA(x)$

ограничена сверху (снизу она всегда ограничена в силу только что рассмотренного свойства).

Этот критерий является следствием теоремы Левина – Шнорра (критерий случайности в терминах монотонной сложности) и будет доказан вместе с ней в разделе 5.6. Но прежде нам необходимо дать определение монотонной сложности (раздел 5.5), для чего подробнее рассмотреть понятие вычислимого отображения пространства Σ в себя (раздел 5.4).

Можно охарактеризовать априорную сложность и как минимальную перечислимую сверху функцию, удовлетворяющую некоторым условиям — подобно тому, как это сделано в теореме 8 (с. 29) для простой сложности и в теореме 62 (с. 116) для префиксной сложности. А именно, справедлива следующая теорема:

Теорема 80. *Функция KA является минимальной с точностью до константы перечислимой сверху функцией K с таким свойством*:

$$\sum_{x \in M} 2^{-K(x)} \leqslant 1$$

для любого множества M попарно несравнимых двоичных слов.

◀ Поскольку слова $x \in M$ несравнимы, соответствующие множества Σ_x не пересекаются и вероятности попадания в них в сумме не превосходят единицы.

С другой стороны, пусть дана произвольная перечислимая сверху функция K, обладающая указанным в теореме свойством. Нам нужно определить перечислимую снизу полумеру, которая была бы не меньше (перечислимой снизу) функции 2^{-K}. Заметим, что функция 2^{-K} не обязана быть полумерой, её значения на словах x, $x0$ и $x1$, вообще говоря, никак не связаны. Поэтому её надо увеличить — в той мере, в которой это неизбежно. А именно, положим $u(x)$ равным точной верхней грани всех сумм вида

$$\sum_{y \in M} 2^{-K(y)}$$

по всем множествам M попарно несравнимых продолжений слова x. Легко проверить, что $u(x)$ действительно будет перечислимой снизу полумерой и будет не меньше $2^{-K(x)}$, что и требовалось. ▶

131 Будем рассматривать функции $b\colon \Xi \to [0,1]$, обладающие следующим свойством: существует мера μ на Ω, для которой $b(x) \leqslant \mu(\Omega_x)$ при всех x.

(**а**) Докажите, что любая полумера на дереве обладает этим свойством.

(**б**) Докажите, что любая перечислимая снизу функция, обладающая этим свойством, мажорируется некоторой перечислимой снизу полумерой на дереве.

5.4. Вычислимые отображения $\Sigma \to \Sigma$

В главе 4 мы рассматривали префиксную сложность, определённую в терминах кратчайших описаний, и априорную вероятность, определённую в терминах вероятностных машин — и одно оказалось логарифмом другого (плюс $O(1)$).

В этой главе мы рассмотрели другой вариант априорной вероятности, и возникает естественный вопрос: какая сложность ей соответствует? Такая сложность

действительно может быть естественно определена и называется монотонной сложностью, правда, она уже не совпадает с логарифмом априорной вероятности на дереве. Но её определение (см. далее раздел 5.5) требует некоторой подготовки, которой и посвящён этот раздел.

Алгоритмы (машины), рассматриваемые нами при определении априорной вероятности (универсальной полумеры) на дереве, состоят из двух частей: датчика случайных чисел, порождающего последовательность случайных битов, и алгоритма, который порождает выходные биты, используя эту последовательность. Сейчас мы изучим подробнее эту вторую составляющую, введя понятие вычислимого отображения пространства Σ (конечных и бесконечных последовательностей нулей и единиц) в себя. Отметим, что мы рассматриваем отображения, определённые на всём Σ (зато среди значений может быть пустое слово, что в некотором смысле заменяет неопределённость).

5.4.1. Непрерывные отображения $\Sigma \to \Sigma$

Пусть дано отображение $f\colon \Sigma \to \Sigma$, определённое на всём Σ. Мы будем называть его *непрерывным*, если выполнены два свойства:

(1) оно монотонно: если $x \in \Sigma$ является началом $y \in \Sigma$, то $f(x)$ является началом $f(y)$;

(2) значение отображения f на бесконечной последовательности ω является объединением (наименьшим общим продолжением) значений $f(x)$ на всех конечных началах x последовательности Ω.

Мы будем использовать обозначение $x \preccurlyeq y$ для отношения «x является началом y»; предполагается, что $x, y \in \Sigma$ могут быть и конечными, и бесконечными. Если $x \preccurlyeq y$ для бесконечной последовательности x, то $x = y$. Требование (1) представляет собой монотонность f с точки зрения частичного порядка $\preccurlyeq$ на Σ. Оно гарантирует, что значения $f(x)$ на конечных началах x последовательности ω будут продолжать друг друга; их объединение (точная верхняя грань в смысле $\preccurlyeq$-порядка) согласно требованию (2) должно совпадать с $f(\omega)$.

132 Покажите, что это определение соответствует стандартному понятию непрерывности для отображений топологических пространств, если Σ снабдить топологией раздела 4.4.3 (с. 104). [Указание: ср. аналогичный результат для непрерывных отображений пространства Σ в $\mathbb{N}_\perp$ в том же разделе.]

С каждым непрерывным отображением $f\colon \Sigma \to \Sigma$ свяжем множество Γ_f, которое естественно назвать *подграфиком* отображения f; его элементами являются пары слов $\langle x, y\rangle$, где x, y конечные слова, для которых $y \preccurlyeq f(x)$. Имеет место следующее простое наблюдение: для любого непрерывного f множество Γ_f обладает следующими тремя свойствами:

(1) $\langle x, \Lambda\rangle \in \Gamma_f$ для любого слова x;

(2) если $\langle x, y\rangle \in \Gamma_f$, то $\langle x', y'\rangle \in \Gamma_f$ при любых $x' \succcurlyeq x$, $y' \preccurlyeq y$;

(3) если $\langle x, y_1\rangle$ и $\langle x, y_2\rangle$ принадлежат Γ_f, то слова y_1 и y_2 сравнимы (одно из них является началом другого).

Первые два свойства очевидны, третье следует из того, что два слова, являющиеся началами конечной или бесконечной последовательности, сравнимы. Следующая теорема показывает, что непрерывные отображения однозначно задаются своими подграфиками.

Теорема 81. *Соответствие* $f \mapsto \Gamma_f$ *является взаимно однозначным соответствием между непрерывными отображениями* $\Sigma \to \Sigma$ *и множествами пар слов, обладающими свойствами* (1) – (3).

◀ Пусть дано множество F пар слов, обладающее свойствами (1) – (3). Требования (1) – (3) гарантируют, что для любого слова x множество тех y, при которых $\langle x, y\rangle \in F$, непусто и состоит из сравнимых слов. Их объединение (точную верхнюю грань) и будем считать значением $f(x)$. Свойство (2) гарантирует, что $x \preccurlyeq x'$ влечёт $f(x) \preccurlyeq f(x')$ (с ростом x рассмотренное множество слов y также растёт). Тем самым можно корректно определить $f(\omega)$ как объединение $f(x)$ для всех слов $x \preccurlyeq \omega$. Построенное отображение f будет непрерывно в смысле данного выше определения. Легко проверить, что построенное соответствие $F \mapsto f$ обратно к $f \mapsto \Gamma_f$. ▶

Непрерывное отображение $f\colon \Sigma \to \Sigma$ будем называть *вычислимым*, если соответствующее множество Γ_f перечислимо. (Вычислимые отображения по определению всегда непрерывны.)

Формально нам не требуется никакого «оправдания» для этого определения, и всё сказанное дальше о «машинно-зависимой» интерпретации вычислимых отображений нигде не потребуется. Тем не менее поучительно понять, какому виду машин (программ) соответствует такое определение.

5.4.2. Монотонные машины с неблокирующим чтением

Будем рассматривать программы, которые используют неблокирующее чтение со входа (можно прочесть очередной бит из очереди входных битов и проверить, пуста ли эта очередь). Такие программы подробно обсуждались в разделе 4.4.2, с. 102. Но теперь будем считать, что машина строит выход по битам, имея команду `OutputBit`(b) с битовым аргументом.

Выходная последовательность такой программы может быть конечной или бесконечной, и зависит, вообще говоря, не только от входной последовательности битов (нажатых клавиш «0» и «1»), но и от моментов нажатия клавиш. Будем называть программу *корректной*, если такой зависимости нет (моменты нажатия клавиш могут влиять на моменты появления битов на выходе, но не на выходную последовательность). Корректная программа задаёт некоторое отображение множества Σ в себя.

Теорема 82. *Любая корректная программа задаёт вычислимое отображение* (*в описанном выше смысле*); *всякое вычислимое отображение может быть задано некоторой корректной программой.*

◀ Пусть имеется некоторая корректная программа M и две входные последовательности x и x', причём $x \preccurlyeq x'$. Покажем, что $M(x) \preccurlyeq M(x')$, где через $M(z)$ обозначается выход программы M на входе z (не зависящий, по предположению, от

моментов времени подачи на вход битов последовательности z). При бесконечном x это заведомо так (поскольку $x = x'$). Пусть x конечно. Подадим x на вход и будем ждать, пока на выходе не появится $M(x)$. Это рано или поздно должно случиться, если $M(x)$ конечно, после чего мы подадим на вход недостающие символы из x'. Полученная после этого на выходе последовательность $M(x')$ неизбежно будет продолжением $M(x)$.

Столь же ясно, что для бесконечного ω значение $M(\omega)$ есть объединение значений $M(x)$ для конечных $x \preccurlyeq \omega$, поскольку в каждый момент вычисления на вход подано лишь конечное число битов последовательности ω.

Множество пар слов $\langle x, y\rangle$, для которых $y \preccurlyeq M(x)$, перечислимо, так как его можно перечислять, моделируя работу машины M на всевозможных входах. Таким образом, каждой корректной машине соответствует некоторое вычислимое (в абстрактном смысле) отображение.

Напротив, пусть имеется произвольное вычислимое отображение f. Построим машину M, которая корректно вычисляет его. А именно, M перечисляет подграфик Γ_f и параллельно с этим читает биты со входа. Как только в Γ_f обнаружится некоторая пара $\langle x, y\rangle$, для которой x является началом входа, мы выдаём недостающие биты слова y (требования (2) и (3) гарантируют, что все обнаруживающиеся слова y будут сравнимы, так что отзывать биты с выхода не придётся). ►

5.4.3. Перечислимость множества вычислимых отображений

Определение вычислимости с корректными программами кажется более естественным, но оно (как и для случая префиксно корректных программ) имеет важный недостаток: нет алгоритма, позволяющего по данной программе определять, будет ли она корректной. Тем не менее можно алгоритмически преобразовать каждую программу в корректную: надо перейти от неё к множеству пар, «скорректировать» это множество и затем обратно перейти к программе. Мы не будем это описывать подробно, поскольку корректные программы для нас — всего лишь мотивировка понятия вычислимого отображения, а ограничимся утверждением о перечислимости множества вычислимых отображений, которое нам понадобится в дальнейшем.

Теорема 83. *Существует перечислимое множество U троек $\langle n, x, y\rangle$ (где n — натуральные числа, а x и y — слова) с такими свойствами*:

(а) *при любом n множество $U_n = \{\langle x, y\rangle \mid \langle n, x, y\rangle \in U\}$ является подграфиком некоторого вычислимого отображения $u_n\colon \Sigma \to \Sigma$ (то есть удовлетворяет свойствам* (1)–(3) *теоремы* 81);

(б) *среди отображений u_n встречаются все вычислимые отображения Σ в себя.*

◄ Рассмотрим универсальное множество W троек того же вида, среди сечений W_n которого встречаются все перечислимые множества пар слов. Далее изменим W, «скорректировав» все сечения W_n. Мы хотим, чтобы после коррекции сечение заведомо удовлетворяло свойствам (1)–(3) (и тем самым было подграфиком вычислимого отображения по теореме 81), и чтобы корректные сечения при этом не менялись.

Коррекция состоит из двух этапов: сначала устраняются «противоречия», а затем восполняются «пробелы». Противоречие образуют две пары $\langle x_1, y_1\rangle$ и $\langle x_2, y_2\rangle$, в которых x_1 сравнимо с x_2, а y_1 не сравнимо с y_2. (Легко видеть, что в подграфике таких пар быть не может.) Противоречия устраняются естественным способом: выбрасываются пары, противоречащие ранее появившимся (и не выброшенным). Полученное множество перечислимо. После этого восполняются пробелы: добавляются все пары вида $\langle x, \Lambda\rangle$; если в множестве есть пара $\langle x, y\rangle$, то добавляются также и все пары $\langle x', y'\rangle$ с $x' \succcurlyeq x$ и $y' \preccurlyeq y$. Восполнение пробелов также сохраняет перечислимость; легко проверить, что полученное после этого множество удовлетворяет нашим требованиям. ►

Это утверждение, которое можно назвать перечислимостью множества всех вычислимых отображений Σ в себя, будет использовано при построении монотонной сложности в следующем разделе.

5.5. Монотонная сложность

При определении монотонной сложности в качестве способов описания (декомпрессоров) рассматриваются вычислимые отображения $D\colon \Sigma \to \Sigma$. *Монотонной сложностью* слова x при данном способе описания D называется наименьшая длина слова y, для которого $x \preccurlyeq D(y)$. Она обозначается $KM_D(x)$.

(Это определение безо всяких изменений переносится и на бесконечные последовательности x, но мы следуем традиции и рассматриваем лишь двоичные слова, если это не оговорено особо.)

$\boxed{\mathbf{133}}$ Докажите, что естественным образом определённая сложность бесконечной последовательности равна пределу неубывающей последовательности сложностей её начальных отрезков.

Теорема 84. *Существует оптимальный способ описания, то есть вычислимое отображение $D\colon \Sigma \to \Sigma$, для которого KM_D минимальна с точностью до константы: для всякого вычислимого $D'\colon \Sigma \to \Sigma$ найдётся такая константа c, что*

$$KM_D(x) \leqslant KM_{D'}(x) + c$$

для любого слова x.

◄ Пусть U — множество троек, сечениями которого являются все подграфики вычислимых отображений и только они (теорема 83, с. 148). Пусть D_n — вычислимое отображение, соответствующее сечению U_n. Определим отображение D формулой

$$D(\hat{n}z) = D_n(z),$$

где $\hat{n}$ — беспрефиксный код числа n (скажем, его двоичная запись с удвоенными цифрами, за которой следует 01), а z — произвольный элемент Σ. В терминах подграфика: рассмотрим множество всех пар $\langle \hat{n}u, v\rangle$, для которых $\langle n, u, v\rangle \in U$. Легко проверить, что действительно получается вычислимое отображение и что если способ описания D' имеет номер n (подграфик D' совпадает с U_n), то $KM_D(x) \leqslant KM_{D'}(x) + l(\hat{n})$ при всех x. ►

Как обычно, мы фиксируем некоторый оптимальный «монотонный способ описания» (вычислимое отображение D, для которого выполнено утверждение этой теоремы) и *монотонной сложностью* слова x называем $KM_D(x)$. Обозначение: $KM(x)$. (В англоязычной литературе, в частности, в книге Ли и Витаньи [85], используется обозначение $Km(x)$, а $KM(x)$ обозначает априорную сложность, $KA(x)$ в наших обозначениях.)

Теорема 85. (а) *Монотонная сложность монотонна*: $KM(x) \leqslant KM(y)$ *при* $x \preccurlyeq y$;

(б) *функция* KM *перечислима сверху*;

(в) $KM(x) \leqslant l(x) + O(1)$;

(г) $KM(x) \leqslant KP(x) + O(1)$;

(д) $KA(x) \leqslant KM(x) + O(1)$;

(е) *бесконечная последовательность нулей и единиц вычислима тогда и только тогда, когда монотонная сложность её начальных отрезков ограничена*;

(ё) *если* $f\colon \Sigma \to \Sigma$ — *вычислимое отображение, то* $KM(f(x)) \leqslant KM(x) + O(1)$ (*скрытая в* $O(1)$ *константа может зависеть от* f, *но не от* x);

(ж) *если* $f\colon \Sigma \to \mathbb{N}_\perp$ — *вычислимое отображение, то* $KP(f(x)) \leqslant KM(x) + O(1)$ (*скрытая в* $O(1)$ *константа может зависеть от* f, *но не от* x).

Поучительно сравнить эти утверждения со свойствами априорной сложности (теорема 79, с. 141). Поскольку монотонная сложность не меньше априорной (утверждение (д)), некоторые свойства априорной сложности сразу же переносятся на монотонную. В частности, мы получаем, что $KP(x \,|\, l(x)) \leqslant KM(x) + O(1)$ и $KP(x) \leqslant KM(x) + KP(l(x)) + O(1)$. Кроме того, отметим, что для вычислимых последовательностей несравнимых слов префиксная и априорная сложности совпадают с точностью до $O(1)$ — и потому совпадают с находящейся между ними монотонной: если $x_0, x_1, \ldots$ — такая последовательность, то $KM(x_i) = KA(x_i) + O(1) = KP(x_i) + O(1)$.

◄ Утверждение (а) непосредственно следует из определения: если $D(u) \succcurlyeq y$, то и $D(u) \succcurlyeq x$ для любого начала x слова y. Можно сказать, что при определении монотонной сложности требуется описать не само слово, а любое из его продолжений, и при удлинении слова эта задача усложняется (множество продолжений становится меньше).

Утверждение (б) сразу следует из определения: подграфик вычислимого отображения перечислим, поэтому перечислимо и множество троек $\langle x, y, r\rangle$, для которых $l(y) < r$ и пара $\langle y, x\rangle$ принадлежит подграфику D, а значит, перечислим и надграфик функции KM как проекция этого множества.

Чтобы доказать (в), достаточно заметить, что тождественное отображение $\Sigma \to \Sigma$, для которого $D(x) = x$ при всех $x \in \Sigma$, вычислимо.

Чтобы сравнить KM и KP (пункт (г) теоремы), достаточно заметить, что любое вычислимое отображение $\Sigma \to \mathbb{N}_\perp$ можно переделать в отображение $\Sigma \to \Sigma$ (отобразив $\mathbb{N}_\perp$ в Σ: элемент $\perp$ переходит в пустое слово). Более формально, если D — оптимальный префиксно корректный способ описания, используемый при определении KP, то его же можно считать вычислимым отображением $\Sigma \to \Sigma$, до-

определив пустым словом на тех аргументах, где D не был определён, и продолжив на бесконечные последовательности по непрерывности.

Для сравнения KM и KA (пункт (д) теоремы) надо вспомнить, с чего мы начинали обсуждение вычислимых отображений: мы говорили, что вероятностный алгоритм представляет собой датчик случайных битов, к выходу которого применяется вычислимое отображение пространства Σ в себя. Пусть D — оптимальный способ описания, используемый при определении монотонной сложности. Рассмотрим вероятностную машину, которая состоит в применении D к последовательности случайных битов. Нас интересует вероятность того, что на её выходе появится слово x или какое-то его продолжение. Ясно, что эта вероятность не меньше $2^{-l(y)}$ для любого слова y, при котором $D(y) \succcurlyeq x$, поскольку датчик случайных битов с вероятностью $2^{-l(y)}$ выдаст последовательность, начинающуюся на y, а применение D после этого даст последовательность, начинающуюся на x. (Подробнее о сравнении KM и KA см. формулировку теоремы 87 и её обсуждение.)

Утверждение (е) в одну сторону прямо следует из соответствующего утверждения теоремы 79, а в другую очевидно: начальные отрезки вычислимой последовательности ω имеют ограниченную монотонную сложность, поскольку можно рассмотреть вычислимое отображение $\Sigma \to \Sigma$, равное ω на всех аргументах.

В пункте (ё) допусти́м и случай, когда $f(x)$ бесконечно (если стремиться этого избежать, то надо говорить о сложности всех конечных начал $f(x)$). Для доказательства достаточно рассмотреть способ описания, являющийся композицией оптимального (использованного при определении монотонной сложности) способа и отображения f.

Наконец, в пункте (ж) надо заметить, что композиция оптимального для монотонной сложности способа описания и f является префиксно корректным способом описания. Кроме того, это можно вывести из аналогичного утверждения про априорную сложность. ►

134 Докажите, что $KM(xy) \leqslant KP(x) + KM(y) + O(1)$ (здесь xy — конкатенация слов x и y). В частности, $KM(xy) \leqslant KP(x) + l(y) + O(1)$. [Указание. Рассмотрим оптимальный беспрефиксный способ описания D_p и оптимальный монотонный способ описания D_m. Теперь положим $D'(uv) = D_p(u)D_m(v)$ (когда D_p прочитает всё, что захочет, непрочитанная часть рассматривается как вход для D_m).]

135 Покажите, что в предыдущей задаче можно заменить $KM(y)$ в правой части на «условную» монотонную сложность $KM(y\,|\,x)$, определив её естественным образом (монотонности по аргументу x не требуется; подробнее см. в главе 6).

136 Докажите, что утверждение (ё) остаётся верным, если заменить KM на KA (в обеих частях). [Указание: отображение f можно применять к выходу вероятностной машины; получится новая вероятностная машина, которая не хуже оптимальной.]

Определение монотонной сложности можно дать и формально более простым (хотя, на наш взгляд, менее естественным) образом. Вот как это делается. Напомним, что через Ξ мы обозначаем множество всех двоичных слов. На этом множестве рассмотрим отношение «быть сравнимым»: x сравнимо с y, если $x \preccurlyeq y$ или

$y \preccurlyeq x$. Перечислимое отношение $D \subset \Xi \times \Xi$ будем называть *корректным*, если оно обладает таким свойством: для всех x_1, x_2, y_1, y_2

$$\langle x_1, y_1\rangle \in D,\ \langle x_2, y_2\rangle \in D \text{ и } (x_1 \text{ сравнимо с } x_2) \Rightarrow (y_1 \text{ сравнимо с } y_2).$$

Далее монотонную сложность слова y относительно D определяем как минимальную длину x, при котором $\langle x, y\rangle \in D$, и доказываем утверждение о существовании оптимального корректного перечислимого отношения.

$\boxed{\textbf{137}}$ Покажите, что такой подход приводит к величине монотонной сложности, отличающейся от определённой нами не более чем на ограниченное слагаемое. [Указание. Подграфик всякого вычислимого отображения $\Sigma \to \Sigma$ является корректным множеством в только что описанным смысле. Напротив, если множество D корректно, то после «восполнения пробелов» (см. доказательство теоремы 83) оно превращается в подграфик вычислимого отображения.]

Сравнивая это определение с определением обычной колмогоровской сложности (где в качестве D берутся графики вычислимых функций, то есть перечислимые и однозначные по второму аргументу множества), можно увидеть отличия между KS и KM. В данном случае мы не требуем однозначности: при одном и том же x в D могут быть пары $\langle x, y\rangle$ с разными y; нужно лишь, чтобы все эти y были согласованы, то есть были различными началами какой-то одной последовательности нулей и единиц. За счёт этого сложность уменьшается. Например, это позволяет всем началам данной вычислимой последовательности иметь ограниченную сложность (скажем, $KM(0^n)$ ограничено, в то время как $KS(0^n) = KS(n)$ может достигать логарифма n).

С другой стороны, мы накладываем и дополнительные ограничения: если x является описанием какого-то y, то сравнимые с x слова могут описывать лишь слова, сравнимые с y. За счёт этого сложность увеличивается. Это особенно хорошо видно, когда мы говорим о несравнимых словах (элементах вычислимой последовательности): монотонная сложность при этом обращается в префиксную, и разница между ней и обычной сложностью также может быть примерно равна $\log n$ для слов длины n.

Суммируя эти наблюдения (и вспоминая, что и априорная сложность, и обычная сложность отличаются от префиксной не более чем на $O(\log n)$ для слов длины n), приходим к такому утверждению:

Теорема 86. *Разница между $KS(x)$ и $KM(x)$ не превышает $O(\log n)$ для слов длины n и может достигать $\log n - O(1)$ в обе стороны для слов длины n при бесконечно многих n.*

Мы вернёмся к вопросу о связи различных вариантов определения сложности в главе 6. Сейчас мы приведём лишь одно утверждение такого рода:

Теорема 87. *Величина $KM(x) - KA(x)$ не ограничена сверху; более того, для бесконечно многих n найдется слово длины n, для которого эта разность не меньше $\log\log n - O(\log\log\log n)$.*

Это доказал недавно А. Дей [36], усилив более ранний результат П. Гача о том, что разность не ограничена сверху [45]. Обе работы основаны на сведении вопроса

к некоторой игре, в которой указывается выигрышная стратегия для одного из игроков.

Вспомним, что и при определении $KM(x)$, и при определении $KA(x)$ выбирается некоторое вычислимое непрерывное отображение $f\colon \Sigma \to \Sigma$. Затем рассматривается прообраз множества всех продолжений слова x при этом отображении. Разница между KA и KM состоит в том, что в первом случае нас интересует мера этого прообраза, а во втором случае — мера наибольшего интервала вида Σ_y, содержащегося в этом прообразе. Из этого описания видно, что $KA_f \leqslant KM_f$, и разница может быть большой, если прообраз «разреженный». Вопрос в том, насколько это возможно для оптимального способа описания.

Аналогичная проблема возникала при сравнении префиксной сложности с логарифмом априорной вероятности на изолированных объектах. Речь шла о выделении места на отрезке $[0,1]$ счётному числу клиентов, см. доказательства теорем 46 (с. 91) и 58 (с. 108). Разница между префиксной сложностью и логарифмом априорной вероятности на $\mathbb{N}$ также возникала из-за того, что в одном случае нас интересовал общий размер прообраза, а в другом случае — длина наибольшего непрерывного участка. Однако тогда за счёт перехода к другому способу описания удавалось обойтись лишь постоянным множителем.

Теперь задача усложняется за счёт того, что клиенты объединены в иерархическую структуру типа дерева. Из-за этого дополнительное ужесточение требований (учёт лишь максимального непрерывного участка памяти, а не общего её количества) приводит к бóльшим потерям памяти.

5.5.1. Доказательство теоремы Гача – Дея

Сначала аккуратно определим используемую игру двух игроков: *Заказчика* и *Менеджера*. Чтобы определить игру, фиксируем некоторое дерево с корнем и рациональное число $d \geqslant 1$. Каждой вершине дерева в каждый момент игры поставлено в соответствие неотрицательное рациональное число, которое мы называем *запросом* этой вершины (в этот момент). При этом запрос любой вершины всегда должен быть не меньше суммы запросов всех её сыновей, а запрос корня дерева — не больше $1/d$.

Запросы делает Заказчик; в ответ Менеджер выделяет место в канторовском пространстве. В каждый момент игры для каждой вершины дерева известно, какое подмножество пространства Ω ей выделено. Это подмножество должно быть объединением конечного числа интервалов (множеств вида Ω_x). Множество, выделенное вершине, должно содержать множества, выделенные её детям, а эти последние не должны пересекаться. (Отсюда следует, что несравнимым вершинам соответствуют непересекающиеся множества.)

Игроки делают ходы по очереди. В начале игры все запросы равны нулю, и все выделенные множества пусты. На очередном ходу Заказчик может увеличить запрос любой вершины дерева (или сразу нескольких), соблюдая ограничения из предыдущего абзаца. Уменьшать запросы нельзя. В свою очередь, Менеджер может увеличивать выделенные вершинам множества. Цель его — удовлетворить запросы в таком смысле: выделенное вершине i множество должно содержать интервал

длины не меньше запроса этой вершины. (Как мы уже говорили, тут важно не общее выделенное место, а сплошная его часть.)

Менеджер проигрывает, если в какой-то момент не может удовлетворить очередной запрос заказчика; если игра продолжается бесконечно, мы считаем, что Менеджер выиграл.

Можно представлять себе Заказчика как иерархически устроенную организацию, которой нужно место: всей организации требуется сколько-то места, при этом известно, сколько из него надо отдать разным лабораториям. Внутри места для лаборатории надо выделить место для её секторов, внутри места для каждого сектора — место для его отделов, и т. д. Перераспределять место нельзя, и учитывается только «сплошное» пространство (максимальный размер интервала, а не сумма их размеров).

Теорема 88 (Гача – Дея). *Для каждого $d \geqslant 1$ для дерева T глубины $O(d)$ и ветвления в каждой вершине $2^{O(d)^{O(d)}}$ Заказчик имеет вычислимую выигрышную стратегию.* (*Вычислимость означает, что есть алгоритм, который, получив на вход d, реализует эту стратегию.*)

Гач доказал аналогичный результат для деревьев бесконечного ветвления глубины $O(d)$, а также для деревьев конечного, но очень большого ветвления. Вклад Дея заключается в улучшении конструкции Гача, позволяющем значительно уменьшить степень ветвления дерева.

Объясним, как из теоремы Гача – Дея следует теорема 87. Из описания игры видно, что увеличение ветвления и высоты дерева облегчает задачу Заказчика. Кроме того, ветвление можно заменить высотой: скажем, дерево с корнем и 2^n его сыновьями можно вложить в бинарное дерево глубины n (суммируя запросы для промежуточных вспомогательных вершин). Дерево из теоремы Гача – Дея можно вложить в бинарное дерево глубины $O(d)^{O(d)}$, так что и для этого бинарного дерева Заказчик может выиграть.

Положим $d = 2^c$, где c — некоторое натуральное число. По теореме Гача – Дея при таких значениях параметров Заказчик имеет вычислимую выигрышную стратегию на бинарном дереве глубины $2^{O(c2^c)}$.

Запустим эту стратегию, а в качестве стратегии Менеджера применим алгоритм перечисления подграфика оптимального вычислимого отображения $f\colon \Sigma \to \Sigma$ из определения монотонной сложности. Точнее, по ходу игры Менеджер перечисляет все пары $\langle y, x\rangle$, для которых $x \preccurlyeq f(y)$, то есть y является описанием x, выделяя вершине x интервал Ω_y и не обращая внимания на x за пределами дерева (когда длина x больше $2^{O(c2^c)}$). Как только все текущие запросы окажутся удовлетворенными, он сообщает Заказчику, что ход сделан (а после ответного хода Заказчика он продолжит перечисление и выделение места, пока не будет удовлетворён и новый запрос). Если этого (удовлетворения всех запросов) не произойдёт, то это значит, что Менеджер так и не сделает ход и проиграет. (Эта стратегия Менеджера, так сказать, соответствует идеалам планового хозяйства: не имеет значения, что просят потребители, промышленность делает то, что запланировано, пока потребители не окажутся удовлетворены — если окажутся.)

Теорема гарантирует, что Менеджер в некоторый момент не сможет сделать хода. Это означает, что существует слово x длины не больше $2^{O(c2^c)}$, запрос которого в некоторый момент становится больше $2^{-KM(x)}$. С другой стороны, в ходе игры Заказчик перечисляет снизу некоторую полумеру μ_c на дереве: значение на x равно верхней грани всех запросов вершины x. Поскольку Заказчик в некоторый момент выигрывает, для этой полумеры для некоторого x выполнено $KM(x) > -\log\mu_c(x)$. Сумма полумер μ_c с весами $2^c/c^2$ является перечислимой снизу полумерой. Поэтому априорная полумера на дереве не меньше, чем $\varepsilon\mu_c(x)2^c/c^2$ (для некоторого положительного ε и всех c, x). Отсюда следует, что для каждого c найдется слово x длины не больше $2^{O(c2^c)}$, для которого

$$KM(x) > KA(x) + c - 2\log c - O(1).$$

Обозначим длину этого слова через n. Тогда c не меньше $\log\log n - O(\log\log\log n)$, а значит

$$KM(x) > KA(x) + \log\log n - O(\log\log\log n).$$

Заметим, что приведённое рассуждение позволяет построить слово с нужным свойством лишь для бесконечно многих n, но не для всех n.

138 Покажите, что для бесконечно многих n выполнено такое свойство: для всех слов x длины n разница $KM(x) - KA(x)$ не превосходит $\log\log\log n$. Здесь тройной логарифм можно заменить на любую неубывающую вычислимую функцию, стремящуюся к бесконечности. [Указание: если объявить какие-то очень редкие длины особо важными, то для слов каждой из этих длин можно выделять место по отдельности, и превышение будет не очень большим.]

А что будет
ства n?

◄ Изложение доказательства теоремы Гача–Дея начнём с неформального обсуждения. Какая главная трудность возникает у Менеджера при выделении места? Пусть Заказчик запросил очень небольшое (но ненулевое) место для некоторой вершины. У Менеджера есть выбор: позаботиться о том, чтобы рядом с выделенным местом был запас свободного пространства (соседние интервалы не выделены никому другому), или не делать этого.

В первом случае есть риск, что Заказчик не будет увеличивать запрос для этой вершины и пустое пространство рядом с её владением окажется потерянным — представим себе, например, что в дальнейшем все запросы будут больше по размеру, чем этот свободный кусочек. Во втором случае, если Заказчик увеличит запрос вершины, занятые соседние участки помешают расширить выделенное пространство и оно окажется потраченным впустую.

Наша стратегия будет эксплуатировать эту трудность, контролируя размер ε-окрестности владений различных вершин. Пусть дано $\varepsilon \leqslant 1$, являющееся целой степенью двойки. Будем называть *ε-окрестностью* подмножества X в канторовском пространстве объединение всех интервалов длины ε, пересекающихся с множеством X. Заказчик будет добиваться, чтобы отношение окрестности выделенного пространства к запросу было бо́льшим (скажем, не меньше некоторого k); поскольку размер окрестности не больше 1, то при запросе больше $1/k$ Менеджер не сможет это поддерживать и проиграет. В соответствии с этим планом мы будем строить

стратегии для Заказчика, добивающиеся того, что в некоторый момент игры запрос корня не превосходит некоторого α, а длина ε-окрестности владений корня не меньше некоторого β. (Эту окрестность мы будем называть «надкушенным пространством» по аналогии с фразой «не съем, так хоть понадкусываю» из анекдота про жадину.)

Более точно, мы будем строить стратегии Заказчика, исходя из следующих параметров:

- дерево, на котором происходит игра;
- ε, указывающий, как измеряется окрестность выделенного множества;
- максимально разрешённый запрос α для корня;
- требуемый размер β надкушенной части (то есть ε-окрестности выделенного пространства).

При некоторых соотношениях параметров выигрышная стратегия для Заказчика существует, то есть он может добиться, чтобы ε-окрестность выделенного Менеджером пространства (для корня дерева — остальные вершины получают части) была не меньше β, при этом запрос для корня не превосходит α.

Пример 1. Пусть ε — степень двойки, $\beta = \varepsilon$, а значение α положительно и много меньше β. Тогда стратегия существует, даже если дерево состоит только из корня: запросим α; что бы нам ни выделили, это непустое множество будет иметь ε-окрестность размера не меньше ε.

Таким образом, получить большой коэффициент «полезного действия» (отношение β/α) несложно, если β мало. Сложность возникает, когда $\beta \gg \varepsilon$, и для этого нам понадобиться строить стратегии индуктивно, соединяя стратегии для разных деревьев и используя внутри одной стратегии рекурсивные вызовы других (на поддеревьях). При индуктивном построении нам будет полезно учитывать отдельно коэффициент полезного действия, добавив ещё один параметр k и требуя, чтобы отношение надкушенного к запрошенному было не меньше k. При $k \leqslant \beta/\alpha$ это выполнено автоматически, но нам будет полезно иметь стратегии, гарантирующие данный коэффициент полезного действия, но оставляющие некоторую свободу для размера запрошенного.

Пример 2. Пусть дерево T имеет m сыновей корня, каждый из которых имеет двух сыновей (внуков корня). Пусть ε — некоторая степень двойки, $\alpha = \beta = m\varepsilon$, а $k = 3/2$ (так что в данном случае k существенно: мы не знаем точного размера запрошенного и надкушенного, но знаем, что второе должно быть минимум в полтора раза больше первого, а $m\varepsilon$ должно лежать между ними).

Стратегия в этом примере строится так. Чтобы поставить Менеджера в трудное положение, выберем у каждого сына корня по одному сыну и запросим для них по $\varepsilon/2$. (Мы указываем запросы для листьев; для остальных вершин они получаются сложением запросов для потомков.) Менеджер должен будет решить, кого из внуков размещать в паре со своим двоюродным братом (выделив им по $\varepsilon/2$ внутри одного интервала размера ε), а кого помещать «поодиночке в двухместный номер» (сохраняя соседний интервал как резерв для брата). Для тех внуков, у которых нет резерва, закажем $\varepsilon/2$ и для их родных братьев — тогда запрос отца этих братьев

достигнет ε и придётся выделить новый интервал размера ε (потому что в старом не было резерва; резервы в других местах тоже малы). Таким образом, для любого сына корня происходит одно из двух: либо при заказе $\varepsilon/2$ было выделен резерв ε, либо при заказе в ε пришлось выделить $(3/2)\varepsilon$. В обоих случаях достигнут коэффициент $3/2$.[1]

Попробуем теперь соединить две стратегии с одним и тем же гарантированным коэффициентом полезного действия k. (Наша цель — сохранить этот коэффициент, но увеличить размер запрошенного и надкушенного.) Пусть дерево T состоит из корня из двух поддеревьев T_1 и T_2. Пусть на дереве T_1 для каких-то $\alpha_1, \beta_1, \varepsilon$ (и выбранного k) есть выигрышная стратегия (для Заказчика, как всегда), а на дереве T_2 есть стратегия с другими параметрами α_2, β_2 и теми же ε и k. Будем применять эти стратегии последовательно: сначала на T_1 применим первую стратегию, а когда будет надкушено достаточно (первая стратегия выиграет), применим на T_2 вторую. (Без ограничения общности можно считать, что во время запросов на T_1 для вершин из T_2 ничего не выделяется, так как это в любом случае можно отложить.) В такой ситуации общий запрос будет не больше $\alpha_1 + \alpha_2$, так как запросы складываются. Беда в том, что надкушенное пространство, в отличие от выделенного, не складывается: размер ε-окрестности объединения двух множеств может быть меньше суммы размеров ε-окрестности каждого из них. Например, во второй игре менеджер может использовать резерв внутри интервалов, надкушенных, но не занятых целиком в первой игре.

Это можно предотвратить, рассматривая разные значения ε_1 и ε_2 для первой и второй стратегии и взяв $\varepsilon_1 \ll \varepsilon_2$. Если при этом вторая стратегия будет запрашивать только интервалы размера ε_1 и больше, то внутри ε_1-окрестности владений дерева T_1 ей ничего выделено быть не может. Неформально говоря, мы накапливаем резервы на разных уровнях (сначала малого размера, потом большого). Правда, остаётся проблема в обратную сторону: хотя владения T_2 теперь не пересекают ε_1-окрестность владений T_1, но владения T_1 вполне могут попадать в ε_2-окрестность владений T_2, и сложения опять не получится. Чтобы выйти из положения, мы снова обобщим игру и будем считать, что перед началом игры какое-то подмножество $A \subset \Omega$ считается *недоступным* менеджеру, и оценивать *новые* надкушенные интервалы. Более точно это описывается так.

В окончательном варианте игра задаётся такими параметрами:

- дерево T, на котором происходит игра;
- подмножество $A \subset \Omega$ (недоступное менеджеру);
- δ (указывающий, как измеряется окрестность недоступного множества);
- ε (указывающий, как измеряется окрестность выделенного множества);

[1] В принципе, достигнуть коэффициента $3/2$ легко, запросив для каждого сына чуть больше $\varepsilon/2$: поскольку выделять можно только интервалы, их размеры будут степенями двойки и менеджеру придётся выделить ε. Однако этот эффект, в отличие от описанного выше, не «масштабируется» (не позволяет увеличивать коэффициент за счёт рекурсивных вызовов), так что мы его использовать и о нём говорить не будем. Ещё можно было бы обойтись без внуков, сначала заказав для детей корня по $\varepsilon/2$, а потом увеличив заказ до ε в тех вершинах, где нет резерва. Мы выбрали вариант с внуками, так как он больше похож на будущую конструкцию для общего случая.

- максимально разрешённый (суммарный) запрос α;
- требуемый размер β;
- требуемый коэффициент полезного действия k.

Заказчик увеличивает запросы на дереве T (от нуля), минимальный ненулевой запрос δ, при этом запрос корня не может быть больше α. Менеджер выделяет место в канторовском пространстве, при этом не используя интервалы, пересекающиеся с A. Заказчик выигрывает, если *размер свеженадкушенной части* (ε*-окрестности выделенного минус* δ*-окрестность* A) *не меньше* β, *а его отношение к запросу корня не меньше* k (или если Менеджер не смог сделать хода, не нарушив правила игры). Как и раньше, увеличение дерева и ε, а также α упрощает задачу Заказчика. Она упрощается также и при уменьшении δ, β или k.

При таком определении обсуждение выше даёт такое утверждение:

Лемма о композиции. *Пусть для дерева* T_1 *и любого недоступного подмножества Заказчик может выиграть в игре с параметрами* $\varepsilon_1, \delta_1, \alpha_1, \beta_1, k_1$, *а для дерева* T_2 *и любого недоступного подмножества он может выиграть в игре с параметрами* $\varepsilon_2, \delta_2, \alpha_2, \beta_2, k_2$. *Пусть при этом* $\varepsilon_1 = \delta_2$ *и* $k_1 = k_2$ (*общее их значение обозначим* k). *Тогда на дереве* T, *состоящем из корня и двух поддеревьев* T_1 *и* T_2, *при любом недоступном подмножестве Заказчик может выиграть в игре с параметрами* $\varepsilon_2, \delta_1, \alpha_1 + \alpha_2, \beta_1 + \beta_2, k$.

Основной (но непосредственно следующий из определений) момент здесь в том что надкушенные (=«свеженадкушенные») в обеих играх части не пересекаются.

На самом деле, мы будем использовать не буквально утверждение леммы, а применять ту же идею в чуть более общей ситуации: деревья не будут заданы заранее, а будут выбираться в ходе игры (выбор следующего дерева будет зависеть от хода игры на предыдущем). Кроме того, мы будем соединять не две стратегии, а много. За это нам придётся заплатить большой разницей между значениями ε и δ в результирующей игре: в каждой следующей игре значение ε увеличивается по сравнению с предыдущим (которое для неё будет входным значением δ). Заметим также, что соединение стратегий не приводит к увеличению коэффициента полезного действия, которого нужно добиться как-то иначе.

Прежде чем показать на примере, как увеличивать коэффициент полезного действия, сделаем одно простое техническое замечание. Без ограничения общности можно потребовать, чтобы в конце игры запрос корня был не меньше β/k. В самом деле, если в результате какой-то непредвиденной экономии он оказался меньше, можно в конце игры искусственно увеличить запрос, не нарушив условий выигрыша (в плановой экономике этот приём назывался «освоением средств»).

Пример 3. Покажем, как можно добиться коэффициента полезного действия $k > 2$. Мы будем действовать как в примере 2, но вместо прямого заказа $\varepsilon/2$ для внуков мы будем рекурсивно вызывать стратегию того же примера 2.

Прежде всего вспомним, чего мы можем добиться с помощью стратегии примера 2. Для данного α и сколь угодно малого ε, делящего α нацело, можно построить стратегию на дереве глубины 2, которая при $\delta = \varepsilon/2$ для любого множества A позволяет получить не менее α (свеже)надкушенного при не более чем α запрошенного с коэффициентом полезного действия не менее 1,5. Для этого достаточно

вспомнить стратегию и понять, что при $\delta = \varepsilon/2$ наличие запрещённого множества не помешает. Дерево имеет $m = \alpha/\varepsilon$ сыновей корня, каждый из которых имеет двух сыновей. Условие делимости нам в дальнейшем не помешает, так как мы будем применять стратегию примера 2 только для α, кратных ε.

Если мы хотим применять к таким стратегиям лемму о композиции, нам придётся позаботиться о согласовании параметров ε: в каждой следующей стратегии параметр должен быть вдвое больше, чем в предыдущей. Важно, что при этом мы можем использовать одно и то же α, конструкция примера 2 позволяет развязать выбор размера запроса и выбор масштаба измерения окрестности. Если знать заранее максимальное число применений (число стратегий, подвергаемых композиции), то можно понять, с какого значения ε начинать. Теперь о значении ε для игр в композиции уже можно не говорить, и мы будем использовать ε для обозначения параметров основной игры (где мы хотим достичь $k > 2$, измеряя ε-окрестность выделенного).

В основной игре мы тоже будем действовать по аналогии с примером 2. Мы обеспечим коэффициент $k > 2$ при любом заданном ε, при любом $\alpha = m\varepsilon$ и при достаточно малом δ, с максимальным размером запроса и минимальным размером надкушенного, равным $\alpha = m\varepsilon$.

Как же мы играем? Корень имеет m сыновей, как и раньше. Теперь нам понадобится больше внуков. Договоримся, скажем, что каждый сын корня имеет в свою очередь 12 сыновей (этого хватит). Вместо того, чтобы просто заказывать для внуков интервалы какого-то размера, мы будем вызывать для них рекурсивно описанные выше стратегии: над каждым внуком будет поддерево высоты 2, так что всего высота дерева будет 4.

Итак, мы в каждый момент смотрим, какие из сыновей корня не имеют *резерва*, понимая под резервом интервал размера ε, в котором им что-то выделено, ничего не выделено другим вершинам (не считая их потомков и предков) и нет точек запрещённого множества. Другими словами, резерв — это интервал, который (1) уже частично занят и (2) ещё может быть использован в будущем, когда запрос увеличится до ε. Выберем одного из таких (не имеющих резерва) сыновей и запустим на нём (а точнее, на некотором ещё не обработанном его сыне) стратегию с $\alpha = \varepsilon/8$. При этом будет запрошено не больше α и, можно считать, не меньше $\alpha/1{,}5 = \varepsilon/12$ (см. выше об освоении средств). Может быть, после этого появится резерв (а может, и не появится, если выделенные интервалы будут рядом с множеством A). Так или иначе, мы повторяем это действие, пока для всех сыновей корня не появятся резервы — с одним уточнением: если уже запрошенное каким-то из них место превысило $(7/8)\varepsilon$ и есть опасность, что следующий запрос пересечёт границу в ε, мы нового запроса не делаем, а добиваемся резерва, просто увеличив заказ этого интервала до ε. Поскольку при каждом вызове стратегии для некоторого сына его запрос увеличивается не менее чем на $\varepsilon/12$, нам хватит двенадцати сыновей у каждого сына корня. По этой же причине общее количество вызовов стратегии будет ограничено некоторым заранее известным числом ($12m$).

Почему эта стратегия даст хороший результат? Для каждого сына корня рассмотрим последнее обращение к нему, после которого образовался резерв. Этот

резерв имеет размер ε и до этого обращения резервом не был. Раз он смог стать резервом, значит, ничего постороннего в нём нет (и не было); раз он резервом не был, значит, он был совершенно пустым. Вывод: то, что было надкушено при всех не-последних обращениях, в резервы не вошло. Оно дало выигрыш в полтора раза, да плюс ещё почти пустые резервы — в итоге получается выигрыш с $k = 20/9 > 2$, если подсчитать аккуратно.

Вот этот подсчёт: обозначим через γ общий запрос при не-последних обращениях. Он даёт вклад $(3/2)\gamma$ в надкушенное, лежащий вне резервов, всего получается $(3/2)\gamma + m\varepsilon$ надкушенного при максимум $\gamma + m(\varepsilon/8)$ запрошенного. (Можно сказать, что часть увеличивается в $3/2$ раза, а часть в 8 раз, и притом вторая часть не очень мала по сравнению с первой, так что в целом коэффициент увеличивается). Технически, из $\gamma \leqslant m\varepsilon$ следует

$$\left[\frac{3}{2}\gamma + m\varepsilon\right] \geqslant \frac{20}{9}\left[\gamma + m\frac{\varepsilon}{8}\right]$$

(что нетрудно проверить, раскрыв скобки и приведя подобные члены). Таким образом, обещанная стратегия с $\alpha = \beta = m\varepsilon$ и $k = 20/9 > 2$ построена.

Теперь общая схема должна быть уже более или менее понятна. Имея стратегию с $k = 20/9$, мы можем её рекурсивно вызывать для внуков корня (тем самым перейдя к дереву глубины 6 с большим ветвлением — можно даже оценить, каким). Такой переход, если действовать аккуратно, может увеличить k почти на 1. В самом деле, если взять достаточно малую долю ε вместо $\varepsilon/8$, то потерями на последних шагах (с образованием резерва) можно пренебречь, и мы получаем резерв $m\varepsilon$ в дополнение к индуктивно полученному увеличению в k раз. При этом, если общий запрос γ близок к максимуму $m\varepsilon$, то резерв прибавляет к коэффициенту полезного действия почти единицу, а если общий запрос оказался меньше, то нам же лучше. Чтобы оценить сверху число внуков корня, надо вспомнить, что мы можем (методом освоения средств) обеспечить увеличение запроса на некоторую фиксированную долю ε. Таким образом, при дереве глубины $O(k)$ и достаточно большого ветвления мы можем обеспечить превышение надкушенного над запрошенным в k раз, что достаточно для теоремы Гача (но не Дея).[1]

Остаётся более аккуратно объяснить детали этого рассуждения. Индукцией по полуцелым $k \geqslant 1$ мы доказываем, что для любых степеней двойки $\varepsilon \leqslant \alpha \leqslant 1$ найдётся степень двойки $\delta \leqslant \varepsilon$ такая, что Заказчик выигрывает в игре с параметрами $\varepsilon, \delta, \alpha, \alpha, k$ и любым недоступным Менеджеру множеством на дереве бесконечного ветвления глубины $4(k-1)$.

База индукции ($k = 1$) очевидна. Индуктивный переход: пусть утверждение верно для k и надо доказать его для $k + 1/2$. Запустим стратегию из примера 3 на дере-

[1] На самом деле мы еще не до конца доказали теорему Гача. Мы построили стратегию с произвольным коэффициентом k, а нужна нам стратегия со сколь угодно большим отношением β/α: нам надо надкусить больше 1 и при этом запросить не больше $1/d$. Такую стратегию легко получить последовательным применением построенных стратегий с подходящими параметрами. Например, применяя их с параметрами $k = 2d$ и $\alpha = 1/(4d)$ по очереди для каждого из сыновей корня исходного дерева до тех пор, пока запрос корня не превысит впервые $1/d - 1/(4d)$, мы получим стратегию с параметрами $\alpha = 1/d$ и $\beta = 2d(1/d - 1/(4d)) = 3/2$, чего вполне достаточно. Для этой стратегии достаточно дерева глубины $4(2d-1)+1$.

ве глубины $4(k+1/2-1)$, только на этот раз будем во внуках запускать стратегию с параметром k, существующую по индуктивному предположению (глубина дерева с корнем во внуках как раз равна $4(k-1)$). В качестве параметра α при рекурсивных вызовах выберем степень двойки, лежащую в интервале $(\varepsilon/(6k), \varepsilon/(3k)]$. (Поскольку правая граница интервала вдвое больше левой, в этом интервале обязательно найдется степень двойки.) При каждом рекурсивном вызове запрос корня увеличивается не менее, чем на $\varepsilon/(6k^2)$, поэтому общее количество рекурсивных вызовов можно ограничить заранее — оно не превзойдет $6mk^2$. Тем самым, можно заранее подобрать параметры ε, δ для всех рекурсивных вызовов. Кроме того, можно ограничить количество внуков — каждому сыну достаточно $6k^2$ сыновей. В момент, когда все сыновья корня сделали резервы, мы достигли своей цели. В самом деле, обозначим через γ общий запрос корня при не-последних рекурсивных вызовах. Он даёт вклад $k\gamma$ в надкушенное, лежащий вне резервов, всего получается $k\gamma + m\varepsilon$ надкушенного при максимум $\gamma + m(\varepsilon/(3k))$ запрошенного. Поскольку $\gamma \leqslant m\varepsilon$, величина первого в $k+1/2$ больше величины второго:

$$k\gamma + m\varepsilon \geqslant (k+1/2)(\gamma + m\varepsilon/(3k)),$$

что проверяется простым раскрытием скобок.

Как видно из доказательства, стратегия будет выигрывать и на деревьях глубины $4(k-1)$ конечного ветвления, зависящего от ε, α, k. Однако полученная таким образом величина этого ветвления намного хуже указанной в теореме 88. Поясним, почему так получается. Дерево для стратегии с параметрами ε, α, k имеет ветвление α/ε в корне. В сыновьях корня ветвление мало, и им можно пренебречь. Чтобы оценить ветвление во внуках корня, нам нам нужно оценить сверху отношение α'/ε' при всех рекурсивных вызовах стратегии Эти вызовы делаются с параметрами $\alpha' \approx \varepsilon/(3k)$ и некоторых ε', которые меняются от вызова к вызову. При этом минимальное ε', с которым будет вызываться стратегия во внуках получается в результате цепи из $(\alpha/\varepsilon)6k^2$ применений преобразования $\varepsilon \mapsto \delta$. Это приводит к тому, что минимальное значение ε' значительно меньше исходного ε. Поэтому и ветвление во внуке, который будет обрабатываться первым, должно быть очень большим (на самом деле, поскольку мы заранее не знаем, кто будет обрабатываться первым, «старший сын» каждого сына корня должен иметь много сыновей).[1]

Таким образом, недостаток нашей стратегии в том, что она делает слишком много рекурсивных вызовов. Оказывается можно обойтись всего лишь $O(k^2)$ вы-

[1]Чтобы оценить ветвление во внуках, нужно индукцией по k оценить сверху отношение ε/δ первых двух параметров стратегии функцией $f_k(\alpha/\varepsilon)$ от k и α/ε. Значение $f_{k+1/2}(\alpha/\varepsilon)$ получается произведением $(\alpha/\varepsilon)6k^2$ значений $f_k(\alpha'/\varepsilon')$. При этом первое ε' можно взять равным α', но остальные должны быть значительно меньше — второе ε' уже в $f_k(1)$ раз меньше первого ε', третье в $f_k(f_k(1))$ меньше второго и так далее. Последний сомножитель в произведении получается $(\alpha/\varepsilon)6k^2$-кратным применением функции f_k к 1. Поэтому

$$f_{k+1/2}(\alpha/\varepsilon) \approx 3k \cdot f_k(1) \cdot f_k(f_k(1)) \cdot f_k(f_k(f_k(1))) \cdot \ldots$$

(в произведении $(\alpha/\varepsilon)6k^2$ сомножителей и мы используем приблизительное равенство, поскольку первое ε' лишь приблизительно равно $\varepsilon/(3k)$). При этом, скажем $f_{3/2}(\alpha/\varepsilon) \equiv 2$ (см. пример 2). Поэтому $f_2(\alpha/\varepsilon)$ не меньше экспоненты от α/ε: $f_2(\alpha/\varepsilon) > 2^{\alpha/\varepsilon}$, а $f_{2,5}(\alpha/\varepsilon)$ уже не меньше, чем башня из двоек высоты α/ε. (Можно было бы воспользоваться стратегией примера 3 и тем, что при индуктивном переходе можно увеличивать значение k почти на 1, но это мало бы помогло — башня двоек возникла бы при немного большем k.)

зовами (вместо $O(k^2\alpha/\varepsilon)$). Важно, что это число не зависит от α/ε. Для этого $k+1/2$-стратегия будет обрабатывать всех сыновей корня, не имеющих резерва, параллельно, а не последовательно, как раньше. Точнее, на каждом шаге для всех сыновей корня, которые не имеют резерва, стратегия будет выбирать по одному из их необработанных сыновей (внуков корня) и для полученного множества вершин делать рекурсивный вызов. Правда, для этого нужно изменить формат стратегии — теперь ее входом надо считать не одно дерево, а семейство одинаковых не связанных друг с другом деревьев. (Вершинам разных деревьев Менеджер обязан выделять непересекающиеся интервалы.) Само по себе это нововведение не поможет — может оказаться, что на каждом шагу только один сын корня не имеет резерва. В этом случае параллельной обработки фактически происходить не будет. С этим можно бороться следующим образом: не будем ждать, пока все сыновья корней сделают резервы. Вместо этого остановимся, когда количество сыновей с резервом составляет достаточно большую долю от общего их числа (во всех деревьях вместе — напомним, что теперь игра происходит в нескольких деревьях). Эту долю можно определить любым числом, большим половины (если мы хотим, чтобы k возросло не меньше, чем на $1/2$), скажем в три четверти. Это приведет к небольшой потере — значение параметра β будет равно трём четвертям от α. В свою очередь, это приведет к потерям при освоении средств — нижняя оценка на суммарный запрос всех корней будет три четвертых от старой оценки (что скажется на небольшом увеличении количества итераций, необходимых для достижения резервов). Освоение средств теперь будет выполняться так: если среднее арифметическое запросов корней оказалось меньше $(3/4)\alpha/k$, мы увеличиваемы запросы для некоторых, чей запрос меньше α, так, чтобы среднее арифметическое стало равно $(3/4)\alpha/k$.

Как и раньше, если запрос некоторого сына корня уже настолько приблизился к ε, что обработка нового его сына может повлечь превышение ε, мы просто увеличиваем запрос этого сына до ε. Как и раньше, это надо делать перед началом очередного рекурсивного вызова.

Таким образом, длина надкушенного будет равна $k\gamma+(3/4)m\varepsilon$ вместо $k\gamma+m\varepsilon$, как раньше (где γ равно надкушенному внуками, обработанными не последними, а m равно общему количеству сыновей корня во всех деревьях), а длина запрошенного останется той же, что раньше — $\gamma+m\varepsilon/(3k)$. Напомним, что $\varepsilon/(3k)$ здесь появилось, как параметр α при рекурсивных вызовах. Поэтому чтобы соотношение надкушенного к запрошенному было не меньше $k+1/2$, надо будет немножко уменьшить этот параметр — $\varepsilon/(6k)$ будет достаточно. Это приведет к увеличению количества итераций еще вдвое, что не страшно. Более подробно: на каждом шаге сумма запросов всех сыновей будет увеличиваться на величину, пропорциональную $m\varepsilon/k^2$. Сумма запросов всех сыновей не может превысить $m\varepsilon$, поэтому количество рекурсивных вызовов не превысит $O(k^2)$.

Перейдя к более детальному изложению, объясним более подробно, какие изменения надо ввести в определение игры и в построение выигрышной стратегии. Параметрами игры, помимо чисел $k,\varepsilon,\delta,\alpha,\beta$, дерева T и недоступного множества A, является число деревьев l. Смысл параметров ε,δ,A не меняется. Параметр α ограничивает сверху запрос каждого корня, а параметр β ограничивает снизу среднее количество свеженадкушенного в расчете на одно дерево — длина свеженадку-

шенного пространства должна быть не менее $l\beta$. Наконец, параметр k ограничивает снизу отношение длины свеженадкушенного пространства к сумме запросов корней всех деревьев.

Ясно, что и теперь можно соединять стратегии. Но теперь у соединяемых стратегий должны быть одинаковые параметры k, T, α. А именно, применим стратегию с параметрами $\varepsilon_1, \delta_1, \alpha, \beta_1, k$ и недоступным множеством A для некоторого подсемейства деревьев (каждое из которых изоморфно T). Затем применим её еще раз с новыми параметрами $\varepsilon_2, \delta_2, \alpha, \beta_2, k$ (и новым недоступным множеством, равным объединению A с владениями обработанного семейства деревьев) для другого, не пересекающегося со первым, семейства. И пусть $\varepsilon_1 = \delta_2$. Тогда в результате мы выиграем игру с параметрами $\varepsilon_2, \delta_1, \alpha, \beta_1 + \beta_2, k$ для объединенного семейства деревьев и недоступного множества A.

Индукцией по полуцелым k мы строим выигрышную стратегию Заказчика для любых $\varepsilon \leqslant \alpha \leqslant 1$, $\beta = (3/4)\alpha$ и некоторых $\delta \leqslant \varepsilon$ и конечного дерева глубины $4(k-1)$, ветвление которого мы оценим потом. Числа $\varepsilon, \alpha, \delta$ должны быть степенями двойки. Стратегия должна выигрывать для любых l и A.

База индукции ($k = 1$) очевидна. Индуктивный переход от k к $k + 1/2$: дерево, как и раньше имеет α/ε сыновей, а каждый из них — $O(k^2)$ сыновей (точное значение определим позже). Стратегия действует так. Сначала мы увеличиваем до ε запросы тех сыновей корней, запрос которых уже превысил $\varepsilon - \varepsilon/6k$. Дав Менеджеру пойти, мы затем проверяем, сколько сыновей корней не имеют резерва. Если их более одной четверти от общего числа, мы для каждого такого сына выбираем любого его необработанного сына. Для получившегося множества внуков мы делаем рекурсивный вызов с параметром k, выбирая в качестве параметра α степень двойки в интервале $(\varepsilon/(12k), \varepsilon/(6k)]$. Затем мы повторяем то же самое до тех пор, пока доля сыновей без резерва не станет меньше одной четверти.

При каждом рекурсивном вызове общий прирост запроса будет как минимум $(3/4)m(\varepsilon/(12k^2))$, где m — общее число сыновей всех корней. При этом сумма запросов всех корней не может стать больше $m\varepsilon$ — мы гарантируем, что запрос любого сына любого корня не превышает ε (если запрос некоторого сына достиг ε, у него появляется резерв). Поэтому общее количество рекурсивных вызовов будет не больше $O(k^2)$. Кроме того, запрос любого корня не превысит α, поскольку у каждого корня α/ε сыновей. Общий размер свеженадкушенного пространства не меньше общей длины всех резервов, которая не меньше $(3/4)m\varepsilon$, как и требовалось. Осталось оценить отношение длины свеженадкушенного к сумме запросов.

Обозначим через γ суммарный рост запросов всех корней, исключая те последние увеличения, которые приводили к образованию резервов. Он даёт вклад $k\gamma$ в надкушенное, лежащий вне резервов, всего получается $k\gamma + (3/4)m\varepsilon$ надкушенного при максимум $\gamma + m(\varepsilon/(6k))$ запрошенного. Поскольку $\gamma \leqslant m\varepsilon$, величина первого в $k + 1/2$ больше величины второго:

$$k\gamma + (3/4)m\varepsilon \geqslant (k + 1/2)(\gamma + m\varepsilon/(6k)),$$

что проверяется простым раскрытием скобок.

Осталось оценить необходимое ветвление дерева T, как функцию k и отношения α/ε (ясно, что для игры важно только отношение этих чисел). Как мы видели

(см. обсуждение после построения стратегии для дерева бесконечного ветвления), для этого сначала нужно понять во сколько раз δ меньше ε. Для этого индукцией по k докажем, что параметр δ можно выбрать в c_k меньшим ε, где c_k зависит только от k, и оценим сверху величину c_k.

Для $k=1$ у нас было $\delta=\varepsilon$, значит $c_1=1$. Стратегия для $k+1/2$ делает не более $O(k^2)$ вызовов k-стратегии. Отношение ε/δ есть произведение аналогичных отношений для этих вызовов и отношения исходного ε к параметру ε' последнего вызова (оно не превосходит $12k$, так как последний вызов делается с параметром α' не меньшим $\varepsilon/(12k)$ и таким же можно взять значение ε'). Таким образом, мы получаем рекуррентное соотношение

$$c_{k+1/2}=O(kc_k^{O(k^2)}),$$

откуда получается $c_k=2^{O(k)^{4(k-1)}}$.

Теперь легко оценить ветвление дерева T, как функцию от k и α/ε. Вспомним, что при построении (k+1/2)-стратегии нам необходимо было ветвление α/ε в корне и $O(k^2)$ в сыновьях корня. Поэтому ветвление на всех нечётных уровнях ограничено $O(k^2)$ (мы считаем, что корень находится на уровне 0). Осталось оценить ветвление на чётных уровнях больше 0. На втором уровне опять ветвление равно отношению параметров α' и ε' применяемых стратегий. Нетрудно заметить, что это отношение не зависит от исходных α и ε. В самом деле, во внуках k-стратегия вызывается с параметром α' не бо́льшим $\varepsilon/(6k)$ и с параметром ε', не меньшим $\varepsilon/c_{k+1/2}$. Поэтому ветвление дерева во внуках корня ограничено величиной $6kc_{k+1/2}$. Аналогичное верно и для всех остальных чётных уровней, только с меньшим k.

Таким образом, k-стратегия будет выигрывать на дереве ветвления

$$\max\{\alpha/\varepsilon, O(k^2), 6(k-1/2)c_k\}=\max\{\alpha/\varepsilon, O(k^2), 2^{O(k)^{4(k-1)}}\}.$$

При больших k второй член меньше третьего и им можно пренебречь.

Теперь мы можем завершить доказательство теоремы Гача – Дея. Пусть d есть степень двойки. Нам нужно построить стратегию с параметрами $\alpha=1/d$ и каким-то $\beta>1$, выигрывающую на одном дереве T небольшой глубины и ветвления. Для этого мы последовательно вызываем построенную стратегию с параметрами $k=2d$ и $\alpha=1/(4d)$ по очереди для каждого из сыновей корня T до тех пор, пока запрос корня впервые не превысит опасный порог $1/d-1/(4d)$. В результате мы запросим не более $1/d$, а надкусим не менее $2d(1/d-1/(4d))>1$. При каждом вызове запрос корня увеличивается не менее, чем на $(3/4)(1/(4d))(1/k)$, поэтому количество вызовов не превзойдёт $O(d^3)$. Поэтому нам достаточно $O(d^3)$ сыновей корня. При этом поддеревья в сыновьях должны годиться для $2d$-стратегии с параметрами $\alpha=1/(4d)$ и $\varepsilon=(1/(4d))/c_{2d}^{O(d^3)}$ Поэтому в сыновьях достаточно деревьев глубины $O(d)$ и ветвления

$$\max\{c_{2d}^{O(d^3)}, 2^{O(k)^{4(k-1)}}\}=2^{O(d)^{O(d)}}.$$

Теорема Гача – Дея доказана. ►

139 Докажите, что в теореме Гача – Дея глубина дерева не может быть меньше $d/4$: на дереве такой глубины (и бесконечного ветвления в каждой вершине) выигрышная стратегия есть у Менеджера, а не у Заказчика.

Верхняя оценка разницы между KM и KA пока что существенно больше нижней: как мы видели, $KM(x)$ (и даже $KP(x)$) превосходит $KA(x)$ не более чем на $O(\log n)$ для слов длины n. Ничего существенно лучшего не известно. Можно, впрочем, заменить n на $KA(x)$, как показывает следующая задача.

140 Докажите, что $KM(x) \leqslant KA(x) + O(\log KA(x))$. [Указание. Верно более сильное утверждение: $KM(x \mid KA(x)) \leqslant KA(x) + O(1)$. В самом деле, если $KA(x) = k$, то x рано или поздно попадёт в растущее дерево слов априорной сложности менее $k+1$. Это дерево в каждый момент имеет ширину (максимальное число попарно несравнимых элементов) не более 2^{k+1}, поэтому его можно покрыть не более чем 2^{k+1} растущими ветвями, следя за максимальными элементами дерева. Подробно см. теорему 127, с. 219.]

5.6. Теорема Левина – Шнорра

Определение априорной сложности гарантирует, что для любой перечислимой снизу полумеры p выполняется неравенство $KA(x) \leqslant -\log p(x) + O(1)$. Оказывается, что если p является мерой, то это неравенство выполнено не только для KA, но и для (вообще, говоря большей) величины KM.

Теорема 89. *Пусть μ — вычислимое распределение вероятностей на Ω и $p(x) = \mu(\Omega_x)$. Тогда существует такая константа c, что*

$$KM(x) \leqslant -\log p(x) + c$$

для любого слова x.

◄ Идею доказательства можно объяснить так: различие между KM и KA возникает из-за невозможности выделить сплошные участки отрезка $[0,1]$ по требованию. А это невозможно из-за того, что мы не знаем, какие из уже предъявленных требований возрастут, а какие нет, и не можем зарезервировать место. Но когда имеется не полумера, а вычислимая мера, то такой проблемы не возникает.

Если эти объяснения непонятны, их можно пропустить — сейчас мы изложим формальное доказательство.

Для каждого двоичного слова x рассмотрим отрезок π_x — часть отрезка $[0,1]$, определяемую по следующим правилам (см. рис. 5.1):

- длина отрезка π_x равна $p(x)$;
- $\pi_\Lambda = [0,1]$ (где Λ — пустое слово);
- для любого слова x отрезок π_x делится некоторой своей точкой на π_{x0} (левая часть) и π_{x1} (правая часть).

Мы будем сравнивать отрезки π_x с аналогичными отрезками для равномерной меры: через I_x мы обозначим отрезок, в котором двоичные записи чисел начинаются на x. Отрезки вида I_x будем называть *двоичными*.

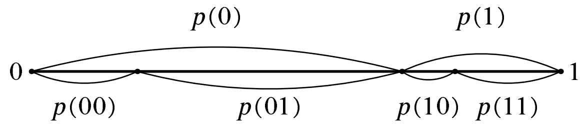


Рис. 5.1. Построение отрезков π_x.

Теперь рассмотрим множество G всех пар слов $\langle x, y\rangle$, для которых (двоичный) отрезок I_x лежит *строго внутри* отрезка π_y. Множество G перечислимо: поскольку функция p вычислима, мы можем находить концы отрезков π_y с любой точностью, и если они строго больше (или меньше) некоторого рационального числа, то рано или поздно это обнаружится. Кроме того, свойство $\langle x, y\rangle \in G$ сохранится, если заменить x на любое его продолжение (отрезок I_x станет меньше) или заменить y на любое начало (отрезок π_y станет больше). Если $\langle x, y_1\rangle \in G$ и $\langle x, y_2\rangle \in G$, то отрезки π_{y_1} и π_{y_2} имеют общую внутреннюю точку (они оба содержат I_x), и потому слова y_1 и y_2 сравнимы. Поэтому по теореме 81 (с. 147) существует вычислимое отображение Σ в себя, подграфик которого равен G. Используем его в качестве декомпрессора D в определении монотонной сложности. При этом $KM_D(y)$ есть минус двоичный логарифм длины самого большого двоичного отрезка, лежащего строго внутри π_y. Остаётся заметить, что (строго) внутри любого отрезка длины h есть двоичный отрезок длины не менее $h/4$, и воспользоваться оптимальностью. ▶

141 Проверьте, что сформулированное утверждение о двоичных отрезках действительно верно. [Указание. Пусть u — степень двойки, для которой $h/4 \leqslant u < h/2$. Тогда отрезок длины h должен пересечься по крайней мере с тремя последовательными отрезками длины u, и содержит внутри себя средний из них.]

Теорема 89 лежит в основе следующего способа оценки сложности (он применялся А. Н. Колмогоровым и его учениками для русских литературных текстов). Читая текст буква за буквой, экспериментатор пытается угадать следующую букву, формулируя свою догадку в виде распределения вероятностей на всех буквах алфавита. Затем открывается следующая буква и к оценке сложности добавляется величина $-\log p$, где p — приписанная этой букве вероятность. Полученная величина (если считать деятельность экспериментатора вычислимой) является верхней оценкой сложности: в самом деле, экспериментатор задаёт вычислимую меру на последовательностях букв (в форме условных вероятностей следующей буквы при известном начальном отрезке), и сложность текста не больше логарифма этой меры.

Конечно, при реальном проведении таких опытов непрактично требовать явного указания вероятности для каждой из букв алфавита; разумно ограничиться каким-то классом предсказаний типа «буква А будет с вероятностью 0,5, остальные гласные равновероятны и в сумме имеют вероятность 0,3, все остальные буквы равновероятны». Отметим ещё, что происходит при этом скорее оценка сложности данного текста при условии жизненного опыта экспериментатора; ничего удивительного, что получится почти что нуль, если экспериментатор знает текст наизусть (и вообще оценка может уменьшиться, если он хорошо знаком с текстами того же автора).

Этот же приём в программах сжатия называется «арифметическим кодированием» (arithmetic coding) и даже был запатентован (много позже экспериментов Колмогорова, в 1970-е годы).

Теперь всё готово для формулировки критерия случайности (в смысле Мартин-Лёфа) в терминах монотонной сложности; последовательность случайна, если доказанное только что неравенство обращается для её начальных отрезков в равенство. Пусть μ — вычислимое распределение вероятностей на пространстве Ω бесконечных двоичных последовательностей и $p(x) = \mu(\Omega_x)$.

Теорема 90 (Левина – Шнорра). *Последовательность $\omega \in \Omega$ случайна по Мартин-Лёфу относительно μ тогда и только тогда, когда*

$$-\log p(x) - KM(x) \leqslant c$$

для некоторого c и для всех начальных отрезков x последовательности ω.

◀ Доказательство состоит из двух частей. Для начала покажем, что если для данной последовательности ω разность $-\log p(x) - KM(x)$ принимает сколь угодно большие значения, то эта последовательность не случайна (множество $\{\omega\}$ является эффективно нулевым).

Пусть дано некоторое число c. Рассмотрим те слова x, у которых разность $-\log p(x) - KM(x)$ больше c. (Эту разность иногда называют *дефектом случайности*, но нужно быть осторожным — это слово употребляется во многих хотя и близких, но всё же различных смыслах; с одним из них мы уже встречались в предыдущей главе в применении к бесконечным последовательностям; другой описан в главе 14.) Множество таких слов обозначим D_c.

Множество D_c перечислимо (поскольку p вычислима, а KM перечислима сверху, разность перечислима снизу).

Лемма 1. *Множество всех последовательностей, у которых некоторое начало принадлежит D_c, имеет меру не больше 2^{-c} (относительно рассматриваемой нами меры μ).*

Неформально говоря, эта лемма верна потому, что на этом множестве мера μ в 2^c раз меньше априорной вероятности на Σ (а последняя в сумме не превосходит единицы). Более формально следует рассуждать так.

Рассматриваемое множество последовательностей есть объединение всех Ω_x при $x \in D_c$. В этом объединении можно оставить лишь минимальные $x \in D_c$, то есть те x, у которых никакие начала не принадлежат D_c. Пусть это будут $x_0, x_1, \ldots$ (мы не утверждаем, что множество минимальных элементов будет перечислимо, так что эта последовательность может и не быть вычислимой).

Для каждого x_i рассмотрим кратчайшее описание p_i (согласно определению монотонной сложности: $x_i \preccurlyeq D(p_i)$, где $D\colon \Sigma \to \Sigma$ — оптимальный способ описания). Тогда $l(p_i) = KM(x_i) < -\log p(x_i) - c$. Кроме того, ни одно из p_i не является началом другого (иначе соответствующие x_i были бы сравнимы). Поэтому $\sum_i 2^{-l(p_i)} \leqslant 1$ (как сумма равномерных мер непересекающихся множеств Ω_{p_i}). Соответствующие $p(x_i)$ в 2^c раз меньше, откуда и получаем требуемое. Лемма доказана.

По предположению рассматриваемая нами последовательность ω покрыта множеством из леммы при любом c. Поэтому, чтобы (в соответствии с определением эффективно нулевого множества) указать покрытие множества $\{\omega\}$ перечислимым семейством интервалов с суммарной μ-мерой не больше 2^{-c}, достаточно перечислить интервалы из D_c.

Тут, правда, возникает техническая трудность (которую мы уже обсуждали, говоря о тестах случайности). Для интервалов из D_c мы знаем, что *мера объединения* не больше 2^{-c} (согласно лемме), в то время как определение требует, чтобы *сумма мер* интервалов не превосходила 2^{-c}. Эту проблему нельзя решить, оставив в D_c минимальные точки, поскольку их множество может уже не быть перечислимым. Вместо этого можно использовать такое утверждение:

Лемма 2. *Всякое перечислимое множество слов $x_0, x_1, \ldots$ можно преобразовать в перечислимое множество несравнимых слов, сохранив неизменным объединение $\bigcup_i \Omega_{x_i}$. Это преобразование эффективно (алгоритм, перечисляющий первое множество, можно вычислимо преобразовать в алгоритм перечисления второго).*

В самом деле, если при перечислении появляется продолжение ранее рассмотренного слова, то его можно просто выбросить (соответствующий интервал в объединении и так покрыт). Если же появляется слово y, являющееся (собственным) началом ранее рассмотренного слова x, то нужно разбить добавляемое множество $\Omega_y \setminus \Omega_x$ в объединение интервалов и заменить y на слова, задающие эти интервалы. Лемма 2 доказана.

После применения леммы 2 мы получаем перечислимое множество несравнимых слов; заметим, что эти слова уже не обязательно принадлежат D_c, но это и не важно. Достаточно, что они задают непересекающиеся интервалы и что объединение этих интервалов (согласно лемме 1) имеет μ-меру не более 2^{-c}.

Для завершения доказательства теоремы Левина – Шнорра осталось показать, что если последовательность принадлежит эффективно нулевому множеству, то разности (между минус логарифмом меры и монотонной сложностью) для её начальных отрезков не ограничены. Идею этого рассуждения можно описать так: для данного малого по мере множества мы строим монотонный способ описания, благоприятствующий последовательностям из этого множества (дающий малую сложность некоторым их начальным отрезкам).

Более подробно. Пусть последовательность ω принадлежит эффективно нулевому (по мере μ) множеству U. Для каждого c можно эффективно указать покрытие множества U интервалами $\Omega_{x_0}, \Omega_{x_1}, \ldots$, имеющими суммарную меру меньше 2^{-c}. Если увеличить меры всех этих интервалов в 2^c раз, то их сумма останется меньше единицы. Применив к последовательности $p_i = 2^c \mu(\Omega_{x_i})$ теорему 59 (с. 112), мы получаем беспрефиксный способ описания, для которого сложность i не превосходит $-\log \mu(\Omega_{x_i}) - c + 2$. Композиция этого способа описания с вычислимым отображением $i \mapsto x_i$ даёт беспрефиксный способ описания D_c, у которого

$$KP'_{D_c}(x_i) \leqslant -\log \mu(\Omega_{x_i}) - c + 2.$$

(Индекс у D_c подчёркивает, что всё построение зависит от c.) Монотонная сложность не больше префиксной, поэтому если разница между минус логарифмом меры

и префиксной сложностью велика (как показывает только что написанное неравенство при больших c), то тем более будет велика разность с монотонной сложностью. Надо только аккуратно соединить построенные способы описания в один.

По аналогии с построением оптимального способа описания будем считать, что $\hat{c}u$ является описанием слова v, если u является описанием слова v относительно D_c. Здесь $\hat{c}$ — самоограниченная запись натурального числа c длины $O(\log c)$. Для построенного способа описания D уже можно написать неравенство, верное при всех c:

$$KP'_D(x_i) \leqslant -\log\mu(\Omega_{x_i}) - c + O(\log c).$$

Поскольку монотонная сложность не превосходит префиксной, можно заменить $KP'_D(x_i)$ на $KM(x_i)$ и заключить, что у всех слов x_i (для данного c) разность между минус логарифмом меры и монотонной сложностью не меньше $c - O(\log c)$. Для каждой последовательности из U можно найти начало такого вида при любых c, поэтому разности между минус логарифмом меры и монотонной сложностью для начальных отрезков этой последовательности не ограничены.

Теорема Левина – Шнорра доказана. ►

142 Покажите, что в первой части доказательства (если разность не ограничена, то последовательность содержится в эффективно нулевом множестве) достаточно перечислимости P сверху, а во второй части — снизу.

Приведённое доказательство теоремы Шнорра по существу доказывает немного больше. Вот что ещё из него можно заключить:

Теорема 91. *В формулировке теоремы можно заменить монотонную сложность $KM(x)$ на априорную сложность $KA(x)$.*

◄ Априорная сложность меньше, поэтому разность при переходе к априорной сложности только увеличится; остаётся доказать первую часть теоремы для априорной сложности. Для этого достаточно заметить при доказательстве леммы 1, что $\sum_i 2^{-KA(x_i)} \leqslant 1$, поскольку эта сумма является суммой априорных вероятностей непересекающихся интервалов Ω_{x_i}. ►

Теорема 92. *В формулировке теоремы можно заменить монотонную сложность $KM(x)$ на префиксную сложность $KP(x)$.*

◄ Здесь, напротив, мы увеличиваем сложность, а не уменьшаем, так что сомнения вызываем лишь вторая часть теоремы. Но в доказательстве мы как раз и получали оценку для префиксной сложности. ►

Именно такой критерий случайности («теорема Левина – Шнорра в форме Чейтина») сейчас наиболее распространён (см., например, [85]; об истории вопроса можно прочесть в обзоре [11]).

С философской точки зрения авторам кажется более естественным использовать монотонную (или априорную) сложность, хотя и префиксная сложность имеет свой резон (см. ниже выражение для дефекта случайности через префиксную сложность). Заметим, например, что при использовании префиксной сложности разность, о которой идёт речь в критерии случайности, может быть и отрицательной: скажем, для равномерной меры величина $-\log\mu(\Omega_x)$ есть длина слова x, а

префиксная сложность может превышать длину (на величину порядка логарифма длины), см. теорему 63, см. 117.

Кроме того, использование монотонной или априорной сложности позволяет усилить теорему следующим образом:

Теорема 93. *Если последовательность ω не случайна (в смысле Мартин-Лёфа) по мере μ, то величина $-\log p(x) - KM(x)$ не только не ограничена для её начальных отрезков, но и стремится к бесконечности.*

◀ Вспомним доказательство: в нём для последовательности слов x_i строился беспрефиксный способ описания, при котором слово x_i имело описание p_i, причём длина p_i (и тем самым префиксная сложность x_i при этом способе описания) не превосходила $-\log\mu(\Omega_{x_i}) - c$. Чтобы получить нужную оценку для монотонной сложности, мы можем использовать (при каждом i) продолжения слова p_i в качестве описаний продолжений слова x_i, при этом длина первых соответствует логарифму меры вторых, как это делалось при доказательстве теоремы 89 (с. 165).

Более формально можно воспользоваться неравенством $KM(xy) \leqslant KP(x) + KM(y\,|\,x)$ (задача 135) и релятивизованным вариантом теоремы 89, утверждающим, что $KM(y\,|\,x) \leqslant -\log\mu_x(\Omega_y)$ для любого вычислимого семейства мер, (вычислимо) зависящего от параметра x. При этом в качестве μ_x нужно взять меру, сосредоточенную на продолжениях слова x, для которой $\mu_x(\Omega_y) = \mu(\Omega_{xy})/\mu(\Omega_x)$.

Для случая равномерной меры (когда $-\log\mu(\Omega_x) = l(x)$) всё упрощается и можно просто сказать, что слово $p_i z$ является описанием слова $x_i z$ для любого слова z. ▶

Для префиксной сложности это рассуждение уже не проходит, но для вычислимой последовательности длин утверждение остаётся верным:

$\boxed{\textbf{143}}$ Пусть A — разрешимое бесконечное множество длин (натуральных чисел), а ω — последовательность. Если $KP(x) \geqslant -\log\mu(\Omega_x) - c$ для некоторого c и для любого начала x последовательности ω, длина которого принадлежит A, то ω случайна. [Указание. В доказательстве теорем 90 и 92 можно измельчать интервалы до нужного размера.]

Мы привели доводы в пользу использования монотонной сложности для характеризации случайности по Мартин-Лёфу. С другой стороны, характеризация случайности в терминах сложности начальных отрезков имеет некоторый дефект: в то время как случайность последовательности нулей и единиц инвариантна относительно вычислимых перестановок индексов (и соответствующего изменения меры), понятие начального отрезка и тем самым критерий случайности в терминах начальных отрезков таковым не является. Как заметил Андрей Румянцев, используя префиксную сложность, можно модифицировать критерий случайности, сделав его инвариантным.

Пусть F — конечное множество натуральных чисел, а ω — последовательность нулей и единиц. Через $\omega(F)$ будем обозначать сужение ω на F, то есть двоичное слово, образованное битами ω_i при $i \in F$ (в порядке возрастания индексов).

Пусть фиксирована вычислимая мера μ на Ω. Для данного множества F и слова Z, длина которого равна числу элементов в F, можно рассмотреть событие $\omega(F) = Z$. Его вероятность будем обозначать $\mu_{F,Z}$.

144 Докажите, что если последовательность ω случайна (в смысле Мартин-Лёфа) относительно меры μ, то найдётся такое c, что

$$KP(F, \omega(F)) \geqslant -\log \mu_{F,\omega(F)} - c$$

для любого конечного множества F.

[Указание. Мера множества последовательностей, где это условие нарушается для фиксированного c, не превосходит 2^{-c}, умноженного на сумму априорных вероятностей всех пар F, Z, то есть не больше 2^{-c}.]

(Заметим, что если F — начальный отрезок, то F восстанавливается по $\omega(F)$, так что мы приходим к прежней формулировке.)

Сформулированное условие является не только необходимым, но и достаточным. Более того, достаточно потребовать его для произвольной вычислимой возрастающей последовательности множеств, покрывающей все индексы.

145 Пусть $F_0 \subset F_1 \subset F_2 \subset \ldots$ — вычислимая последовательность конечных множеств, в объединении дающих всё $\mathbb{N}$. Если для некоторой последовательности ω и для некоторого c неравенство

$$KP(F_i, \omega(F_i)) \geqslant -\log \mu_{F_i,\omega(F_i)} - c$$

выполнено при всех i, то последовательность ω случайна в смысле Мартин-Лёфа по мере μ.

[Указание: удобно переставить индексы и считать, что F_i — начальные отрезки. Далее нужно воспользоваться задачей 143: повторить доказательство теоремы Левина – Шнорра, но использовать только интервалы нужных длин (измельчая неформатные).]

Из этого утверждения следует, например, что двумерная последовательность битов (отображение $\mathbb{Z}^2 \to \{0,1\}$) случайна в смысле Мартин-Лёфа по равномерной мере (все биты независимы, нули и единицы равновероятны) тогда и только тогда, когда любой квадрат $N \times N$ с центром в начале координат (при всех нечётных N) имеет префиксную сложность не меньше $N^2 - O(1)$.

Отметим также ещё одну причину, делающую естественным появление в теореме Левина – Шнорра префиксной сложности. Оказывается, что можно указать количественное уточнение этой теоремы в терминах ограниченного в среднем дефекта случайности (см. раздел 3.5):

146 Пусть μ — вычислимая мера на пространстве Ω. Покажите, что функция

$$t(\omega) = \sum_{x \preccurlyeq \omega} \frac{m(x)}{\mu(x)}$$

(сумма по всем конечным началам x последовательности ω; через $m(x)$ обозначается *дискретная* априорная вероятность слова x, а $\mu(x)$ обозначает меру интервала Ω_x) является универсальным ограниченным в среднем тестом случайности.

[Указание. Будем представлять себе перечислимую снизу функцию как сумму характеристических функций интервалов с коэффициентами. Добавление нового члена в такую сумму можно представлять себе как увеличение «веса» двоичного слова, если значением функции на последовательности считать сумму весов вдоль неё. Веса слов в таком случае должны быть перечислимой снизу функцией u на словах, а условие на интеграл записывается как $\sum_x \mu(x)u(x) \leqslant 1$. Максимальная такая функция отличается от $m(x)$ множителем $1/\mu(x)$. Ещё надо уточнить, что при $\mu(x) = 0$ для некоторого x дробь $m(x)/\mu(x)$ считается бесконечной.]

147 Покажите, что в предыдущей задаче можно заменить сумму на максимум (точную верхнюю грань) и тем самым получить количественный вариант критерия случайности с префиксной сложностью. (Например, для случая равномерной меры ограниченный в среднем дефект случайности равен $\sup_n[n - KP(\omega_0 \ldots \omega_{n-1})]$.)

[Указание. Перечислимую снизу функцию, которая принимает рациональное значение a в некотором эффективно открытом множестве и равна нулю снаружи, можно представить с помощью весов, равных a и размещённых в несравнимых вершинах. Любую перечислимую функцию t можно представить с точностью до постоянного множителя как сумму $t(\omega) = \sum t_k(\omega)$, где $t_k(\omega) = 2^k$, если $t(\omega) > 2^k$, и $t_k(\omega) = 0$ в противном случае. Представив t_k как описано выше, мы добьёмся, чтобы в сумме предыдущей задачи все слагаемые были разными степенями двойки. Тогда сумма совпадает с наибольшим слагаемым с точностью до $O(1)$-множителя.]

Утверждение последней задачи было доказано в давней статье Гача [44]. По ходу доказательства в ней получается и утверждение задачи 146, которое недавно было переоткрыто Миллером и Ю и получило название “ample excess lemma” [105].

148 (**а**) Покажите, что если последовательность ω случайна по вычислимой мере μ, то разница $-\log\mu(x) - KP(x)$ не только ограничена сверху для начальных отрезков последовательности ω, но и стремится к минус бесконечности с ростом длины начального отрезка. Другими словами, если $KP(x) \leqslant -\log\mu(x) + c$ для некоторого c и для бесконечно многих начальных отрезков некоторой последовательности ω, то эта последовательность не случайна.

(**б**) Покажите, что если $KP(x) \leqslant -\log\mu(x) + \log l(x) + c$ для некоторого c и для *всех* начальных отрезков последовательности ω, то эта последовательность не случайна (всюду имеется в виду случайность по Мартин-Лёфу).

[Указание: в обоих случаях утверждение вытекает из леммы об “ample excess” (задача 146).]

Случай равномерной меры заслуживает того, чтобы выписать формулировки всех доказанных нами теорем и задачи 148 специально для этого случая (через $(\omega)_n$ мы обозначаем начальный отрезок последовательности ω, имеющий длину n):

Теорема 94. (а) *Оценка сверху*: *для любого слова x выполнены неравенства*

$$KA(x) \leqslant KM(x) + O(1) \leqslant l(x) + O(1).$$

(б) *Критерий случайности*: *последовательность ω случайна* (*в смысле Мартин-Лёфа*) *по равномерной мере тогда и только тогда, когда для её начальных*

отрезков эти неравенства обращаются в равенства:

$$KA((\omega)_n) = KM((\omega)_n) + O(1) = n + O(1).$$

(в) *Для неслучайной (в смысле Мартин-Лёфа) по равномерной мере последовательности ω разность $n - KM((\omega)_n)$, и тем более $n - KA((\omega)_n)$, стремится к бесконечности.*

(г) *Последовательность ω случайна (в смысле Мартин-Лёфа) по равномерной мере тогда и только тогда, когда $KP((\omega)_n) \geqslant n - c$ для некоторого c и для всех n. В этом случае $KP((\omega)_n) - n \to +\infty$ при $n \to \infty$. (Таким образом, нижний предел разности $KP((\omega)_n) - n$ равен $-\infty$ для неслучайной последовательности и $+\infty$ для случайной; промежуточные варианты невозможны.)*

(д) *Последовательность ω случайна (в смысле Мартин-Лёфа) по равномерной мере тогда и только тогда, когда $KP(F, \omega(F)) \geqslant |F| - c$ для некоторого c и для всех конечных множеств F.*

Другой вариант утверждения (г): последовательность ω случайна тогда и только тогда, когда сумма $\sum_n 2^{n-KP((\omega)_n)}$ конечна (задача 146).

Имеется ещё один критерий случайности по Мартин-Лёфу для случая равномерной меры, который интересен тем, что в нём можно ограничиться использованием обычной колмогоровской сложности (и обойтись без монотонной или префиксной). Даже удивительно, что этот критерий был найден только недавно (см. [105]), поскольку этим занимались ещё в конце 1960-х годов и подошли к нему очень близко (см. [196, 99]), а его доказательство достаточно просто и естественно и не выходит из круга идей, хорошо известных в то время.

Теорема 95. *Пусть $f\colon \mathbb{N} \to \mathbb{N}$ — вычислимая функция и ряд $\sum 2^{-f(n)}$ сходится. Пусть ω — случайная (в смысле Мартин-Лёфа) по равномерной мере последовательность. Тогда*

$$KS((\omega)_n \,|\, n) \geqslant n - f(n) - O(1)$$

(найдётся c, при котором для всех n выполнено неравенство $KS((\omega)_n \,|\, n) \geqslant n - f(n) - c$).

◄ Пусть утверждение теоремы не выполнено. Это значит, что для любого c найдётся такое n, что

$$KS((\omega)_n \,|\, n) < n - f(n) - c.$$

Другими словами, каково бы ни было c, последовательность ω покрыта некоторым интервалом Ω_x, для которого

$$KS(x \,|\, n) < n - f(n) - c,$$

где n — длина x. Для каждого n таких интервалов не более $2^{n-f(n)-c}$, поэтому их суммарная мера не более $2^{-f(n)}2^{-c}$ при данном n и не более

$$2^{-c}\left(\sum_n 2^{-f(n)}\right)$$

для интервалов всех длин. Поэтому последовательность ω образует эффективно нулевое множество: выбирая нужное c, получаем покрытие сколь угодно малой меры. Значит, она не случайна. (Заметим, что сумма ряда $\sum 2^{-f(n)}$ может быть невычислимой; это не важно, так как мы можем использовать любую верхнюю оценку для этой суммы.) ►

Замечание. В рассуждении мы использовали только перечислимость сверху, поэтому эта теорема верна при $f(n) = KP(n)$: для любой случайной по равномерной мере последовательности ω выполнено неравенство $KS((\omega)_n\,|\,n) \geqslant n - KP(n) - O(1)$. Как мы увидим в теореме 98, это условие является необходимым и достаточным.

Из теоремы 95 следует, например, что для случайной последовательности (по равномерной мере) обычная сложность начальных отрезков не меньше $n - 2\log n - O(1)$ или даже $n - \log n - 2\log\log n - O(1)$, поскольку соответствующие ряды сходятся.

Чем меньше функция f, тем ближе ряд к расходящемуся и тем более сильным является утверждение теоремы. Оказывается, что при некоторых f оно становится критерием случайности.

Теорема 96. *Существует всюду определённая вычислимая функция f с натуральными аргументами и значениями, для которой $\sum_n 2^{-f(n)} < \infty$, обладающая таким свойством: если для некоторой последовательности ω и для некоторого c при всех n выполнено неравенство*

$$KS((\omega)_n\,|\,n) \geqslant n - f(n) - c,$$

то последовательность ω случайна (в смысле Мартин-Лёфа) по равномерной мере.

◄ Нам нужно доказать, что любая неслучайная последовательность (то есть любая последовательность, содержащаяся в наибольшем эффективно нулевом множестве) имеет «простые» начальные отрезки. При этом критерий «простоты» (точнее, функцию f) нам предстоит выбирать самим.

Как это делается? Пусть, скажем, у нас есть некоторое семейство интервалов с суммой длин не более ε. Пусть F — множество задающих их двоичных слов (семейство состоит из интервалов Ω_u при $u \in F$). Рассортируем слова множества F (и тем самым интервалы семейства) по длинам и подсчитаем для каждого n суммарную меру интервалов, соответствующих словам длины n. Выберем функцию f таким образом, чтобы эта мера была примерно равна $2^{-f(n)}$. (Поскольку по условию $f(n)$ должно быть целым, то это равенство нельзя сделать точным, но можно добиться ошибки не более чем в 2 раза. Для простоты мы не будем этого учитывать.) Тогда по построению сумма ряда $\sum_n 2^{-f(n)}$ не превосходит ε. С другой стороны, в F имеется примерно $2^{n-f(n)}$ слов длины n и каждое из них может быть описано (при известном n и других параметрах конструкции) своим номером, требующим $n - f(n)$ битов, что даёт верхнюю оценку сложности для слов из F. При этом всякая покрытая нашими интервалами последовательность имеет начальный отрезок в множестве F.

Как это соображение можно применить в нашей ситуации? Рассмотрим наибольшее эффективно нулевое множество. Для любого $\varepsilon > 0$ имеется его покрытие интервалами с суммой длин не больше ε, и описанным способом из этого покрытия получается функция f с $\sum_n 2^{-f(n)} \leqslant \varepsilon$. Далее нам нужно как-то соединить такие функции для разных ε в одну. Будем делать это следующим образом.

Для каждого $c = 0, 1, 2, \ldots$ рассмотрим покрытие с суммой мер меньше 2^{-3c} и построим по нему функцию f описанным способом, а затем уменьшим её на $2c$. Таким образом для каждого c мы получим функцию f_c, для которой

$$\sum_n 2^{-f_c(n)} < 2^{-c}$$

(вместо 2^{-3c} мы получаем 2^{-c}, поскольку уменьшили функцию на $2c$), и множество слов F_c, содержащее $2^{n-f_c(n)-2c}$ слов длины n, причём любая неслучайная последовательность имеет начальный отрезок в F_c.

Далее мы определим $f(n)$ равенством

$$2^{-f(n)} = \sum_c 2^{-f_c(n)},$$

тогда

$$\sum_n 2^{-f(n)} = \sum_n \sum_c 2^{-f_c(n)} = \sum_c \sum_n 2^{-f_c(n)} \leqslant \sum_c 2^{-c} \leqslant 1,$$

и при этом $f(n) \leqslant f_c(n)$ при любых n и c. С другой стороны, множество F_c перечислимо равномерно по c (это гарантируется определением эффективно нулевого множества), и потому любое слово x длины n в F_c может быть задано (при известных n и c) своим порядковым номером в перечислении, который содержит $n - f_c(n) - 2c$ битов:

$$KS(x \mid n, c) \leqslant n - f_c(n) - 2c + O(1),$$

откуда

$$KS(x \mid n) \leqslant n - f_c(n) - 2c + O(\log c) < n - f(n) - c$$

для любого слова $x \in F_c$ длины n (при достаточно бóльших c).

Пусть теперь ω — неслучайная последовательность. Из сказанного следует, что она (для каждого c) имеет начало в F_c; если n — длина этого начала (и c достаточно велико), то

$$KS((\omega)_n \mid n) < n - f(n) - c,$$

что противоречит предположению теоремы.

Это рассуждение, однако, ещё не доказывает теорему — в ней требуется, чтобы функция f была вычислима, в то время как F_c мы можем лишь перечислять (и ни в какой момент не можем быть уверены, что больше слов данной длины не будет). Вспомним, однако, что нас интересуют покрытия интервалами, и в этих покрытиях можно заменить один большой интервал Ω_z на много маленьких Ω_{zt},

если в качестве t взять все слова некоторой фиксированной длины. За счёт этого можно добиться вычислимости функции f_c — надо договориться, что минимально разрешённая длина выдаваемого при перечислении слова растёт со временем. Тогда количество выдаваемых слов длины n является вычислимой функцией от n, так как они могут появляться только до некоторого заранее известного момента. Это рассуждение можно провести параллельно для всех c, и тогда сумма $\sum_c 2^{-f_c(n)}$ также будет вычислима по n.

Наконец, имеется ещё совсем техническая трудность: мы строим функцию f с натуральными значениями, поэтому надо её ещё округлить. ►

Две предыдущие теоремы вместе дают критерий случайности, в котором используется обычная (а не монотонная или префиксная) сложность. Этот критерий имеет некоторый «запас». А именно, можно заменить условную сложность $KS((\omega)_n \,|\, n)$ на безусловную сложность $KS((\omega)_n)$ или на условную префиксную сложность $KP((\omega)_n \,|\, n)$.

В самом деле, при такой замене сложность только увеличится, поэтому адаптации требует лишь теорема 96. Вариант с префиксной сложностью: надо воспользоваться тем, что для любого конечного множества A и любого элемента $x \in A$ справедливо неравенство $KP(x \,|\, A) \leqslant \log_2 |A| + O(1)$ (можно рассмотреть беспрефиксное кодирование словами длины $\log_2 |A|$). Немного сложнее удалить условие n (для обычной, не префиксной сложности). В этом случае слово $x \in F_c$, имеющее длину n, надо задавать не его порядковым номером в множестве $F_{c,n}$ слов длины n из множества F_c, а номером во всём F_c (расположенном в порядке возрастания длин). Соответственно число битов, необходимое для этого, есть

$$\log(|F_{c,0}| + |F_{c,1}| + \ldots + |F_{c,n}|),$$

и всё будет в порядке, если последний член $|F_{c,n}|$ больше всех предыдущих вместе взятых (тогда порядковый номер увеличится не более чем в два раза от добавления предыдущих членов). Ясно, что этого можно достичь тем же приёмом (заменяя короткое слово на все его продолжения некоторой длины). Заметим, что это делается при данном c, то есть условие c остаётся, но это не страшно (в итоге оно даёт вклад $O(\log c)$).

Отсюда получаем такой результат:

Теорема 97. *Случайность последовательности ω (в смысле Мартин-Лёфа) по равномерной мере равносильна тому, что для любой вычислимой функции f с конечной суммой ряда $\sum_n 2^{-f(n)}$ выполнено свойство*

$$KS((\omega)_n) \geqslant n - f(n) - O(1).$$

Этот критерий использует простую колмогоровскую сложность и именно его обычно называют «теоремой Миллера – Ю».

Недостатком этого критерия можно считать то, что в нём есть квантор по f. И хотя этот квантор можно переставить (найдётся f, которая позволяет отвергнуть все неслучайные последовательности; именно такова формулировка теоремы 96),

всё равно хотелось бы как-нибудь обойтись без f. Это можно сделать, но тогда снова приходится упоминать префиксную сложность:

Теорема 98. *Последовательность ω является случайной (в смысле Мартин-Лёфа) относительно равномерной меры тогда и только тогда, когда*

$$KS((\omega)_n) \geqslant n - KP(n) - O(1).$$

◀ Если ряд $\sum_n 2^{-f(n)}$ сходится (для вычислимой функции f), то $KP(n) \leqslant \leqslant f(n) + O(1)$. Поэтому приведённое в теореме условие на ω сильнее, чем в теореме 97.

Таким образом, остаётся доказать утверждение в другую сторону, что мы уже обсуждали в замечании после теоремы 95. ▶

Заметим, что и в этой теореме можно заменить $KS((\omega)_n)$ на $KS((\omega)_n \,|\, n)$ (годится то же самое рассуждение).

$\boxed{\textbf{149}}$ Убедитесь в этом.

Этот результат был доказан в работе Гача [44] (с. 391).

$\boxed{\textbf{150}}$ Покажите, что в теореме 96 нельзя положить $f(n) = 2\log n$. [Указание. Из теоремы 95 следует, что для случайной последовательности ω выполняется более сильное неравенство $KS((\omega)_n) \geqslant n - \log n - 2\log\log n - O(1)$. Поэтому, если разбавить её очень редкими нулями на вычислимых местах, то получится последовательность, для которой $KS((\omega)_n) \geqslant n - 2\log n - O(1)$, но не случайная. Аналогичное рассуждение показывает, что не годится никакая функция с вычислимо сходящимся рядом $\sum 2^{-f(n)}$.]

Всё сказанное до сих пор, однако, не отвечает на естественный вопрос: а нельзя ли вообще обойтись без функции f и потребовать, чтобы $KS((\omega)_n)$ было больше $n - O(1)$ (как для монотонной сложности)?

Такое требование было бы самым естественным, но, как сразу же заметил Мартин-Лёф, ничего хорошего не получается: если в последовательности встречаются сколь угодно длинные участки из нулей, то есть начальный отрезок длины n, который кончается на k нулей (при большом k). Его сложность не больше $n - k + 2\log k + O(1)$ (для беспрефиксного кодирования числа k достаточно $2\log k$ битов), и разница с длиной будет $k - 2\log k - O(1)$. Если же в последовательности нет идущих подряд k нулей (при некотором k), то сложность её начальных отрезков длины n ограничена $\alpha n + O(1)$ при некотором $\alpha < 1$.

Можно оценить неизбежную разницу между длиной и сложностью более точно (см. [196, 99]):

Теорема 99. *Найдётся такая константа c, что для любой последовательности $\omega \in \Omega$ при бесконечно многих n выполнено неравенство*

$$KS((\omega)_n) \leqslant n - \log n + c.$$

◀ Для каждого n выделим среди слов длины n некоторую часть, составляющую $1/n$ от всех слов длины n. При этом сделаем это так, чтобы у любой бесконечной

последовательности было бесконечно много выделенных начал и чтобы множество выделенных слов было разрешимо.

Почему это возможно? Ряд $\sum 1/n$ расходится, поэтому можно его разделить на бесконечное число групп, в каждой из которых сумма больше единицы. С помощью слов каждой группы покроем все последовательности в один слой (каждая последовательность имеет начало из этой группы) — это можно сделать в порядке возрастания длин, покрывая ещё не покрытые слова. (Строго говоря, есть проблема с округлением, поскольку $2^n/n$ — не целое число, но и после округления ряд будет расходиться.)

Любое выделенное слово длины n задаётся (при известном n) своим порядковым номером, который имеет $n - \log n$ битов. Поэтому его условная сложность (при известном n) не больше $n - \log n + O(1)$. Более того, если нумеровать все выделенные слова подряд, то порядковый номер не сильно увеличивается (число выделенных слов растёт почти как геометрическая прогрессия с показателем 2, и добавление всех предыдущих слов увеличивает их число не более чем в $O(1)$ раз).

Отсюда и следует утверждение теоремы. ▶

151 Предложите другое доказательство, использующее такое простое соображение: начало длины k любой последовательности можно рассматривать как двоичную запись некоторого числа N (можно приписать вначале единицу, чтобы не потерять нули), и следующие за этим началом N битов последовательности позволяют восстановить все $k + N$ битов.

152 Покажите, что утверждение теоремы верно не только для некоторого c, но и для всех c (в том числе отрицательных). [Указание. Если ряд $\sum 2^{-f(n)}$ расходится, то можно немного увеличить функцию f, сохранив это свойство: найдётся функция g, для которой $g(n) - f(n) \to \infty$ и ряд $\sum 2^{-g(n)}$ расходится.]

153 Покажите, что в теореме 99 (в варианте с условной сложностью) можно заменить логарифм на любую вычислимую функцию f, для которой ряд $\sum 2^{-f(n)}$ расходится.

В статье Мартин-Лёфа [99] без доказательства (со ссылкой на его неопубликованную работу) приводится утверждение с безусловной сложностью и произвольной вычислимой функцией f с расходящимся рядом. Этот результат (также без доказательства и со ссылкой на Мартин-Лёфа) приведён и в обзоре [196]. (Как его доказать, мы не знаем.)

Отметим также, что утверждение теоремы 95 приведено в статье [99] в немного другом варианте:

154 Докажите, что если последовательность ω случайна (в смысле Мартин-Лёфа по равномерной мере), а функция f вычислима и *ряд* $\sum 2^{-f(n)}$ *вычислимо сходится*, то $KS((\omega)_n \mid n) \geqslant n - f(n)$ для всех n, кроме конечного числа. [Указание. Если ряд вычислимо сходится, а неравенство нарушается бесконечно много раз, то хвосты ряда можно использовать для получения покрытий сколь угодно малой меры.]

А что будет, если требовать большой сложности начального отрезка не для всех (достаточно длинных) отрезков, а для бесконечного их числа? В той же статье Мартин-Лёфа [99] приводятся следующие результаты:

$\boxed{\textbf{155}}$ Докажите, что для почти всех (по равномерной мере) последовательностей ω найдётся c, при котором $KS((\omega)_n \,|\, n) \geqslant n - c$ для бесконечно многих n.

[Указание: если это не так, то для всякого c найдётся N, начиная с которого начальные отрезки длины n имеют сложность меньше $n - c$. Для данных c и N множество таких последовательностей имеет меру меньше 2^{-c}; с ростом N множество растёт и объединение по N имеет меру не больше 2^{-c} по непрерывности.]

$\boxed{\textbf{156}}$ Пусть $KS((\omega)_n \,|\, n) \geqslant n - c$ для некоторой константы c и бесконечно многих n. Тогда последовательность ω случайна в смысле Мартин-Лёфа (по равномерной мере). [Указание. Если ω покрыта одним из интервалов с общей мерой меньше 2^{-c}, то достаточно длинные её начальные отрезки (при известной длине) можно описать порядковым номером среди слов данной длины, попадающих в один из этих интервалов, для чего достаточно $2\log c + n - c$ битов.]

$\boxed{\textbf{157}}$ Покажите, что утверждение предыдущей задачи можно усилить, заменив условную сложность на безусловную $KS((\omega)_n)$. [Указание: можно воспользоваться задачей 6 или, что ещё проще, задачей 55.]

Таким образом, получается некоторое множество полной меры, содержащееся в множестве всех случайных по Мартин-Лёфу последовательностей. (Если бы дополнение этого множества было не просто нулевым, но и эффективно нулевым, то мы получили бы критерий случайности по Мартин-Лёфу — но это не так.)

Сравнительно недавно выяснилось, что это за множество — оказывается, это множество всех последовательностей, случайных в смысле Мартин-Лёфа с оракулом $\mathbf{0}'$. Такие последовательности иногда называют «2-случайными» (а случайные по Мартин-Лёфу в обычном смысле, без оракула, называют «1-случайными»). См. [103, 128]; простое доказательство приведено в [9]. Аналогичный критерий 2-случайности можно сформулировать и в терминах префиксной сложности: надо потребовать, чтобы $KP((\omega)_n)$ было не меньше $n + KP(n) - c$ для некоторого c и для бесконечно многих n ([197], простое доказательство в [198]).

5.7. Случайное число Ω

Интересное применение критерия случайности составляет следующая теорема. Пусть m — максимальная перечислимая снизу полумера на множестве натуральных чисел (например, $m(x) = 2^{-KP(x)}$, или распределение на выходе универсального вероятностного алгоритма, см. главу 4). Г. Чейтин предложил рассмотреть число

$$\Omega = \sum_n m(n)$$

(которое можно назвать общей вероятностью остановки универсального вероятностного алгоритма) и сделал такое интересное наблюдение:

Теорема 100. *Двоичная запись числа Ω является случайной по Мартин-Лёфу последовательностью относительно равномерной меры.*

Заметим, что значение Ω зависит от выбора конкретной максимальной перечислимой снизу полумеры, но утверждение теоремы остаётся верным при любом таком выборе.

◀ Пусть нам известны первые n двоичных знаков числа Ω. Они образуют число Ω_n, которое есть приближение снизу к Ω с погрешностью не более 2^{-n}. Будем перечислять снизу приближения к $m(0), m(1), \ldots$ параллельно, постепенно добавляя всё новые $m(i)$, пока найденные числа в сумме не дадут больше $\Omega_n - 2^{-n}$. (Такое рано или поздно случится, так как сумма ряда равна Ω и строго больше выбранной нами границы.) Составим список всех чисел i, которые встречаются в этой сумме.

Заметим, что в этом списке встречаются все числа i, у которых $m(i) \geqslant 2 \cdot 2^{-n}$ (поскольку если бы такое число было пропущено, то погрешность была бы больше 2^{-n}), и в том числе все i с $KP(i) < n - c$ (для некоторого c, зависящего от выбора функции m, но не от n). Поэтому минимальное число, не входящее в этот список, имеет сложность не меньше $n - c$. Значит, и сам этот список, и двоичное разложение числа Ω_n, по которому этот список построен, имеют сложность не меньше $n - c'$ для некоторого другого c' и для всех n. (Заметим, что это двоичное разложение определяет n, так что значение n не надо указывать отдельно.) Остаётся воспользоваться критерием случайности с префиксной сложностью (см. теоремы 92 и 94). ▶

Как мы уже говорили, можно сразу определять случайность действительных чисел в смысле Мартин-Лёфа (относительно обычной меры на прямой), если в определении эффективно нулевого множества требовать наличия алгоритма, который по каждому рациональному $\varepsilon > 0$ выдаёт покрытие интервалами с рациональными концами и суммарной мерой не более ε.

158 Покажите, что определённая таким образом случайность равносильна случайности последовательности знаков в двоичном разложении.

159 Покажите, что квадрат (синус, экспонента) случайного действительного числа случаен. [Указание. Прообраз множества меры нуль имеет меру нуль, и это можно эффективизировать.]

160 Всегда ли сумма двух случайных действительных чисел случайна? [Указание: они могут быть «зависимы».]

На самом деле число Ω (точнее, числа типа Ω, поскольку для разных максимальных полумер они разные) — не просто любопытный курьёз. Оказывается, что класс таких чисел имеет несколько эквивалентных описаний [17, 70]. Мы приведём их, следуя [84].

5.7.1. Сводимость и полнота по Соловею

Напомним, что действительное число α *перечислимо снизу*, если оно является пределом возрастающей вычислимой последовательности рациональных чисел. (Эквивалентное определение: если множество рациональных чисел, меньших α,

перечислимо.) Желая как-то сравнивать скорость сходимости таких последовательностей, введём следующее определение.

Пусть $a_i \to \alpha$ и $b_i \to \beta$ — две строго возрастающие вычислимые последовательности, сходящиеся соответственно к α и β. Мы говорим, что a_i «лучше сходится» к α, чем $b_i \to \beta$, если существует всюду определённая вычислимая функция h, для которой

$$\alpha - a_{h(i)} \leqslant \beta - b_i$$

для любого i. (Точнее следовало бы говорить «не хуже» вместо «лучше».)

Другими словами, мы требуем, чтобы для любого члена второй последовательности можно было указать член в первой последовательности, который имеет ту же (или меньшую) погрешность приближения. Очевидно, это отношение рефлексивно и транзитивно (возьмём композицию сводящих функций). Заметим также, что на самом деле это свойство самих перечислимых снизу чисел α и β, и не зависит от выбора последовательностей. В самом деле, если две вычислимые возрастающие последовательности сходятся к одному и тому же пределу, то для любого члена одной из них можно алгоритмически найти больший член второй, просто подождав его появления.

Это позволяет дать такое определение: мы говорим, что $\alpha \preccurlyeq_1 \beta$ (для двух перечислимых снизу чисел α и β), если возрастающая вычислимая последовательность $a_i \to \alpha$ сходится лучше, чем аналогичная последовательность $b_i \to \beta$.

Как мы уже говорили, выбор последовательностей тут не играет роли; можно не упоминать их и в определении, сказав, что $\alpha \preccurlyeq_1 \beta$, если существует частичная вычислимая функция φ, определённая на всех рациональных числах $r < \beta$, для которой

$$\varphi(r) < \alpha \quad \text{и} \quad \alpha - \varphi(r) \leqslant \beta - r$$

для всех $r < \beta$.

$\boxed{\textbf{161}}$ Докажите, что перечислимое снизу число α вычислимо тогда и только тогда, когда $\alpha \preccurlyeq_1 \beta$ для любого перечислимого снизу числа β.

Вот ещё одна полезная эквивалентная формулировка:

Теорема 101. *Соотношение $\alpha \preccurlyeq_1 \beta$ выполнено тогда и только тогда, когда разность $\beta - \alpha$ перечислима снизу, то есть $\beta = \alpha + \rho$ для некоторого перечислимого снизу ρ.*

◀ Заметим прежде всего, что если α и ρ перечислимы снизу, то $\alpha \preccurlyeq_1 \alpha + \rho$. В самом деле, имея $s < \alpha + \rho$, мы можем подождать, пока сумма растущих приближений к α и ρ не превзойдёт s. В этот момент точность текущего приближения к α будет не хуже, чем к $\alpha + \rho$.

Остаётся доказать обратное: *если $\alpha \preccurlyeq_1 \beta$, то $\rho = \beta - \alpha$ перечислимо снизу.* В самом деле, пусть вычислимая возрастающая последовательность b_n сходится к β. Найдём для каждого её члена приближение a_n к α с такой же или меньшей погрешностью. Тогда $b_n - a_n \leqslant \beta - \alpha$ и $b_n - a_n \to \beta - \alpha$. Правда, последовательность $b_n - a_n$ может не быть возрастающей, но это легко поправить, заменив её n-й член на максимум из первых n членов. ▶

Частным случаем этой теоремы является такое утверждение: пусть даны два вычислимых ряда $\sum a_i$ и $\sum b_i$ из неотрицательных рациональных чисел (кроме, быть может, a_0 и b_0, которые могут быть и отрицательными). Если $a_i \leqslant b_i$ при всех $i > 0$, то для их сумм α и β выполнено $\alpha \preccurlyeq_1 \beta$.

Верно и обратное утверждение: если $\alpha \preccurlyeq_1 \beta$, то можно найти ряды $\sum a_i = \alpha$ и $\sum b_i = \beta$ с такими свойствами ($0 \leqslant a_i \leqslant b_i$ при $i \geqslant 1$). В самом деле, $\beta = \alpha + \rho$ для перечислимого снизу ρ; представим $\alpha = \sum a_i$ и $\rho = \sum r_i$ и возьмём $b_i = a_i + r_i$.

$\boxed{\mathbf{162}}$ Покажите, что верно и более сильное утверждение и что не только ряд a_i (как видно из приведённого рассуждения), но и ряд b_i можно выбирать произвольно. А именно, если $\alpha \preccurlyeq_1 \beta = \sum b_i$, где $b_i \geqslant 0$, то найдётся представление $\alpha = \sum a_i$, где $0 \leqslant a_i \leqslant b_i$ при всех $i > 0$. (Все ряды вычислимы.) [Указание. Будем строить последовательность a_i индуктивно, сохраняя такой инвариант: текущее приближение (снизу) к α имеет не бо́льшую погрешность, чем текущее приближение к β. Получив следующее приближение к β, мы применяем соотношение $\alpha \preccurlyeq_1 \beta$, чтобы найти не худшее приближение к α. Если возникающее при этом a_i окажется не больше b_i, то и хорошо; если больше, то его можно уменьшить до b_i без нарушения инварианта.]

Из сказанного следует, что для любого перечислимого снизу α мы имеем

$$\alpha \preccurlyeq_1 2\alpha \preccurlyeq_1 3\alpha \preccurlyeq_1 \ldots,$$

поскольку все разности (равные α) перечислимы снизу. Отсюда же видно, что обратные неравенства неверны, если α не вычислимо (если оба числа α и $-\alpha$ перечислимы снизу, то α вычислимо). Можно сказать, что наше отношение $\preccurlyeq_1$ «слишком чувствительно»: оно различает α и 2α, хотя их двоичные разложения отличаются всего лишь сдвигом. Как его огрубить? Наиболее естественный способ, видимо, допустить умножение на константу.

Будем говорить, что перечислимое снизу число α *сводится по Соловею* к перечислимому снизу числу β, и писать $\alpha \preccurlyeq \beta$, если $\alpha \preccurlyeq_1 c\beta$ для некоторого целого $c > 0$. (Можно ввести обозначение: $\alpha \preccurlyeq_c \beta$, если $\alpha \preccurlyeq_1 c\beta$. Тогда $\alpha \preccurlyeq \beta$, если $\alpha \preccurlyeq_c \beta$ для некоторого c.) Это отношение, очевидно, тоже рефлексивно и транзитивно.

Теорема 102. *Среди перечислимых снизу чисел существует наибольшее в смысле сводимости по Соловею.*

◄ Действуем как с перечислимыми полумерами: складываем с некоторыми коэффициентами w_i (годятся $w_i = 2^{-i}$) все перечислимые снизу числа α_i в интервале $[0, 1]$. Тогда сумму $\alpha = \sum w_i\alpha_i$ можно представить как $w_i\alpha_i$ плюс перечислимое снизу число (сумма остальных членов ряда), откуда и следует требуемое. ►

Наибольшие в смысле этого порядка числа называются *полными по Соловею.*

Можно даже определить некоторую количественную характеристику для полных по Соловею чисел, которую можно было бы назвать «дефектом полноты»: для числа β дефект полноты определяется как наименьшее c, при котором $\alpha \preccurlyeq_1 c\beta$. Здесь α — некоторое фиксированное полное по Соловею число; функция дефекта

зависит от выбора α, но определена с точностью до ограниченного множителя. Дефект β конечен тогда и только тогда, когда β является полным.

Оказывается, что полные по Соловею числа и только они могут появляться как «число Ω» в описанной выше конструкции [161, 17].

Теорема 103. *Сумма любой максимальной перечислимой снизу полумеры на множестве натуральных чисел является полным по Соловею перечислимым снизу числом. Наоборот, любое полное по Соловею положительное перечислимое снизу число в* $(0,1)$ *является суммой некоторой максимальной перечислимой снизу полумеры.*

◀ Докажем, что произвольное перечислимое снизу число α сводится к сумме произвольной максимальной полумеры. Число α есть сумма вычислимого ряда $\sum a_i$ с неотрицательными (кроме первого) рациональными членами. (Говоря о вычислимости ряда, мы подразумеваем вычислимость функции $i \mapsto a_i$, а не суммы этого ряда.) С точностью до константы (которая не играет роли в определении полноты) этот ряд мажорируется нашей перечислимой снизу полумерой. Разница между полумерой и рядом перечислима снизу, и потому α сводится по Соловею к сумме полумеры.

С другой стороны, пусть число $\alpha \in (0,1)$ полно по Соловею. Как представить его в виде суммы максимальной перечислимой снизу полумеры? Возьмём какую-нибудь максимальную полумеру $m_0, m_1, \ldots$. В силу полноты $\sum m_i \preccurlyeq_1 c\alpha$ при некотором c, и потому $\alpha = \sum m_i/c + \tau$ для перечислимого снизу τ. Поделив все m_i на c и добавив τ (скажем) к m_0, получим максимальную полумеру с суммой в точности α. ▶

Числа типа Ω имеют и ещё одно описание (помимо полноты по Соловею): это в точности перечислимые снизу *случайные* числа в интервале $(0,1)$. Доказательство содержит несколько шагов, и сейчас мы к нему перейдём.

5.7.2. Полные по Соловею числа случайны

Случайность полных по Соловею чисел следует из уже доказанных результатов: мы знаем, что такое число является суммой максимальной полумеры, и что такая сумма случайна. (Строго говоря, это рассуждение применимо только к числам от 0 до 1, но любое число может быть приведено в этот интервал сдвигом и масштабированием.)

Мы приведём сейчас (следуя подстрочному примечанию в статье Левина [84]) другое доказательство того же факта, которое не использует теорему Левина–Шнорра и вообще не апеллирует к понятию сложности.

Для начала напомним, что существование перечислимого снизу случайного числа может быть доказано без упоминания Ω (см. задачу 86). Поэтому достаточно доказать, что свойство случайности «наследственно вверх»: если α случайно и $\alpha \preccurlyeq \beta$, то и β случайно.

Без ограничения общности можно считать, что $\alpha \preccurlyeq_1 \beta$ (умножение на рациональную константу не меняет случайности). Мы должны доказать, что из неслучайности β следует неслучайность α. Для этого мы преобразуем покрытие β

перечислимым семейством интервалов в покрытие α другим семейством той же или меньшей меры. Пусть b_i — возрастающая вычислимая последовательность рациональных чисел с пределом β. Будем следить параллельно за ростом b_i и за появлением новых интервалов в покрытии для β. Если в покрытии появляется интервал, который целиком левее текущего b_i, то он к делу не относится (β в нём заведомо не лежит). Если в покрытии появляется интервал, который целиком правее текущего b_i, то можно отложить его рассмотрение до момента, когда b_i в него зайдёт (если этого никогда не случится, то интервал не покрывает β). Наконец, если интервал покрытия содержит b_i, мы вместо него рассматриваем интервал той же длины, который начинается с приближения a_j к α, которое имеет не большую погрешность, чем b_i (как приближение к β). Легко видеть, что если один из интервалов покрытия содержит β, то он рано или поздно будет преобразован в интервал той же длины, покрывающий α. Поэтому из неслучайности β следует неслучайность α, что и требовалось доказать.

Итак, случайность наследуется вверх и потому полные по Соловею числа случайны.

Замечание. Вторую часть рассуждения можно переформулировать так: пусть α и β перечислимы снизу; если одно из них случайно, то сумма $\alpha + \beta$ случайна. Оказывается, что верно и обратное: если перечислимые снизу α и β оба не случайны, то и сумма $\alpha + \beta$ не случайна. (Мы дадим ниже несколько доказательств этого факта.)

5.7.3. Критерий случайности в терминах предсказаний

Прежде чем доказывать обратное утверждение (что случайные перечислимые снизу числа полны по Соловею), сделаем небольшое отступление и дадим критерий случайности для перечислимых снизу чисел в терминах некоторой игры.

Наблюдатель глядит на возрастающую последовательность рациональных чисел (которые сообщаются ему по очереди). Время от времени он делает предсказания такого типа: «последовательность в будущем возрастёт по сравнению с текущим значением не более чем на δ». Здесь δ — неотрицальное рациональное число (выбираемое наблюдателем). Будем считать, что наблюдатель выигрывает, если

(1) одно из его предсказаний остаётся верным навсегда;

(2) общая сумма всех использованных им чисел δ мала (меньше рационального $\varepsilon > 0$, которое сообщается наблюдателю до начала игры).

Правила игры не требуют, чтобы в каждый момент было действующее текущее предсказание, но этого несложно добиться, каждый раз делая предсказания с нулевым (или с малым и быстро убывающим) δ. Отметим также, что каждое предсказание можно отложить, и это не ухудшит дела. Поэтому, в частности, мы можем считать, что следующее предсказание делается лишь после того, как предыдущее оказалось опровергнутым.

В терминах этой игры можно дать критерий случайности:

Теорема 104. *Пусть a_i — вычислимая возрастающая последовательность рациональных чисел, сходящаяся к (перечислимому снизу) числу α. Наблюдатель*

имеет вычислимую стратегию, обеспечивающую ему выигрыш в описанной игре, если и только если α не случайно.

◀ Вычислимая стратегия в такой игре сразу же даёт последовательность отрезков, имеющих сумму длин не больше ε и покрывающих α; отрезки легко преобразовать в интервалы чуть большей длины.

С другой стороны, если α не случайно, то существует последовательность интервалов с малой суммой мер, его покрывающая. Эти интервалы можно использовать для предсказания. А именно, когда мы обнаруживаем, что текущее приближение покрыто одним из интервалов, мы предсказываем, что оно никогда не выйдет из этого интервала. Когда и если оно выйдет, снова ожидаем попадания в один из интервалов и т. д. При этом мы будем брать самый первый (раньше других появившийся) из интервалов, покрывающих текущее приближение. Это гарантирует, что рано или поздно мы выберем интервал, покрывающий α, и соответствующее ему предсказание останется верным навсегда. ▶

Утверждение теоремы может быть переформулировано без использования игровой терминологии.

Теорема 105. *Пусть a_i — вычислимая возрастающая последовательность рациональных чисел, сходящаяся к α. Число α не случайно тогда и только тогда, когда для каждого рационального $\varepsilon > 0$ можно эффективно указать вычислимую последовательность неотрицательных рациональных чисел $h_0, h_1, \ldots$, для которой $\sum_i h_i < \varepsilon$ и $\alpha \leqslant a_i + h_i$ для некоторого i.*

◀ Это соответствует игре, в которой предсказания делаются на каждом шаге. Как мы говорили, это не важно, так как можно использовать нулевые h_i.) ▶

Мы рассматривали также критерий случайности по Мартин-Лёфу, предложенный Соловеем (теорема 31, с. 76); в нём шла речь не о покрытии для каждого ε, а об одном покрытии конечной меры, покрывающем α бесконечно много раз. Аналогичную модификацию допускает и предыдущая теорема:

Теорема 106. *Пусть a_i — вычислимая возрастающая последовательность рациональных чисел, сходящаяся к α. Число α не случайно тогда и только тогда, когда существует вычислимая последовательность неотрицательных рациональных чисел h_i с конечной суммой $\sum_i h_i$, для которой $\alpha \leqslant a_i + h_i$ для бесконечно многих i.*

◀ Если α не случайно, то мы можем применить предыдущую теорему к $\varepsilon = 1, 1/2, 1/4, 1/8, \ldots$, и затем сложить все полученные последовательности, предварительно сдвинув i-ю последовательность вправо на i (от этого её свойства не нарушатся). Каждая из последовательностей даст по крайней мере одно i, для которого $\alpha \leqslant a_i + h_i$, и благодаря сдвигу таких i будет бесконечно много.

С другой стороны, если $\alpha \leqslant a_i + h_i$ для бесконечно многих i, то мы получаем последовательность интервалов с конечной суммой длин, которая покрывает α бесконечно много раз (точнее говоря, отрезков, но их можно увеличить до интервалов). Остаётся воспользоваться критерием Соловея — или вспомнить его доказательство: множество точек, покрытых этими интервалами в m или более слоёв, эффективно открыто и имеет меру $O(1/m)$. ▶

Из доказанных только что критериев случайности сразу же вытекает такое следствие (на первый взгляд удивительное). Рассмотрим вычислимую последовательность положительных рациональных чисел a_i, для которой ряд $\sum a_i$ сходится. *Случайность суммы не изменится, если каждый член умножить на ограниченный отделённый от нуля множитель.* (В самом деле, все h_i можно умножить на константу.)

Докажем теперь обещанное утверждение:

Теорема 107. *Если α и β перечислимы снизу и не случайны, то и их сумма $\alpha + \beta$ не случайна.*

◀ Естественная идея: попытаться воспользоваться критерием с предсказаниями: если в игре с α мы предсказываем рост h, а в игре с β мы предсказываем рост k, то в игре с $\alpha + \beta$ можно предсказать рост $h + k$. Но тут возникает трудность: если возобновлять такие предсказания после их нарушения, то одно и то же значение h может комбинироваться с разными k и войдёт в сумму многократно.

Выход состоит в том, чтобы использовать предсказания одного и того же размера для α и β. Чтобы сделать предсказание для $a_i + b_i$ (после того, как предыдущее окажется негодным), мы ожидаем, пока текущие приближения a_i и b_i попадут внутрь соответствующих покрытий (по предположению α и β не случайны и покрыты перечислимыми семействами интервалов малых мер). Возьмём максимальные h и k, при которых $(a_i, a_i + h)$ и $(b_i, b_i + k)$ целиком содержатся в (обнаружившейся части) соответствующих покрытий. Наше предсказание будет интервалом $(a_i + b_i, a_i + b_i + 2\delta)$, где $\delta = \min(h, k)$ (другими словами, мы утверждаем, что $a_i + b_i$ в будущем возрастёт не более чем на 2δ).

Остаётся проверить две вещи. Во-первых, убедимся, что одно из предсказаний останется верным навсегда. В самом деле, α и β покрыты какими-то интервалами, вместе со своими σ-окрестностями для какого-то $\sigma > 0$. Текущие приближения a_i и b_i рано или поздно попадут в эти интервалы, если после этого нам придётся делать предсказание, то h и k будут выбраны не меньше σ, интервал предсказания покроет $\alpha + \beta$, и это предсказание уже не нарушится никогда.

Во-вторых, мы должны убедиться, что общая сумма длин предсказаний мала. В самом деле, когда интервал предсказания $(a_i + b_i, a_i + b_i + 2\delta)$ становится непригодным, это гарантирует, что либо a_i, либо b_i с тех пор увеличилось как минимум на δ. Значит, общая мера покрытия справа от текущих приближений уменьшилась по крайней мере на δ, что гарантирует оценку на сумму всех предсказаний (удвоенная сумма мер покрытий). В этом рассуждении мы используем то, что оба интервала $(a_i, a_i + \delta)$ и $(b_i, b_i + \delta)$ целиком входят в покрытие, поэтому важно, что мы брали минимум из h и k. ▶

Вернёмся к критерию случайности из теоремы 105. Предлагаемое там условие может быть ослаблено в двух отношениях. Во-первых, вместо вычислимой последовательности h_i можно рассматривать перечислимую снизу последовательность. Во-вторых, можно заменить h_i на весь остаток (хвост ряда) $h_i + h_{i+1} + \ldots$:

Теорема 108. *Пусть дана вычислимая возрастающая последовательность рациональных чисел a_i с пределом α. Предположим, что для каждого рационального $\varepsilon > 0$ можно эффективно указать перечислимую снизу последова-*

тельность неотрицательных действительных чисел h_i, для которой $\sum h_i < \varepsilon$ и $\alpha \leqslant a_i + h_i + h_{i+1} + \ldots$ для некоторого i. Тогда число α не случайно.

◀ Представим себе, что для каждого i есть маляр, которому выдано ведёрко с h_i краски и велено красить действительную прямую, идя направо от a_i, пропуская уже покрашенные участки и крася всё остальное. (Поскольку h_i перечислимы снизу, надо представлять себе, что краска появляется в ведёрке постепенно, и используется по мере появления и по мере надобности.)

Покрашенная зона представляет собой объединение перечислимого семейства отрезков общей длины не более $\sum h_i$ (общее количество краски). Если $\alpha < a_i + h_i + h_{i+1} + \ldots$, то точка α будет покрашена, так как вся краска в правой части неравенства накладывается в один слой справа от a_i и не поместится левее α. (В условии теоремы было нестрогое неравенство, но это не важно, поскольку все h_i можно, скажем, удвоить. По тем же причинам не важно, что мы покрыли α отрезками, а не интервалами.) ▶

Отсюда можно вывести ещё одну формулировку критерия случайности:

Теорема 109. *Пусть $\alpha = \sum r_i$ — сумма вычислимого ряда из неотрицательных рациональных чисел (и тем самым перечислимо снизу). Число α не случайно тогда и только тогда, когда для каждого $\varepsilon > 0$ можно эффективно указать перечислимое множество $W \subset \mathbb{N}$, для которого (1) $\sum_{i \in W} r_i < \varepsilon$ и (2) W имеет конечное дополнение (содержит все натуральные числа, начиная с некоторого).*

◀ Если α не случайно, то его можно покрыть интервалами малой меры и включить в W те i, для которых отрезок $[r_0 + \ldots + r_{i-1}, r_0 + \ldots + r_i]$ покрывается целиком. В обратную сторону утверждение прямо следует из теоремы 108, надо положить $a_i = r_0 + \ldots + r_{i-1}$ и $h_i = r_i$ при $i \in W$ (и $h_i = 0$ при $i \notin W$). ▶

Из этого критерия сразу видно, что сумма двух неслучайных перечислимых снизу чисел неслучайна (надо взять пересечение множеств W_1 и W_2 для каждого из них), так что мы получаем новое доказательство теоремы 107.

Использованный при доказательстве теоремы 108 приём оказывается полезным и в других ситуациях, например, в следующей задаче (это рассуждение сообщил авторам Л. Бьенвеню, оригинальное рассуждение в [70] существенно сложнее).

163 Пусть U — эффективно открытое подмножество отрезка, имеющее меру меньше 1 и содержащее все неслучайные последовательности (скажем, одно из множеств универсального теста Мартин-Лёфа). Докажите, что мера U является перечислимым снизу случайным числом.

[Указание. Пусть мера множества U равна α. Имея покрытие числа α интервалами малой меры, мы можем построить покрытие наименьшего числа вне U интервалами той же меры. Как только текущее приближение к α попадает в некоторый интервал, мы предполагаем, что оно не выйдет из этого интервала и красим имеющееся на данный момент дополнение к U слева направо. Если наше предположение окажется истинным (а когда-то оно окажется), то мы действительно покрасим наименьшее число вне U (покрашенная часть есть объединение отрезков, а не интервалов, но это не важно).]

5.7.4. Случайные перечислимые снизу числа полны

Теперь уже несложно доказать обратную импликацию [70]: любое перечислимое снизу случайное число является полным по Соловею. Рассмотрим два перечислимых снизу числа α и β и сходящиеся к ним возрастающие вычислимые последовательности $a_i \to \alpha$ и $b_i \to \beta$. Рассмотрим приращения $h_i = a_{i+1} - a_i$ первой последовательности. Попробуем применить их в качестве предсказаний для второй. Другими словами, мы сдвигаем интервал (a_0, a_1) так, чтобы его левый конец пришёлся на b_1. Затем мы ждём, пока b_i выйдет из полученного интервала. Пусть b_{i_1} — первый такой член. Сдвинем (a_1, a_2) так, чтобы его левый конец пришёлся на b_{i_1}, и подождём появления b_{i_2} справа от этого интервала. И так далее.

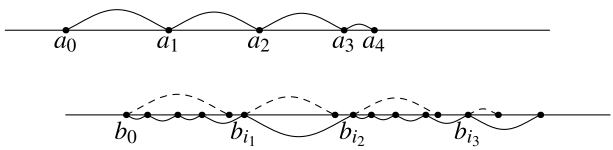


Рис. 5.2. Приращения для a_i используются как предсказания для β.

Есть две возможности:

(1) предсказатель выигрывает, то есть какой-то из сдвинутых интервалов содержит все оставшиеся b_i, очередное i_k не определено;

(2) процесс продолжается бесконечно.

Во втором случае $\alpha \preccurlyeq_1 \beta$, поскольку разность $\beta - \alpha$ оказалось представленной в виде суммы вычислимого ряда (промежутки между сдвинутыми интервалами). Отсюда мы заключаем, что при полном α и неполном β этот случай невозможен.

Следовательно, при полном α и неполном β имеет место первый случай, и мы получаем последовательность отрезков с конечной суммой длин, покрывающих β. Повторив это рассуждение для $\alpha/2$ (которое тоже полное), мы получим другую такую последовательность, для $\alpha/4$ третью и так далее. Суммы длин для этих последовательностей не превосходят $\alpha, \alpha/2, \alpha/4, \ldots$, поэтому можно эффективно указать покрытие числа β со сколь угодно малой суммой. (Само α не вычислимо, но это нам и не нужно, достаточно верхней оценки.)

Эти рассуждения можно провести для любого неполного β, взяв в качестве α произвольное полное перечислимое снизу число. Тем самым мы доказали, что неполное число β не является случайным, что завершает доказательство обещанного результата:

Теорема 110. *Перечислимое снизу действительное число случайно тогда и только тогда, когда оно полно по Соловею.*

5.7.5. Медленная сходимость и функции Соловея

Мы уже встречались с несколькими результатами такого типа: предел возрастающей вычислимой последовательности рациональных чисел случаен тогда и только

тогда, когда она сходится медленно. Сейчас мы рассмотрим ещё несколько результатов в этом направлении [7, 53].

Пусть дан вычислимый ряд $\sum r_i$ с неотрицательными членами и конечной суммой. Тогда r_i не превышает $O(m(i))$, где $m(i)$ — априорная вероятность натурального числа i (в смысле главы 4), а префиксная сложность $KP(i)$, равная $-\log m(i)$, не превышает $-\log r_i$ (с точностью до аддитивной константы). Будем говорить, что ряд *медленно сходится в смысле Соловея*, если эта верхняя оценка для r_i является $O(1)$-точной при бесконечно многих i, то есть если $r_i \geqslant \varepsilon m(i)$ при некотором $\varepsilon > 0$ и бесконечно многих i. Переформулировка: ряд быстро сходится (не является медленно сходящимся в смысле Соловея), если $r_i/m(i) \to 0$.

Исторически *функцией Соловея* (Solovay function) называют вычислимую верхнюю оценку $S(i)$ для префиксной сложности $KP(i)$, если эта оценка точна в бесконечном числе мест, то есть если $KP(i) \leqslant S(i) + c$ для некоторого c и всех i, а также $KP(i) \geqslant S(i) - c$ для некоторого c и бесконечно многих i. Таким образом, ряд $\sum r_i$ медленно сходится в смысле Соловея тогда и только тогда, когда $-\log r_i$ является функцией Соловея. (Обычно, говоря о функциях Соловея, имеют в виду функции с целыми значениями, так что в данном случае надо ещё произвести округление.) Мы приведём несколько результатов, связывающих случайность с медленной сходимостью, по большей части следуя [7, 53].

Теорема 111. *Сумма вычислимого сходящегося ряда с неотрицательными рациональными членами случайна тогда и только тогда, когда этот ряд медленно сходится в смысле Соловея.*

Другими словами, сумма ряда $\alpha = \sum r_i$ не случайна тогда и только тогда, когда $r_i/m(i) \to 0$.

◄ Пусть $r_i/m(i) \to 0$. Тогда для произвольного $\varepsilon > 0$ мы можем положить $h_i = \varepsilon m(i)$ и получить перечислимую снизу последовательность, которую можно использовать в теореме 108. Следовательно, α не случайно.

Приведём альтернативное рассуждение, показывающее, что α не является полным по Соловею. Вспомним доказательство теоремы 103: если $r_i \leqslant m(i)$, то $\sum r_i \preccurlyeq_1 \sum m(i)$. А если $r_i \leqslant cm(i)$, то $\sum r_i \preccurlyeq_c \sum m(i)$. Это остаётся верным, если неравенство $r_i \leqslant cm(i)$ верно не для всех i, а только начиная с некоторого. Следовательно, если $r_i/m(i) \to 0$, то для $\alpha = \sum r_i$ выполнено свойство $\alpha \preccurlyeq_c \sum m(i)$ при любом $c > 0$. Может ли тогда α быть полным? Нет, и вот почему. Если оно полно, то $\sum m(i) \preccurlyeq_d \alpha$ при некотором d. Но тогда $\alpha \preccurlyeq_{cd} \alpha$ при всех $c > 0$, в том числе и сколь угодно малых. Выходит, что $\alpha \preccurlyeq_{1/2} \alpha$, а отсюда следует вычислимость α: по любому приближению снизу к нему можно указать вдвое более точное!

Осталось доказать обратное утверждение: предположив, что число $\alpha = \sum r_i$ не случайно, установить, что $r_i/m(i) \to 0$. Представим себе отрезок $[0, \alpha]$, разбитый на участки длиной $r_0, r_1, \ldots$ (слева направо). Имея покрытие числа α интервалами с малой суммой мер, мы рассмотрим те из участков, которые покрыты целиком (вместе с концами). Это перечислимое множество, а их суммарная мера не превосходит меры покрытия. Если мера покрытия не больше 2^{-2n}, то мы можем умножить соответствующие r_i на 2^n и их сумма не превзойдёт 2^{-n}. Заметим также, что все участки, начиная с некоторого, покрыты. Таким способом мы получаем полумеру

M^n, для которой $M^n(i)/r_i \geqslant 2^n$ для всех достаточно больших i, а сумма $\sum_i M^n(i)$ не превосходит 2^{-n}. Сложив теперь все M^n, мы получим перечислимую снизу полумеру M, для которой $r_i/M(i) \to 0$. Следовательно, это верно и для максимальной полумеры. ▶

Отсюда вытекает (третье по счёту) доказательство того, что сумма двух неслучайных перечислимых снизу чисел неслучайна: сумма двух последовательностей, сходящихся к нулю, тоже сходится к нулю.

Ещё отсюда следует, что функции Соловея существуют (что само по себе не очевидно, если посмотреть на определение). Достаточно взять перечислимое снизу случайное число и представить его в виде суммы вычислимого ряда с положительными рациональными членами. Более того, это рассуждение позволяет построить неубывающую функцию Соловея, поскольку ряд можно взять с невозрастающими членами (разбивая каждое слагаемое на равные части, меньшие предыдущих частей).

$\boxed{\textbf{164}}$ Пусть U — оптимальный беспрефиксный способ описания. Рассмотрим функцию $f(p,x,n)$, которая равна $l(p)$, если U на входе p останавливается за n шагов и даёт x, и равна (скажем) $2l(p) + 2l(x) + 2\log n$ в противном случае. Покажите, что $f(p,x,n)$ является верхней оценкой для $KP(p,x,n)$, которая точна, когда p является кратчайшим описанием x и его обработка требует n шагов. Выведите отсюда существование функций Соловея.

Ещё одно следствие: медленная сходимость ряда в смысле Соловея — свойство не самого ряда, а его суммы. Это выглядит странно: формулировка свойства, согласно которой *бесконечно много членов ряда приближаются к верхней границе, заданной априорной вероятностью*, говорит о его членах, и кажется, что разбив члены на меньшие, можно это свойство нарушить. Тем не менее это не так; в следующем разделе мы попытаемся понять механизм этого явления, дав прямое доказательство (и получив заодно некоторое усиление результатов этого раздела).

5.7.6. Свойство Соловея для ряда определяется его суммой

Отметим для начала такой простой факт: медленная сходимость ряда сохраняется при вычислимой перестановке его членов. Это сразу следует из того, что вычислимая перестановка π меняет априорную вероятность не более чем в константу раз: $m(\pi(i)) = \Theta(m(i))$. Более интересен вопрос о группировке членов. Поскольку мы хотим разрешить бесконечные группы, рассмотрим двумерный вычислимый ряд $\sum_{i,j} a_{ij}$ с неотрицательными рациональными членами. Его можно разбить на группы:

$$\alpha = \sum_{i,j} a_{ij} = (a_{00} + a_{01} + \ldots) + (a_{10} + a_{11} + \ldots) + \ldots = \sum_i A_i,$$

где $A_i = \sum_j a_{ij}$.

Мы хотим показать, что ряды $\sum A_i$ и $\sum a_{ij}$ одновременно обладают (или не обладают) свойством Соловея. (Мы уже видели, что это свойство сохраняется при перестановке, поэтому корректно определено для второго ряда.)

Тут, правда, нужны некоторые дополнительные пояснения и условия. Во-первых, $\sum A_i$ не является вычислимым рядом, а только перечислим снизу. Это, однако, не мешает распространить на него определение медленной сходимости: перечислимые ряды тоже оцениваются сверху априорной вероятностью, и мы требуем, чтобы эта оценка была $O(1)$-точна на бесконечном множестве. Дав такое определение, мы видим, что интересующее нас утверждение неверно без дополнительных предположений: пусть, например, все ненулевые члены входят в группу A_1; тогда ряд $\sum A_i$ сходится с максимально возможной скоростью, а ряд $\sum a_{ij}$ может сходится медленно. Следующая теорема указывает такие дополнительные предположения:

Теорема 112. *Пусть каждая группа A_i ряда содержит лишь конечное число ненулевых членов. Тогда свойства $A_i/m(i) \to 0$ и $a_{ij}/m(i,j) \to 0$ равносильны.*

Здесь $m(i,j) = 2^{-KP(i,j)+O(1)}$ — априорная вероятность пары $\langle i,j\rangle$ (или её кода при каком-то вычислимом кодировании).

◀ Вспомним прежде всего, что $m(i) = \sum_j m(i,j)$ с точностью до ограниченного множителя. (В самом деле, правая часть есть перечислимый снизу сходящийся ряд. С другой стороны, уже первый член $m(i,0)$ равен $m(i)$ с интересующей нас точностью.) Поэтому $a_{ij}/m(i,j) \to 0$ влечёт $A_i/m(i) \to 0$: лишь для конечного число пар $\langle i,j\rangle$ выполнено неравенство $a_{i,j} > \varepsilon m(i,j)$, и эти пары входят лишь в конечное число групп.

Сложнее обратное утверждение: $A_i/m(i) \to 0$ влечёт $a_{ij}/m(i,j) \to 0$. (Именно здесь нам важно, что в каждой группе лишь конечное число ненулевых членов.) Нам нужно построить перечислимую снизу функцию $\tilde{m}(i,j)$ с конечной суммой, для которой $a_{ij}/\tilde{m}(i,j) \to 0$, используя тот факт, что $A_i/m(i) \to 0$. Естественная идея: разбить $m(i)$ в сумму $\sum \tilde{m}(i,j)$ в той же пропорции, в которой A_i разбивается в $\sum a_{ij}$. Для этого, однако, мы должны уметь вычислять A_i, а не только перечислять снизу. Это было бы легко, если бы мы знали, сколько ненулевых членов есть среди a_{ij} при данном i, но эта информация не вычислима, так что нам придётся действовать иначе. (Для группировки ряда в группы с конечным числом членов этот способ проходит.)

Будем действовать «с другого конца». Мы можем положить $\tilde{m}(i,j)$ равным ca_{ij} для некоторой константы c, пока это не нарушает свойства $\sum_j \tilde{m}(i,j) \leqslant m(i)$. Глядя на рост $m(i)$ в правой части, мы позволяем левой части подрасти соответствующим образом, в благоприятных условиях до ca_{ij}. Если (согласно нашему предположению) $A_i/m(i) \to 0$, то для любой константы c при всех достаточно больших i выполнено неравенство $cA_i \leqslant m(i)$ и граница $m(i)$ не помешает росту. Следовательно, $a_{ij}/\tilde{m}(i,j) \leqslant 1/c$ при всех достаточно больших i, и потому это верно для всех пар $\langle i,j\rangle$, кроме конечного числа (при каждом i лишь конечное число ненулевых членов). Таким образом, для каждого c мы построили свою полумеру $\tilde{m}_c(i,j)$, для которой $a_{ij}/\tilde{m}_c(i,j) \leqslant 1/c$ почти всюду (а сумма не больше $\sum m(i) \leqslant 1$). Остаётся выполнить это построение при всех $c = 2^{2^n}$ и сложить $m_{2^{2n}}$ с коэффициентами 2^{-n}. ▶

Как следствие этой теоремы мы получаем другое доказательство того, что свойство Соловея для вычислимого ряда зависит только от суммы ряда. В самом деле, если для двух вычислимых рядов $\sum_i a_i = \sum_j b_j$, то эти ряды могут быть получе-

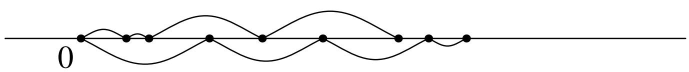


Рис. 5.3. Два ряда с одинаковой суммой могут быть представлены как результаты различных группировок третьего.

ны различной группировкой слагаемых (отметим все точки деления и группируем интервалы между ними по-разному, рис. 5.3).

Помимо альтернативного доказательства инвариантности, мы получаем некоторое усиление теоремы 111: в ней рациональные числа r_i (члены ряда) были вычислимы, а теперь их можно задавать перечислением снизу, при условии, что каждое из них меняется (возрастает) лишь конечное число раз. Точнее говоря, мы требуем, чтобы $r_i = \lim_n r(i,n)$, где r — вычислимая функция двух натуральных аргументов с рациональными значениями, неубывающая по n и принимающая при каждом i (и всевозможных n) лишь конечное число значений. Тогда неслучайность числа $\sum r_i$ равносильна условию $r_i/m(i) \to 0$. В самом деле, каждое r_i можно представить как сумму вычислимого ряда из приращений, в котором лишь конечное число членов отлично от нуля. После этого замена ряда на двумерный не меняет ни свойства Соловея (как мы видели в предыдущей теореме), ни суммы (очевидным образом).

Вспомним, что перечислимая сверху функция $f(n)$ с натуральными аргументами является верхней оценкой для $KP(n)$ с точностью до $O(1)$, если и только если $\sum_n 2^{-f(n)} < \infty$ (теорема 62). Теперь мы можем дополнить это утверждение:

Теорема 113. *Эта оценка точна для бесконечно многих n ($KP(n) \geqslant f(n) - c$ для некоторого c и бесконечно многих n), если и только если сумма $\sum_n 2^{-f(n)}$ случайна.*

◀ В самом деле, убывающие целочисленные верхние оценки для $f(n)$ дают как раз возрастающие (и меняющиеся конечное число раз) нижние оценки для $2^{-f(n)}$, так что можно сослаться на только что доказанное утверждение. ▶

5.7.7. Регуляторы сходимости и функция $BB(n)$

Мы уже видели несколько разных уточнений интуитивной идеи «медленно сходящегося ряда». Но, пожалуй, самый естественный подход мы ещё не пробовали. Он состоит в следующем. Если $a_n \to \alpha$, то для всякого $\varepsilon > 0$ найдётся N, для которого $|a_n - \alpha| < \varepsilon$ при всех $n > N$. Минимальное такое N (рассматриваемое как функция от ε) называют *регулятором сходимости* a_n к α, и сходимость естественно считать медленной, если регулятор быстро растёт (с убыванием ε). Сейчас мы покажем, как в этих терминах можно характеризовать свойство Соловея.

Вспомним, что в разделе 1.2 (с. 30) мы определили $B(n)$ как наибольшее число, сложность которого не превосходит n. Там речь шла о простой сложности (мы ещё не рассматривали других), но аналогичное определение можно дать и для префиксной сложности: обозначим через $BP(n)$ наибольшее число, префиксная сложность которого не превосходит n.

165 Фиксируем оптимальный беспрефиксный декомпрессор, задающий префиксную сложность. Покажите, что максимальное время его работы на описаниях длины не более n заключено между $BP(n-c)$ и $BP(n+c)$ при некотором c и всех n. [Указание: годится то же рассуждение, что и для простой сложности (см. раздел 1.2).]

Докажем теперь эквивалентность различных вариантов понятия «медленно сходящегося ряда».

Теорема 114. *Вычислимый ряд с неотрицательными рациональными членами $\sum r_i$ обладает свойством Соловея ($\Leftrightarrow$ имеет случайную сумму) тогда и только тогда, когда его регулятор сходимости быстро растёт: $N(2^{-m}) > BP(m-c)$ для некоторого c и для всех m.*

◀ Пусть $\alpha = \sum r_i = \lim a_i$, где $a_i = r_0 + \ldots + r_{i-1}$. Предположим, что число α случайно. Мы должны показать, что из $|\alpha - a_i| < 2^{-m}$ следует $KP(i) > m - O(1)$, и тем самым установить, что $N(2^{-m}) \geqslant BP(m - O(1))$. Поскольку $KP(i) \geqslant KP(a_i) + O(1)$, достаточно доказать, что любое рациональное 2^{-m}-приближение к α имеет префиксную сложность не меньше $m - O(1)$. Это немного более сильное утверждение, чем в теореме Левина – Шнорра, где речь не о произвольном приближении, а о первых m битах. Тем не менее оно может быть доказано примерно тем же способом. Вот как это делается.

Пусть c — произвольное натуральное число. Рассмотрим эффективно открытое множество U_c, которое строится так. Для каждого рационального r возьмём окрестность с центром в r и радиусом $2^{-KP(r)-c}$; множество U_c представляет собой объединение всех таких окрестностей. Поскольку функция KP перечислима сверху, это объединение эффективно открыто. Общая длина всех окрестностей равна $2 \cdot 2^{-c} \sum_r 2^{-KP(r)} \leqslant 2^{-(c-1)}$. Если бы α принадлежало бы всем U_c, то оно не было бы случайным. Значит, найдётся U_c, которому оно не принадлежит — а это как раз означает, что префиксная сложность всех 2^{-m}-приближений к α не меньше $m - O(1)$.

В другую сторону мы можем использовать теорему Левина – Шнорра без каких-либо изменений: если $N(2^{-m}) > BP(m-c)$, то $KP(i) \geqslant m - c$ для любого i, при котором a_i приближает α с ошибкой не более 2^{-m}. Следовательно, m-битовое начало двоичного разложения α имеет сложность не менее $m - O(1)$: зная его, мы можем дождаться появления a_i, которое отличается от α не более чем на 2^{-m}. ▶

Замечание. Мы установили эквивалентность двух понятий «медленной сходимости» для вычислимых рядов с неотрицательными членами. Однако доказательство не совсем прямое: в качестве промежуточного этапа используется понятие случайности. Было бы интересно найти прямое доказательство, а также связать сводимость по Соловею (для произвольных двух чисел, а не только полных) со скоростью роста регуляторов.

Переводя определение $BP(m)$ на язык априорной вероятности, мы можем определить $BP(m)$ как минимальное N, для которого все $n > N$ имеют априорную вероятность менее 2^{-m}. Но раз уж мы перешли к априорной вероятности, то логичнее рассмотреть другое понятие: обозначим через $BP'(m)$ минимальное N, для

которого *сумма априорных вероятностей* всех $n > N$ меньше 2^{-m}. Вообще говоря, $BP'(m)$ может быть больше $BP(m)$, но оказывается, что $BP'(m)$ тоже может быть использовано для характеризации случайности:

Теорема 115. *Пусть a_i — вычислимая возрастающая последовательность рациональных чисел, которая сходится к случайному числу α. Тогда $N(2^{-m}) \geqslant BP'(m-c)$ для некоторого c и для всех m.*

◀ Числа $i > N(2^{-m})$ имеют (с точностью до константы) ту же априорную вероятность, что и соответствующие a_i. Поэтому достаточно доказать, что сумма априорных вероятностей всех рациональных чисел в 2^{-m}-окрестности числа α есть $O(2^{-m})$. (В самом деле, при $i > N(2^{-m})$ число a_i попадает в эту окрестность.)

Как обычно, мы будем рассуждать с другого конца и покроем все α, для которых это свойство неверно, семейством интервалов малой меры. Зафиксируем некоторое c и рассмотрим все интервалы с рациональными концами, обладающие следующим свойством: *сумма априорных вероятностей всех рациональных чисел внутри интервала по крайней мере в $c/2$ раз превышает его длину*. Мы докажем, что мера объединения всех таких интервалов есть $O(1/c)$ (на самом деле не превосходит $4/c$).

Достаточно доказать это для конечного набора интервалов с этим свойством. Более того, можно считать, что в этом наборе нет ни одного лишнего интервала (от выбрасывания которого объединение не уменьшится). Упорядочим интервалы набора в порядке возрастания левых концов:

$$(l_0, r_0), (l_1, r_1), (l_2, r_2), \ldots,$$

где $l_0 \leqslant l_1 \leqslant l_2 \leqslant \ldots$; легко видеть, что правые концы также неубывают (иначе один из интервалов был бы лишним): $r_0 \leqslant r_1 \leqslant r_2 \leqslant \ldots$ Теперь заметим, что если взять интервалы через один, то они будут непересекающимися (если два интервала через один пересекаются, то интервал между ними, очевидно, лишний). Соответственно мы получаем два семейства непересекающихся интервалов: с чётными номерами и с нечётными номерами. Рассмотрим одно из семейств; в нём длина каждого интервала в $c/2$ раз меньше, чем сумма априорных вероятностей всех рациональных чисел внутри него, и потому сумма длин всех интервалов не больше $2/c$. Всего для двух семейств получается оценка $4/c$, как мы и обещали.

Остаётся заметить, что случайное α не покрывается этими интервалами при некотором c. ▶

С этим утверждением связаны различные естественные вопросы (на которые авторы не знают ответа). Насколько сильно могут отличаться $BP(m)$ и $BP'(m)$? Можно ли в описании $BP(m)$ в терминах времени работы машин использовать префиксно-корректные машины вместо беспрефиксных? Можно ли вывести последний результат из теоремы Левина – Шнорра с априорной сложностью? Можно ли использовать аналогичные методы, чтобы установить, что $\sum_{x\in X} 2^{-l(x)}$ случайно для всякого беспрефиксного множества X, которое содержит область определения оптимального беспрефиксного декомпрессора?

Возвращаясь к «философскому смыслу» числа Ω, заметим, что его можно рассматривать как бесконечный аналог объектов сложности n, рассмотренных в теореме 15 (с. 36). Между ними есть и более формальная связь.

Теорема 116. *Пусть Ω_n — двоичное слово, составленное из первых n битов двоичной записи числа Ω. Тогда Ω_n обладает свойствами, указанными в теореме 15 с точностью до логарифмической поправки в значении n: любой из указанных в этой теореме объектов может быть алгоритмически получен из Ω_n при известном n и наоборот (в обоих случаях — с поправкой $O(\log n)$ в значении n).*

◀ Мы уже видели, что по Ω_n можно построить число t, большее $BP(n)$ (номер приближения, которое превысит Ω_n). Разница между префиксной и обычной сложностью (из-за которой $B(n)$ может быть больше $BP(n)$) компенсируется изменением n на $O(\log n)$.

Теперь покажем, что верно и обратное: зная $B(n)$ и n, можно найти $\Omega_{n-O(\log n)}$. В самом деле, сделаем $B(n)$ шагов приближения снизу к Ω. Надо проверить, что в полученном приближении снизу по крайней мере $n - O(\log n)$ знаков будут «окончательными» (то есть будут совпадать с соответствующими знаками числа Ω). Если это не так, то существует конечная двоичная дробь β из $n - O(\log n)$ знаков, разделяющая достигнутое приближение и окончательное значение Ω. Эта дробь имеет сложность $n - O(\log n)$; с другой стороны, зная β, мы можем найти число, большее $B(n)$ — достаточно подсчитать число шагов, после которых приближение к Ω впервые пересекает границу β. При достаточно большой константе в $O(\log n)$ эти два обстоятельства противоречат друг другу. ▶

Отсюда видно, что знание $n + O(\log n)$ битов в числе Ω позволяет ответить на любой вопрос о завершении работы программы длины не более n. Поскольку любой перечислимый вопрос, например, вопрос о выводимости формул длины n, может быть представлен в таком виде, то — при достаточном художественном воображении — можно, следуя Чейтину, назвать число Ω «числом мудрости», в котором содержится информация обо всём на свете.

Не вдаваясь в философию, можно отметить, что мы доказали, что число Ω эквивалентно $\mathbf{0}'$ по Тьюрингу (это значит, что можно вычислять биты Ω, имея оракул, отвечающий на вопрос об остановке любых программ, и наоборот).

5.8. Эффективная размерность Хаусдорфа

В классической теории меры хорошо известно понятие *размерности Хаусдорфа* (часто упоминаемое в связи с «фракталами»). Его можно определить так. Пусть $\alpha > 0$ — некоторое действительное число. Будем говорить, что множество A является *α-нулевым*, если для всякого $\varepsilon > 0$ его можно покрыть счётным семейством интервалов I_k с

$$\sum_k \mu(I_k)^\alpha < \varepsilon.$$

Это определение предполагает, что A является подмножеством некоторого пространства, в котором выделен класс «интервалов» с определённой на них мерой μ.

Мы ограничимся случаем канторовского пространства, считая, что интервалами являются множества Ω_x, состоящие из бесконечных продолжений слова x. Мера интервала Ω_x равна $2^{-l(x)}$.

Несколько простых замечаний.

(1) Подмножество α-нулевого множества является α-нулевым множеством.

(2) При $\alpha = 1$ мы получаем обычное определение множества меры нуль.

(3) При $\alpha > 1$ любое подмножество $A \subset \Omega$ является α-нулевым (его можно покрыть 2^n интервалами, соответствующими 2^n словам длины n, и сумма α-степеней их мер стремится к нулю при $n \to \infty$).

(4) Если $0 < \alpha < \alpha'$, то любое α-нулевое множество является и α'-нулевым (мера $\mu(I)$ интервала I не превосходит 1, и потому $\mu(I)^{\alpha'} < \mu(I)^{\alpha}$).

166 Предложите естественное определение α-нулевого множества для подмножеств прямой и проверьте, что множество $A \subset [0, 1]$ является α-нулевым тогда и только тогда, когда множество двоичных записей всех чисел из A является α-нулевым согласно определению. [Указание: по существу надо проверить, что разрешение рассматривать произвольные промежутки на прямой, а не только половины, четверти и т.п., не меняет класса нулевых множеств.]

Из наших замечаний следует, что для любого множества $A \subset \Omega$ есть некоторая граница $d \in [0, 1]$ с таким свойством: при $\alpha > d$ множество A является α-нулевым, а при $\alpha < d$ — нет. (При $\alpha = d$ множество может быть α-нулевым, а может и не быть.) Эта граница называется *хаусдорфовой размерностью* множества A.

167 *Канторово множество* на отрезке получается, если выбросить из отрезка среднюю треть $(1/3, 2/3)$, у каждого из двух полученных отрезков выбросить среднюю треть и так далее. Докажите, что получится компактное множество, гомеоморфное Ω и имеющее хаусдорфову размерность $\log_3 2$.

[Указание. Верхняя оценка для размерности получается, если в качестве покрытия использовать все отрезки, оставшиеся к некоторому шагу — точнее, немного бóльшие их интервалы. Чтобы получить нижнюю оценку, заметим, что (1) в силу компактности можно рассматривать конечные покрытия интервалами; (2) если есть покрытие канторова множества, то можно посмотреть отдельно на части этого покрытия, покрывающие первую и последнюю треть множества; любую из них можно увеличить в три раза до покрытия целого множества в силу самоподобия; если α меньше границы, то одно из покрытий будет лучше исходного, и так можно увеличивать размер интервалов, в конце концов придя к противоречию.]

168 Предложите естественное определение размерности множеств в пространстве ($\mathbb{R}^3$) по Хаусдорфу. Объясните, почему мера тел равна 3, поверхностей — 2, линий — 1, а точек — 0. Покажите, что в пространстве существует множество любой размерности от 0 до 3.

Теперь понятно, каким будет эффективный аналог понятия хаусдорфовой размерности (см. [163, 130]). Будем говорить, что множество $A \subset \Omega$ является *эффективно α-нулевым* (для данного $\alpha > 0$), если существует алгоритм, который по рациональному $\varepsilon > 0$ указывает покрытие множества A интервалами, у которых сумма α-степеней мер не превосходит ε.

Как и в классическом случае, при увеличении α (и при уменьшении A) это свойство сохраняется. Принципиальная разница с классическим определением (как и для нулевых множеств) возникает дальше:

Теорема 117. *При любом рациональном $\alpha > 0$ существует наибольшее по включению эффективно α-нулевое множество.*

◀ Доказательство этой теоремы повторяет аналогичное рассуждение для эффективно нулевых множеств в главе 3. Счётное объединение α-нулевых множеств (классических) является α-нулевым; аналогичным образом объединение перечислимого семейства эффективно α-нулевых множеств является эффективно α-нулевым. С другой стороны, при рациональном (и даже при любом вычислимом) α можно эффективно перечислить все эффективно α-нулевые множества (точнее, алгоритмы, порождающие необходимые покрытия) — надо перечислять все, при этом корректируя излишние меры при необходимости. ▶

169 Покажите, что это наибольшее множество состоит из тех и только тех последовательностей ω, для которых разность $\alpha n - KP((\omega)_n)$ не ограничена сверху. [Указание: рассуждение аналогично доказательству теоремы Левина – Шнорра в варианте для префиксной сложности.]

Следующий результат А. Ходырева не используется в дальнейшем (для определения эффективной хаусдорфовой размерности достаточно ограничиться рациональными α), но любопытен сам по себе. Пусть α — произвольное действительное число.

Теорема 118. *Наибольшее эффективно α-нулевое множество существует тогда и только тогда, когда α перечислимо снизу.*

◀ Пусть α перечислимо снизу. Это значит, что постепенно мы получаем всё более и более точные приближения к α с недостатком. Тем самым требования к покрытию (сумма α-степеней мер любого конечного числа интервалов меньше ε) постепенно ослабевают, и мы можем «снять с полки» (как говорили про фильмы во время перестройки) и пропустить на выход забракованные ранее, а ныне признанные безопасными интервалы. Легко понять, что если алгоритм удовлетворяет требованиям для предельного α, то рано или поздно любой порождённый им интервал будет пропущен.

Напротив, пусть существует наибольшее эффективно α-нулевое множество. Рассмотрим соответствующий ему алгоритм. Его можно использовать для получения нижних оценок на α. В самом деле, если этот алгоритм при некотором ε выдал интервалы, для которых сумма β-степеней мер больше ε, то это означает только одно: выбранное нами β меньше α. Пробуя разные β, разные ε и разные конечные наборы интервалов, мы получим перечислимое множество оценок снизу для α.

Остаётся лишь показать, что эти оценки могут сколь угодно близко подходить к α. Предположим, что это не так и что они все меньше некоторого $\alpha' < \alpha$. В этом случае всякое эффективно α-нулевое множество будет и эффективно α'-нулевым, чего не бывает, как мы увидим дальше (существуют множества любой эффективной размерности, см. задачу 170, с. 199). ▶

Теперь естественно дать такое определение. *Эффективной хаусдорфовой размерностью* множества $A \subset \Omega$ называется точная нижняя грань всех α, при которых A является эффективно α-нулевым. Это число находится в промежутке $[0, 1]$ и не меньше (классической) хаусдорфовой размерности. (Первоначально определение эффективной хаусдорфовой размерности было предложено в терминах вычислимых мартингалов, см. [93, 100], где установлены приводимые ниже свойства размерности; об этом варианте определения см. далее в разделе 9.10.)

В своё время мы говорили, что свойство множества быть эффективно нулевым парадоксальным образом зависит не от того, много ли в нём элементов, а от того, какие это элементы — случайные или нет. Подобным образом обстоит дело и с эффективной размерностью.

Теорема 119. *Эффективная хаусдорфова размерность множества есть точная верхняя грань эффективных хаусдорфовых размерностей его элементов.*

(Говоря об эффективной хаусдорфовой размерности точки $\omega \in \Omega$, мы имеем в виду размерность синглетона $\{\omega\}$.)

◀ Очевидно, размерность множества не меньше размерности любого его подмножества. Остаётся проверить, что если размерности всех синглетонов, образованных элементами некоторого множества A, меньше некоторого рационального r, а r' больше r (и тоже рационально), то размерность множества в целом не превосходит r'. А это сразу следует из теоремы 117: все синглетоны лежат в наибольшем эффективно r'-нулевом множестве, поэтому A является его подмножеством и имеет размерность не больше r'. ▶

Как же найти эффективную размерность синглетона? Оказывается, она имеет следующее простое описание в терминах колмогоровской сложности.

Теорема 120. *Эффективная хаусдорфова размерность множества $\{\omega\}$, где $\omega = \omega_0\omega_1\omega_2\ldots$, равна*

$$\liminf_{n\to\infty} \frac{KS(\omega_0\omega_1\ldots\omega_{n-1})}{n}.$$

(В формулировке теоремы говорится о простой колмогоровской сложности начальных отрезков. Однако это не принципиально: поскольку разница между различными вариантами сложности имеет порядок $O(\log n)$ для слов длины n, при делении на n и переходе к пределу она исчезает.)

◀ Надо доказать неравенство в две стороны.

Предположим, что нижний предел меньше некоторого рационального r. Надо проверить, что множество $\{\omega\}$ является эффективно r'-нулевым для любого рационального $r' > r$.

Рассмотрим для каждого n все слова длины n, имеющие сложность меньше rn. Таких слов не больше $O(2^{rn})$. Условие про нижний предел гарантирует, что при бесконечно многих n начальный отрезок последовательности ω оказывается в числе таких слов. Рассмотрим все интервалы Ω_z, порождённые этими словами z. Мера каждого из Ω_z в степени r' есть $2^{-r'n}$, и суммарная r'-мера равна $O(2^{(r-r')n})$, что с ростом n убывает как геометрическая прогрессия. Ряд $\sum_n 2^{(r-r')n}$ сходится, и если

отбросить некоторый его начальный кусок (рассматривать слова длины N и более), то суммарная r'-мера этих слов может быть сделана сколь угодно малой (выбором достаточно большого N). Между тем по предположению о нижнем пределе остаток будет по-прежнему покрывать ω. В одну сторону неравенство доказано.

Напротив, пусть $\{\omega\}$ имеет эффективную размерность меньше r для некоторого рационального r. Покажем, что нижний предел, о котором идёт речь в теореме, не превосходит r.

По определению для каждого рационального $\varepsilon > 0$ можно эффективно перечислять последовательность интервалов, про которую известно, что один из них содержит ω и сумма ряда из r-степеней длин не больше ε. Сделаем это для $\varepsilon = 1, 1/2, 1/4, \ldots$ и получим последовательность интервалов, у которых сумма r-степеней длин ограничена и которые покрывают ω бесконечно много раз. Другими словами, мы нашли вычислимую последовательность слов $x_0, x_1, \ldots$ с такими свойствами:

- $\sum 2^{-rl(x_i)} < \infty$;
- x_i является началом ω при бесконечно многих i.

Из первого условия следует, что $m(i) \geqslant c2^{-rl(x_i)}$ при некотором положительном c и всех i (здесь m — наибольшая полумера на натуральных числах, рассмотренная в главе 4). Переходя к логарифмам, получаем оценку для префиксной сложности $KP(x_i) \leqslant KP(i) + O(1) \leqslant rl(x_i) + O(1)$. Остаётся заметить, что длины x_i стремятся к бесконечности (поскольку ряд сходится), что среди x_i имеется бесконечно много начал последовательности ω и что простая сложность не превосходит префиксной. (А также вспомнить определение нижнего предела и заметить, что если последовательность бесконечно много раз не превосходит r, то её нижний предел не превосходит r.) ▶

170 Выведите из этой теоремы, что для любого действительного $\alpha \in [0, 1]$ существует множество (и даже одноэлементное множество) с эффективной размерностью α. [Указание: сложность начальных отрезков можно увеличивать, добавляя случайные биты, и уменьшать, добавляя нулевые биты.]

171 Покажите, что для эффективно замкнутого подмножества в пространстве Ω (дополнение которого представляет собой объединение перечислимого семейства интервалов) хаусдорфова размерность совпадает с эффективной хаусдорфовой размерностью. [Указание: покрытия в силу компактности можно считать конечными, и искать их перебором.]

172 Найдите хаусдорфову размерность канторовского множества (задача 167), используя предыдущую задачу и определение эффективной размерности в терминах синглетонов.

173 Покажите, что для всякого $\alpha \in [0, 1]$ существует множество с (классической) хаусдорфовой размерностью α. [Указание. Можно взять множество последовательностей, у которых на некоторых местах стоят нули.]

174 Докажите, что определение эффективной хаусдорфовой размерности не изменилось бы, если вместо наличия для каждого ε покрытия с суммой степеней

меньше ε мы бы требовали наличия одного покрытия с конечной суммой степеней, но покрывающего каждый элемент множества бесконечно много раз. [Указание. Если такое покрытие есть для некоторого α, то для большего α' те же самые интервалы дадут меньшие члены ряда, и уменьшение будет тем больше, чем длиннее двоичное слово. Остаётся заметить, что можно выбросить из покрытия все короткие двоичные слова, так как оно было в бесконечное число слоёв.]

Мы вернёмся к понятию эффективной размерности, когда будем говорить о связи случайности с эффективными мартингалами (раздел 9.5 и далее). Там же будет приведено доказательство теоремы 120 (по существу то же самое) на языке мартингалов.

5.9. Случайность по различным мерам

5.9.1. Переход от одной меры к другой

Понятие случайности зависит от меры: по разным мерам случайны разные последовательности. Например, усиленный закон больших чисел гарантирует, что для различных значений p множества случайных по бернуллиевой мере B_p последовательностей не пересекаются, поскольку предел частоты единиц у них разный. Тем не менее с точки зрения алгоритмических свойств случайные последовательности относительно вычислимых мер устроены в некотором смысле одинаково, с одной оговоркой.

Оговорка такая: если некоторая последовательность имеет положительную вероятность по вычислимой мере, то она случайна по этой мере (поскольку, очевидно, не может содержаться в нулевом множестве). Такая последовательность обязательно вычислима. Это можно доказать, сославшись на пункт (ж) теоремы 79, а также с помощью такого простого рассуждения. Пусть $\{\omega\}$ имеет положительную вероятность ε относительно вычислимой меры μ. Посмотрим на μ-меры множеств Ω_x, где x — начала ω. С ростом x они убывают, оставаясь большими или равными ε. Дождёмся момента, когда они станут меньше $1{,}1\varepsilon$. У каждого такого начала x лишь одно из слов $x0$ и $x1$ имеет меру больше $0{,}9\varepsilon$, и его можно найти, раз мера вычислима, поэтому с этого места последовательность можно вычислимо продолжать.

Исключим этот особый случай и будем рассматривать только *безатомные* вычислимые меры, для которых одноэлементные множества имеют меру 0. Тогда множества случайных последовательностей по двум таким мерам находятся в естественном взаимно однозначном соответствии.

Теорема 121. *Пусть μ_1 и μ_2 — две безатомные вычислимые меры. Тогда существует взаимно однозначное соответствие между множествами МЛ-случайных по μ_1 и μ_2 последовательностей, которое в обе стороны задаётся вычислимым отображением $\Sigma \to \Sigma$.*

Другими словами, существуют две машины M_{12} и M_{21} с оракулом, при этом если оракулом является случайная по μ_1 последовательность ω, то M_{12}^{ω} будет случайной по μ_2 последовательностью, и наоборот, причём эти отображения (на случайных последовательностях) взаимно обратны.

◀ Следуя [196], построим взаимно однозначное соответствие между последовательностями, случайными по мере μ_1, и случайными точками отрезка $[0,1]$ (по равномерной мере). А именно, как мы уже не раз делали, разобьём отрезок $[0,1]$ на два: левый π_0 длиной $\mu_1(\Omega_0)$ и правый π_1 длиной $\mu_1(\Omega_1)$, каждый из них разобьём на две части аналогичным образом и так далее. Последовательности ω поставим в соответствие общую точку всех отрезков π_x, где x — начала ω. Условие безатомности гарантирует, что эта общая точка единственна.

Это отображение является изоморфизмом пространств с мерой, а именно пространства Ω с мерой μ_1 и отрезка $[0,1]$ с равномерной мерой. Оно не является взаимно однозначным, так как концам отрезков соответствуют сразу две последовательности, но это множество меры нуль. Вычислимость меры гарантирует, что при этом эффективно нулевые множества соответствуют эффективно нулевым, а случайные последовательности — случайным, и тут уже нет неоднозначности (концы отрезков вычислимы и не случайны, как и соответствующие им последовательности вида $x000\ldots$ и $x111\ldots$).

Остаётся сделать так для обеих мер μ_1 и μ_2 и воспользоваться отрезком с равномерной мерой на нём как промежуточным шагом. Вычислимость соответствующих отображений также следует из вычислимости мер μ_1 и μ_2. ▶

В теории вычислимости говорят, что две последовательности (два множества натуральных чисел: последовательность отождествляется с множеством позиций, где в ней единицы) *эквивалентны по Тьюрингу* (лежат в одной тьюринговой степени), если каждая из них вычислима относительно другой (= с оракулом для другой). Классы эквивалентности называются *тьюринговыми степенями*. Доказанная теорема показывает, в частности, что множество тьюринговых степеней случайных по (безатомной вычислимой) мере μ последовательностей не зависит от выбора меры μ.

175 Докажите, что любая последовательность, случайная по некоторой вычислимой мере P (не обязательно безатомной), либо вычислима, либо эквивалентна по Тьюрингу последовательности, случайной по равномерной мере. [Указание. Построим вложенные отрезки для последовательности (обозначим её ω), как это было описано выше, и рассмотрим их общую точку. Если она не единственна, то ω вычислима. Если общая точка z единственна, то она случайна по мере Лебега на отрезке и вычислима с оракулом для ω. Напротив, ω вычислима, если даны приближения к z: поскольку мы знаем, что z случайна, то z не может совпадать с (вычислимыми) границами отрезков.]

5.9.2. Последовательности, не случайные по вычислимым мерам

Пусть дана некоторая последовательность ω. Всегда ли можно подобрать вычислимую меру, относительно которой последовательность ω будет случайной? Оказывается, что нет.

Теорема 122. *Существует бесконечная последовательность нулей и единиц, не случайная ни по какой вычислимой мере на Ω.*

Последовательности, случайные по некоторой вычислимой мере, были названы «правильными» в статье [196] (в английском переводе этой статьи использован термин “proper”).

Есть разные способы построить «неправильную» последовательность. Наиболее наглядный, пожалуй, использует *априорный дефект случайности*.

Критерий случайности (в смысле Мартин-Лёфа) по вычислимой мере P можно переформулировать следующим образом. Для каждого слова x рассмотрим величину

$$d_P(x) = -\log_2 P(\Omega_x) - KA(x).$$

Назовём её *дефектом случайности* двоичного слова x относительно вычислимой меры P. Объяснение этого названия: последовательность ω случайна (в смысле Мартин-Лёфа) тогда и только тогда, когда дефекты случайности её начальных отрезков ограничены.

Слова «дефект случайности» могут употребляться в самых разных смыслах; мы уже говорили об ограниченных в среднем и по вероятности дефектах для бесконечных последовательностей, а в главе 14 мы будем говорить о дефекте случайности элемента внутри конечного множества. Но в этом разделе, говоря о дефекте случайности, мы имеем в виду именно функцию d_P, определённую выше.

Приведённое определение предполагает, что $P(\Omega_x) > 0$; если $P(\Omega_x) = 0$, то дефект естественно считать бесконечным.

Дефект случайности всегда неотрицателен (с точностью до константы, см. теорему 89).

$\boxed{\textbf{176}}$ Докажите, что для любого слова x хотя бы одно из слов $x0$ и $x1$ имеет не больший дефект случайности, чем само x. (Вычислимая мера, относительно которой вычисляется дефект, фиксирована.)

Согласно этой задаче, мы можем начать с любого слова с конечным дефектом случайности (ненулевой мерой) и добавлять к нему бит за битом, не увеличивая дефекта. В результате, согласно критерию случайности, получится случайная в смысле Мартин-Лёфа последовательность (по выбранной заранее мере, относительно которой измеряется дефект).

Введя понятие дефекта, вернёмся к доказательству теоремы 122.

◄ Чтобы построить «неправильную» последовательность ω, нам надо позаботиться, чтобы для каждой вычислимой меры P у ω были начальные отрезки со сколь угодно бо́льшим дефектом относительно P. Получается счётное число требований: для каждой меры P и для каждой константы c мы должны обеспечить начальный отрезок, у которого дефект относительно P не меньше c.

Применяя диагональную конструкцию, мы строим искомую последовательность постепенно: на каждом шаге мы прибавляем к уже построенной части кусок, после которого выполнено очередное требование. Таким образом, нужно проверить, что для любого конечного слова x и любой вычислимой меры P существует продолжение y слова x, имеющее сколь угодно большой дефект случайности относительно P. В самом деле, будем продолжать x так, чтобы добавление очередного бита уменьшало меру P не менее чем (скажем) в полтора раза. Это можно сделать

вычислимо, поэтому сложность будет расти медленно, а мера убывать быстро, и в какой-то момент мы достигнем требуемого. ▶

По существу то же самое рассуждение можно изложить с помощью так называемых «генерических» последовательностей. Напомним, что подмножество A пространства Ω называется *всюду плотным*, если оно пересекается с любым интервалом. Знаменитая *теорема Бэра* утверждает, что пересечение счётного семейства всюду плотных открытых множеств (открытые множества — объединения интервалов) непусто (и даже всюду плотно).

177 Докажите это утверждение, взяв произвольное конечное слово и прибавляя к нему фрагменты так, чтобы попасть внутрь очередного открытого всюду плотного множества.

Рассмотрим теперь эффективно открытые множества (объединения перечислимых семейств интервалов) и отберём из них всюду плотные. Получится счётное семейство всюду плотных открытых множеств, которое по теореме Бэра имеет непустое (и даже всюду плотное) пересечение. Назовём последовательности, входящие в это пересечение, *генерическими*. Генерические последовательности можно неформально описать как «не подчиняющиеся никаким законам», если под законом понимать утверждение, запрещающее хотя бы одно конечное продолжение у любой конечной последовательности, и нарушения которого могут быть эффективно установлены (множество не удовлетворяющих закону последовательностей эффективно открыто).

178 Докажите, что никакая генерическая последовательность не удовлетворяет усиленному закону больших чисел. [Указание. Множество последовательностей, у которых есть начало длины более N, на 99% состоящее из единиц, является эффективно открытым и всюду плотным. Аналогично для нулей вместо единиц.]

179 Докажите, что генерическая последовательность не может быть вычислимой. [Указание. Множество всех последовательностей, отличных от данной вычислимой, эффективно открыто и всюду плотно.]

В отличие от случайности, определение генерической последовательности не предполагает никакой меры.

180 Докажите, что генерическая последовательность не может быть случайной ни по какой вычислимой мере. [Указание: достаточно построить эффективно открытое всюду плотное множество сколь угодно малой меры. Для этого можно из двух половинок каждого интервала выбирать меньшую (или чуть-чуть большую) по мере.]

В статье [196] указан другой метод построения «неправильной» последовательности: как говорит замечание после определения 4.4 в этой статье, несложно показать, что характеристическая последовательность универсального перечислимого множества не является случайной ни по какой вычислимой мере. Правда, там не указано, в каком смысле понимается универсальность. Скорее всего утверждение, которое авторы этой статьи имели в виду, вытекает из следующего результата:

181 Покажите, что существует перечислимое множество, характеристическая последовательность которого не случайна ни по какой вычислимой мере.

[Указание. Поскольку сложности начальных отрезков будут логарифмическими для любого перечислимого множества, достаточно гарантировать, что их меры быстро убывают. Разделим натуральный ряд на счётное число арифметических прогрессий; i-я из них будет отвечать за i-ю вычислимую меру. Мы будем добиваться, чтобы последовательность битов, записанная на месте i-й прогрессии, была не случайна относительно проекции i-й меры на эти координаты, выбирая направление, в котором мера заметно убывает. (Далее мы используем теорему 123). Это построение предполагает, что i-й алгоритм действительно задаёт меру; поскольку мы на самом деле этого не знаем, получается перечислимое (а не разрешимое) множество.]

Интересно, что не все перечислимые неразрешимые множества обладают указанным в этой задаче свойством:

182 Постройте перечислимое неразрешимое множество, характеристическая функция которого случайна по некоторой вычислимой мере.

[Указание (Л. Бьенвеню). Пусть a_i — вычислимая последовательность рациональных чисел, плотная в $[0, 1]$. Построим вычислимое отображение пространства Ω в себя: последовательности α, интерпретируемой как двоичная дробь в $[0, 1]$, ставится в соответствие последовательность ω, у которой $\omega_i = 1$ при $a_i < \alpha$ и $\omega_i = 0$ при $a_i > \alpha$. (При этом ω_i не определено, если α встречается среди $a_1, \ldots, a_i$.) Это отображение определено почти всюду по равномерной мере. Образ равномерной меры — вычислимая мера на Ω, а образ перечислимого снизу случайного числа — случайная по этой мере последовательность, являющаяся характеристической функцией перечислимого неразрешимого множества. (Для неразрешимости важна плотность a_i в $[0, 1]$.) См. разделы 5.7 и 5.9.3.]

Мы построили несколько примерно последовательностей, не случайных ни по какой вычислимой мере. Можно усложнить задачу и спросить, существуют ли тьюринговы степени, не содержащие случайных по вычислимой мере последовательностей. (Здесь уже всё равно, какую вычислимую меру рассматривать, поскольку тьюринговы степени одни и те же, так что можно говорить о равномерной мере без ограничения общности.) Оказывается, что и такое бывает:

183 Докажите, что существует невычислимый оракул A с таким свойством: никакая вычислимая относительно A последовательность не случайна по равномерной мере.

[Указание. Этот оракул надо строить диагональным методом, по частям. Во-первых, можно всегда добавить кусок, который гарантирует, что он не вычисляется i-й машиной Тьюринга. С другой стороны, можно добавить кусок, который гарантирует, что i-я машина Тьюринга с этим оракулом либо вычисляет не всюду определённую последовательность, либо вычисляет последовательность, в которой есть начало большого дефекта. В самом деле, если свободную часть оракула можно подобрать так, чтобы машина вычислила очень длинную последовательность, то сделаем это первым встретившимся способом и получим простое длинное начало. Если же так подобрать нельзя, то машина заведомо не вычислит всюду определённой последовательности.]

Утверждение последней задачи следует из двух интересных (но трудных) результатов, выходящих за рамки нашей книги. Первый из них доказал В. Вьюгин [186]: можно построить вероятностную машину, которая с положительной вероятностью выдаёт последовательности, обладающие указанным в задаче свойством.

Это кажется парадоксальным: свойство состоит в том, что к последовательности нельзя подобрать меру, которая её «объясняет» (относительно которой она случайна). С другой стороны, такая последовательность с положительной вероятностью получается в ходе алгоритмического процесса — не есть ли это то самое объяснение, которого мы ищем? Дело тут в том, что этот алгоритмический процесс с положительной вероятностью может дать конечную последовательность, так что не задаёт никакой вычислимой меры.

Во-вторых, утверждение этой задачи можно вывести из недавних (но уже знаменитых) результатов о «низких множествах», см. книги Ниса [127] и Доуни – Хиршфельдта [37]: существует перечислимое неразрешимое множество A, добавление которого в качестве оракула не меняет класса случайных последовательностей (а также не меняет префиксной сложности, хотя сейчас это нам не важно). Соответственно никакая A-вычислимая последовательность не может быть случайной, поскольку она очевидно не является A-случайной. (Отметим, что здесь мы получаем множество A с дополнительным условием перечислимости.)

Таким образом, мы видим, что по самым разным причинам бывают последовательности, к которым не сводятся по Тьюрингу никакие случайные. Но в обратную сторону ситуация другая: всякая последовательность сводится по Тьюрингу к случайной по равномерной мере последовательности (см. ниже теорему 126, с. 214). Доказательство этой теоремы показывает также, что всякая тьюрингова степень выше $\mathbf{0}'$ (к которой сводится проблема остановки) содержит случайную последовательность, см. задачу 190 на с. 216.

$\boxed{\mathbf{184}}$ Покажите, что существует последовательность, эквивалентная (по Тьюрингу) случайной по равномерной мере последовательности, но не случайная ни по какой вычислимой мере. [Указание. На чётных местах поместим генерическую последовательность, а на нечётных — ту случайную по равномерной мере, к которой она сводится. Надо использовать ещё тот факт, что если последовательность случайна по мере P, то подпоследовательность её членов с чётными номерами случайна относительно проекции P на эти координаты, см. теорему 123.]

Последовательности, не случайные ни по какой вычислимой мере, в некотором смысле аналогичны нестохастическим по Колмогорову конечным объектам (см. раздел 14.2). Более того, несложно показать, что если последовательность случайна по некоторой вычислимой мере, то её начальные отрезки являются стохастическими (задача 348, с. 475).

5.9.3. Случайность по образу меры

Вернёмся к тому, с чего мы начали эту главу: рассмотрим вероятностную машину, которая состоит из датчика случайных («честных») битов и алгоритмического преобразования этих битов в выходную последовательность (конечную или бесконечную).

Предположим, что с вероятностью 1 получается бесконечная последовательность, и тем самым распределение выходных последовательностей есть некая вычислимая мера μ.

Зададим себе такой (слегка философский) вопрос: про какие последовательности мы готовы поверить, что они появились на выходе такого устройства? Тут есть два ответа.

Во-первых, у нас как раз для этого есть определение случайности по Мартин-Лёфу относительно меры μ, и можно объявить «правдоподобными» случайные в этом смысле последовательности.

Во-вторых, мы можем посмотреть вовнутрь машины и спросить: а в какие последовательности мы готовы поверить на выходе датчика случайных битов? естественно считать, что в МЛ-случайные по равномерной мере. Соответственно, на выходе мы ожидаем их образы.

Оказывается, наша философская совесть чиста и эти два ответа на самом деле совпадают. Вот точные формулировки.

Пусть μ — вычислимое распределение вероятностей на Ω, а $f\colon \Sigma \to \Sigma$ — непрерывное вычислимое отображение. Рассмотрим образ меры μ при этом отображении, то есть меру ν на пространстве Σ, для которой

$$\nu(U) = \mu(f^{-1}(U))$$

при $U \subset \Sigma$. Другими словами, ν — это распределение случайной величины $f(\omega)$, если ω — случайная величина с распределением μ. Вообще говоря, ν не обязана быть сосредоточена на бесконечных последовательностях, то есть в нашей терминологии может быть не мерой, а полумерой. Предположим, однако, что это не так и что ν является мерой. (В этом случае, как мы видели, ν задаёт вычислимую меру на Ω.)

Теорема 123. (а) *Если ω — случайная по мере μ последовательность, то её образ $f(\omega)$ будет бесконечной последовательностью, случайной по мере ν.*

(б) *Всякая случайная по мере ν последовательность может быть получена таким образом (равна $f(\omega)$ для некоторой случайной по мере μ последовательности ω).*

(Говоря о случайности в этой теореме, мы имеем в виду случайность в смысле Мартин-Лёфа.)

◄ Прежде всего докажем, что образ μ-случайной последовательности не может быть конечным словом. В самом деле, рассмотрим бесконечные последовательности, входящие в прообраз этого конечного слова z, то есть в прообраз $\Sigma_z \setminus (\Sigma_{z0} \cup \Sigma_{z1})$. Прообраз Σ_z есть эффективно открытое множество (объединение перечислимого семейства интервалов), прообраз $\Sigma_{z0} \cup \Sigma_{z1}$ — также эффективно открытое множество, являющееся подмножеством первого. Нам надо доказать, что прообраз разности (то есть разность прообразов) не содержит случайных последовательностей, то есть является эффективно нулевым относительно меры μ множеством. Это следует из такой общей леммы:

Лемма 1. *Пусть μ — вычислимая мера на Ω, а $U \subset V$ — два эффективно открытых множества, причём $\mu(V \setminus U) = 0$. Тогда $V \setminus U$ является*

эффективно нулевым множеством (не содержит случайных последовательностей).

Доказательство. Очевидно, достаточно рассмотреть один интервал I множества V (ограничив U этим интервалом). По мере перечисления интервалов множества U обнаруживается всё бóльшая часть интервала I, входящая в U. При этом мера обнаруженной части стремится к мере интервала I (непрерывность меры), и рано или поздно мы дождёмся момента, когда разность по мере будет меньше ε, какое бы $\varepsilon > 0$ нам бы ни задали. Лемма 1 доказана. (Заметим в скобках, что это рассуждение годится для любого открытого V; эффективность мы не использовали.)

Чтобы завершить доказательство пункта (а), осталось показать, что образ μ-случайной последовательности ω не может быть ν-неслучайным, то есть содержаться в эффективно ν-нулевом множестве. Но если бы $f(\omega)$ можно было покрыть интервалами сколь угодно малой меры, то беря f-прообразы этих интервалов, мы бы получили покрытие последовательности ω сколь угодно малой меры (прообраз эффективно открытого множества открыт и по определению имеет ту же меру). Утверждение (а) доказано.

$\boxed{\mathbf{185}}$ Докажите количественный вариант этого утверждения: ограниченный в среднем дефект случайности последовательности $f(\omega)$ по мере ν превосходит ограниченный в среднем дефект случайности последовательности ω по мере μ не более чем на константу, зависящую от мер и от отображения, но не от последовательности ω. (В этой задаче, говоря о дефектах случайности, мы имеем в виду определение раздела 3.5 для бесконечных последовательностей.)

Докажем теперь утверждение (б), используя понятие дефекта (для конечных последовательностей, в смысле определения на с. 202). Пусть последовательность τ случайна по мере ν. Это значит, что дефекты её начальных отрезков относительно этой меры ограничены. Теперь применим следующую лемму, которую можно рассматривать как аналог утверждения (б) для конечных последовательностей:

Лемма 2. *Пусть u — произвольное двоичное слово, для которого $\nu(\Omega_u) > 0$. Тогда существует двоичное слово w, для которого $u \preccurlyeq f(w)$ (слово u является началом $f(w)$) и $d_\mu(w) \leqslant d_\nu(u) + O(1)$.*

(Имеется в виду, что константа в $O(1)$ не зависит от слова u, но может зависеть от выбора f, μ и ν; через d_μ и d_ν обозначены соответствующие дефекты.)

Доказательство. Рассмотрим прообраз множества Σ_u при отображении f. Это эффективно открытое множество в Σ, которое мы обозначим через F_u. По построению μ-мера множества F_u (сосредоточенная на бесконечных последовательностях) равна $\nu(\Omega_u)$. Если дефект $d_\nu(u)$ мал, то эта мера не может быть сильно меньше априорной вероятности слова u.

Рассмотрим теперь априорную вероятность множества F_u, то есть вероятность того, что универсальная вероятностная машина M на выходе даст элемент из F_u, или, другими словами, вероятность того, что машина $f \circ M$ (к выходу M дополнительно применяем f) даст некоторое продолжение слова u. Сравнивая машину $f \circ M$ с универсальной, заключаем, что априорная вероятность множества F_u лишь в константу раз превышает априорную вероятность слова u, которая не более чем

в $2^{d_\nu(u)}$ раз превышает $\nu(u)$, которое равно μ-мере множества F_u. Таким образом мы получаем неравенство для двух мер множества F_u (априорной вероятности, которую мы обозначим через A, и меры μ):

$$\frac{A(F_u)}{\mu(F_u)} \leqslant O(2^{d_\nu(u)}).$$

Множество F_u представимо в виде объединения (пусть неперечислимого) непересекающихся интервалов, и ясно, что для одного из интервалов Σ_w выполнено аналогичное неравенство:

$$\frac{A(\Sigma_w)}{\mu(\Sigma_w)} \leqslant O(2^{d_\nu(u)}).$$

Поскольку $\Sigma_w \subset F_u$, то $f(w) \succcurlyeq u$, а неравенство даёт оценку $d_\mu(w) \leqslant d_\nu(u) + O(1)$, что и требовалось. Лемма 2 доказана.

Продолжая доказательство утверждения (б) теоремы, применим доказанную лемму к начальным отрезкам $t_0, t_1, t_2, \ldots$ случайной по мере ν последовательности τ (имеющим длины $0, 1, 2, \ldots$). Их ν-дефекты ограничены, и по лемме можно найти последовательность $w_0, w_1, \ldots$ слов с ограниченными μ-дефектами, для которых $f(w_i)$ является продолжением t_i. Если бы w_i ещё продолжали друг друга, то мы достигли бы желаемого. Но это ниоткуда не следует. Тем не менее, применяя обычное рассуждение с компактностью, можно из последовательности w_i выделить либо бесконечную подпоследовательность, состоящую из равных слов, либо бесконечную подпоследовательность, сходящуюся к некоторой бесконечной последовательности ω (последнее означает, что любой начальный отрезок ω является начальным отрезком всех членов подпоследовательности, начиная с некоторого).

В первом случае последовательность τ вычислима, будучи образом конечного слова w, встречающегося среди w_i бесконечно много раз. Это не противоречит случайности, если эта вычислимая последовательность имеет (как одноэлементное множество) положительную ν-меру, но в этом случае в качестве ω можно взять любое бесконечное μ-случайное продолжение слова w (оно существует, так как μ-дефект слова w конечен и $\mu(\Omega_w) > 0$).

Осталось рассмотреть второй случай, когда бесконечная подпоследовательность последовательности w_i сходится к некоторой последовательности ω.

Здесь нам понадобится некоторая подготовка, так что мы сделаем отступление и докажем, что дефект почти монотонен. Вспомним критерий случайности Левина – Шнорра (теоремы 91 и 93). Мы видели, что для случайных в смысле Мартин-Лёфа последовательностей дефект их начальных отрезков ограничен, а для неслучайных — стремится к бесконечности. В частности, не бывает последовательностей, у которых определённый таким образом дефект неограничен, но не стремится к бесконечности. Почему так происходит? На этот вопрос даёт ответ следующая теорема.

Теорема 124. *Пусть фиксирована вычислимая мера P. Тогда существует такая константа c, что для любого слова x и для любого его продолжения y выполняется неравенство*:

$$d_P(y) \geqslant d_P(x) - 2\log d_P(x) - c.$$

Смысл этой теоремы: любое продолжение конечной последовательности с большим дефектом имеет (почти столь же) большой дефект. Или: любое начало (конечной) последовательности с малым дефектом имеет (почти столь же) малый дефект.

◄ Для любого k рассмотрим перечислимое множество всех конечных последовательностей, у которых дефект больше k. Все их бесконечные продолжения образуют открытое множество S_k, у которого P-мера не больше 2^{-k}. Рассмотрим теперь меру P_k на Ω, которая равна нулю вне S_k и в 2^k раз превосходит P внутри S_k. (Формально говоря, мера множества U равна $2^k P(U \cap S_k)$). Отметим, что P_k не является вычислимой мерой в нашем смысле слова, поскольку мера всего пространства не равна 1 (и даже не обязана быть вычислимой, а всего лишь перечислима снизу). Но P_k вполне можно считать перечислимой снизу полумерой (оставшуюся до 1 вероятность считаем мерой пустого слова).

Рассмотрим теперь сумму

$$\sum_k \frac{1}{2k^2} P_k,$$

которая задаёт некоторую перечислимую снизу полумеру S (двойка в знаменателе добавлена, чтобы сумма ряда $\sum 1/(2k^2)$ была меньше единицы; меру пустого слова надо вновь увеличить). При этом

$$-\log S(x) \leqslant -\log P(x) - k + 2\log k + O(1)$$

для всех слов x, являющихся продолжениями какого-либо слова с дефектом больше k. Поскольку S не больше априорной вероятности на дереве (с точностью до $O(1)$-множителя), получаем требуемое утверждение.

В этом рассуждении мы предполагали, что дефект x конечен, то есть что $P(\Omega_x) \neq 0$. Но если это не так, то для любого продолжения y мера $P(\Omega_y)$ также нулевая, и дефект бесконечен. ►

Вернёмся к доказательству теоремы 123. Мы остановились на последовательности слов (подпоследовательности последовательности w_i), которая сходится к некоторой точке $\omega \in \Omega$. Все w_i имеют малые дефекты. Заметим, что в этом случае:

(1) любой начальный отрезок ω является начальным отрезком одного из слов w_i, а их μ-дефекты ограничены, поэтому по теореме 124 и μ-дефекты начальных отрезков ω ограничены, и потому ω является μ-случайной;

(2) по ранее доказанному в пункте (а) последовательность $f(\omega)$ бесконечна;

(3) последовательность $f(\omega)$ не может иметь начальных отрезков, не являющихся начальными отрезками τ, поскольку в этом случае у ω был бы начальный отрезок, образ которого несовместим с τ, и этот отрезок был бы (в силу сходимости) начальным отрезком сколь угодно далёких w_i, образы которых имеют сколь угодно длинное общее начало с τ.

Полученное противоречие завершает доказательство пункта (б) теоремы. ►

Теорема 123 имеет и другие доказательства. Следующее рассуждение было предложено Ан. А. Мучником:

186 Дайте прямое доказательство утверждения (б), используя лишь определение эффективно нулевого множества.

[Указание. Для данного ε рассмотрим покрытие Z_ε наибольшего эффективно μ-нулевого множества интервалами суммарной μ-меры меньше ε; пусть F — замкнутое множество бесконечных последовательностей, оказавшихся непокрытыми. Все последовательности из F случайны, поэтому функция f всюду определена на F (значения — бесконечные последовательности). Образ F будет замкнутым множеством, и его мера больше $1-\varepsilon$ (прообраз содержит F); дополнение является открытым множеством, имеет меру не больше ε и покрывает все точки, не имеющие прообразов в F. Единственная оставшаяся проблема — надо эффективизировать доказательство теоремы о компактности образа компактного множества при непрерывном отображении и заключить, что дополнение к образу F не просто открыто, но и эффективно открыто равномерно по ε.]

Аналогичное рассуждение позволяет получить и количественную версию утверждения (б), показывающую, что оценка задачи 185 точна: ограниченный в среднем ν-дефект последовательности ω совпадает с точностью до $O(1)$-слагаемого с точной нижней гранью ограниченных в среднем μ-дефектов всех f-прообразов этой последовательности.

187 Докажите утверждение, которое можно рассматривать как конечный аналог пункта (а) теоремы 123: если u и w — два двоичных слова, причём $u \preccurlyeq f(w)$, то

$$d_\nu(u) \leqslant d_\mu(w) + 2\log d_\mu(w) + O(1).$$

[Указание. Последовательности с большим ν-дефектом покрываются множеством малой ν-меры, поэтому их прообразы покрываются множеством малой μ-меры и потому имеют большой μ-дефект. Отметим, что это утверждение является обобщением теоремы 124.]

Теорема 123 имеет довольно неожиданные следствия. Приведём несколько примеров.

188 Пусть последовательность ω — случайна в смысле Мартин-Лёфа относительно бернуллиевой меры (все испытания независимы) с вероятностью единицы $1/3$. Докажите, что существует последовательность ω', случайная относительно равномерной меры, которая получается из ω заменой некоторых нулей на единицы. [Указание. Рассмотрим случайную последовательность независимых случайных точек, равномерно распределённых на $[0,1]$, точнее, последовательность случайных битов, записанных в двумерную таблицу, строки которой — бесконечные двоичные дроби. Заменим точки, большие $2/3$, на единицы, а остальные на нули. Доказанная теорема утверждает, что получится последовательность, случайная по бернуллиевой мере с вероятностью единицы $1/3$, и что всякая случайная по этой мере последовательность может быть получена таким образом. Аналогично для границы $1/2$, откуда и следует утверждение задачи.]

Другое следствие теоремы 123 и его обобщения мы сейчас обсудим.

5.9.4. Теорема Ламбальгена

Рассмотрим вероятностную машину, которая получает с помощью датчика случайных битов последовательность $\omega_0\omega_1\omega_2\ldots$, и выдаёт на выход последователь-

ность $\omega_0\omega_2\omega_4\ldots$ (пропуская каждый второй бит). Ясно, что распределение вероятностей на выходе такой машины будет равномерным, и теорема 123 даёт такие два утверждения:

(а) если $\omega_0\omega_1\omega_2\ldots$ — случайная по равномерной мере последовательность, то и последовательность $\omega_0\omega_2\omega_4\ldots$ случайна по той же мере;

(б) для всякой случайной по равномерной мере последовательности $\omega_0\omega_2\omega_4\ldots$ можно найти такую последовательность $\omega_1\omega_3\omega_5\ldots$, что их смесь $\omega_0\omega_1\omega_2\ldots$ будет случайной.

Первое несложно усмотреть непосредственно, а второе совсем не так очевидно. Это утверждение можно переформулировать для пар последовательностей: если α — случайная по равномерной мере последовательность, то можно найти последовательность β, для которой пара $\langle\alpha,\beta\rangle$ будет случайной (в естественном смысле, относительно независимых равномерных мер на координатах).

Возникает вопрос: хорошо, такая β существует, но какими она должна обладать свойствами, чтобы пара $\langle\alpha,\beta\rangle$ была случайной? Ясно, что β сама должна быть случайной (по уже доказанному), но этого мало. Скажем, нельзя положить $\beta=\alpha$, поскольку пара $\langle\alpha,\alpha\rangle$ не случайна. (Ей соответствует последовательность $\omega_0\omega_1\ldots$, в которой каждый бит повторяется дважды.)

Ответ на этот вопрос такой: последовательность β должна быть случайной по равномерной мере, даже если разрешить использовать (в определении случайности) α в качестве оракула. Это утверждает следующая теорема Ламбальгена [73]. Пусть P и Q — два вычислимых распределения на пространстве Ω. Рассмотрим произведение этих распределений; получится вычислимое распределение на $\Omega\times\Omega$ (которое изоморфно Ω, и определения случайности на него очевидно переносятся).

Теорема 125. *Пара последовательностей $\langle\xi,\eta\rangle$ случайна по Мартин-Лёфу относительно распределения $P\times Q$ тогда и только тогда, когда выполнены два условия*:

(1) *последовательность ξ случайна в смысле Мартин-Лёфа по мере P*;

(2) *последовательность η случайна в смысле Мартин-Лёфа с оракулом ξ по мере Q.*

Говоря о случайности с оракулом, мы имеем в виду, что алгоритм, перечисляющий покрытие, теперь имеет ξ в качестве оракула (от этого перечислимых множеств и неслучайных последовательностей может стать больше, а случайных — меньше).

Можно ещё отметить, что ξ и η входят симметрично, поэтому условие (1) можно было бы усилить, разрешив η в качестве оракула. Но несимметричный вариант кажется более естественным, его можно прочесть так: «выбрать пару случайно означает случайно выбрать первый член пары, а затем, зная первый член, случайно выбрать второй».

◄ Докажем сначала, что для случайной пары выполнены условия (1) и (2).

(1) Если последовательность ξ не случайна и покрывается интервалами малой P-меры, то эти интервалы (умноженные на всё Ω по второй координате) дают покрытие пары $\langle\xi,\eta\rangle$ той же меры (относительно $P\times Q$). Можно также сослаться на теорему 123.

(2) Пусть для всякого ε можно перечислять (с оракулом ξ) семейство интервалов малой Q-меры, покрывающее η. Этот же процесс можно запустить и с любым другим оракулом, и он будет порождать какие-то интервалы, используя конечную информацию относительно оракула.

Таким образом мы получим (по данному ε) некоторое перечислимое (без оракула) семейство прямоугольников с таким свойством: Q-мера сечения, в котором первая координата равна ξ, не больше ε. Такое семейство легко переделать в семейство прямоугольников общей меры не больше ε, если принудительно урезать прямоугольники так, чтобы в любом сечении мера была не больше ε и при этом сечения, где это неравенство и так выполнялось, не менялись. Получаем противоречие со случайностью пары $\langle\xi,\eta\rangle$.

Теперь докажем, что если пара $\langle\xi,\eta\rangle$ не случайна, то не выполнено одно из условий (1) и (2). Пусть имеется семейство прямоугольников с общей мерой не больше ε, покрывающее $\langle\xi,\eta\rangle$. Пусть U — объединение этих прямоугольников. Посмотрим, при каких значениях первой координаты x) сечение $U_x = \{y \mid \langle x,y\rangle \in U\}$ имеет меру больше $\sqrt{\varepsilon}$. Соответствующее подмножество (одномерное, по первой координате) имеет меру не больше $\sqrt{\varepsilon}$ и его можно представить в виде перечислимого объединения интервалов. При этом для точки $\langle\xi,\eta\rangle$, покрытой исходным семейством прямоугольников, верно одно из двух: либо ξ покрыта построенным семейством интервалов общей меры не более $\sqrt{\varepsilon}$, либо $\langle\xi,\eta\rangle$ покрыта семейством прямоугольников, которые в ξ-сечении имеют Q-меру не больше $\sqrt{\varepsilon}$. Во втором случае η покрыта ξ-перечислимым семейством интервалов суммарной Q-меры не более $\sqrt{\varepsilon}$.

Хотелось бы сделать так для каждого ε и заключить, что либо ξ не случайна, либо η не случайна с оракулом ξ. Первое гарантировано, если при всех ε имеет место первый вариант (ξ покрыта интервалами); второе — если при всех ε случается второе. Но что делать, если при некоторых ε случается одно, а при других — другое?

Тут помогает такой трюк: выполним указанное построение для $\varepsilon = 2^{-2k}$ при всех k. Тогда для каждого k получится семейство $V(k)$ интервалов по первой координате, имеющих общую P-меру не более $2^{-k} = \sqrt{2^{-2k}}$. Теперь есть две возможности: либо при бесконечно многих k это семейство покрывает ξ, либо для всех достаточно больших k это семейство не покрывает ξ.

В первом случае для каждого K объединим все семейства $V(k)$ при $k \geqslant K$; получится семейство интервалов с мерой не более $2 \cdot 2^{-K}$, и оно покрывает ξ уже при любом K, так что ξ не случайна.

Во втором случае есть способ (для всех k, начиная с некоторого, пусть нам и неизвестного) получить ξ-перечислимое покрытие для η, имеющее малую меру, и потому η не случайна с оракулом ξ. ▶

Эта теорема также имеет количественный вариант (см. [183]): можно доказать, что ограниченный в среднем дефект случайности пары $\langle\xi,\eta\rangle$ относительно меры $P \times Q$ равен сумме дефекта ξ относительно P и дефекта η относительно Q с оракулом ξ и дополнительным условием, равным первому дефекту (округлённому до ближайшего целого числа). Правда, чтобы придать точный смысл этому утвержде-

нию, нужно определить понятие дефекта с оракулом и условием (как функцию последовательности, оракула и условия), но это можно сделать и получить таким образом формулу, аналогичную формулу для префиксной сложности пары. (Интересно, что прямого способа вывести это из формулы для префиксной сложности пары, используя выражение для дефекта через префиксную сложность, не видно.)

Естественно попытаться обобщить эту теорему на случай зависимых величин и доказать что-нибудь в таком роде: пара $\langle \xi, \eta \rangle$ случайна относительно вычислимого распределения на $\Omega \times \Omega$, если ξ случайна относительно проекции этого распределения на первую координату (эту проекцию называют *маргинальным* распределением), а η случайна относительно условного распределения (при данном значении первой координаты) с оракулом ξ. Однако это не так просто. Во-первых, надо определить условное распределение (что можно сделать, если ξ случайно, как доказал Хайято Такахаши); во-вторых, условная вероятность может не быть вычислимой, и надо как-то определять случайность. Эти вопросы разбираются в его статьях [164, 165], но критерия совсем без дополнительных ограничений пока получить не удалось.

5.9.5. Теорема Кучеры – Гача

Вернёмся к образу множества случайных последовательностей при вычислимых отображениях. Вопрос о том, какие последовательности могут (не вызывая отторжения гипотезы о случайности) появиться на выходе вероятностной машины, имеет смысл и в общем случае, без предположения о том, что вероятность появления конечной последовательности равна нулю. Пусть, как и прежде, имеется вычислимое распределение вероятностей μ на множестве Ω, а также вычислимое непрерывное отображение $f\colon \Sigma \to \Sigma$. Тогда возникает распределение вероятностей на Σ, являющееся образом μ при отображении f. Мы, однако, не предполагаем, что оно сосредоточено на бесконечных последовательностях, так что получается некоторая перечислимая снизу полумера ν, вообще говоря, не являющаяся мерой.

Мы можем теперь рассмотреть образы μ-случайных последовательностей при отображении f. Как связано это множество с полумерой ν? Можно ли утверждать, что оно определяется полумерой ν (как это было для случая мер, когда оно состояло из ν-случайных последовательностей)? Может быть, верен аналог теоремы Левина – Шнорра и образы μ-случайных последовательностей — это те и только те последовательности, у начальных отрезков которых отношение (непрерывной) априорной вероятности к ν ограничено?

К сожалению, рассуждения, использованные для случая мер, не работают, когда ν является лишь полумерой. (Проблемы возникают в обе стороны.) Более того, как недавно обнаружили Л. Бьенвеню и К. Портер, образ множества μ-случайных последовательностей вообще не определяется полумерой ν:

$\boxed{\mathbf{189}}$ Покажите, что существуют вычислимые отображения $f_1, f_2\colon \Sigma \to \Sigma$, которые порождают одну и ту же выходную полумеру (как образ равномерной меры на Ω), но для которых образы $f_1(R)$ и $f_2(R)$ множества R случайных по Мартин-Лёфу последовательностей (относительно равномерной меры) различны.

[Указание. Будем рассматривать машины f_1 и f_2, которые печатают на выходе только нули (конечное или бесконечное число нулей). Такая машина, ограниченная на Ω, описывается убывающей последовательностью равномерно эффективно открытых множеств $A_1 \supset A_2 \supset \ldots$, где A_i состоит из тех входов, для которых на выходе появляется i или больше нулей. Выходное распределение машины при этом задаётся мерами множеств A_i. Остаётся построить две последовательности вложенных множеств, у которых меры одинаковы, но у одной пересечение содержит случайную точку, а у второй не содержит. Это делается так: рассмотрим перечислимое снизу случайное число ω и вычислимую возрастающую последовательность рациональных чисел r_i с пределом ω, и возьмём вложенные интервалы $(r_i, \omega + 1/i)$. Их пересечение содержит ω, но если взять интервалы тех же длин с левой границей в нуле, то пересечение будет пусто (а если взять с серединой в $1/2$, пересечение будет состоять только из неслучайного числа $1/2$).]

Тем не менее следующий факт (который бы следовал из сформулированного выше «критерия», будь он верным: достаточно было бы взять в качестве f универсальную вероятностную машину, выходное распределение которой совпадает с непрерывной априорной вероятностью!) всё же можно доказать (это сделали Кучера [69] и Гач (см. [46]):

Теорема 126. *Пусть α — произвольная последовательность нулей и единиц. Тогда существует случайная (в смысле Мартин-Лёфа) по равномерной мере последовательность ω, а также вычислимое отображение $f\colon \Sigma \to \Sigma$, для которого $f(\omega) = \alpha$.*

Для знакомых с теорией вычислимых функций можно было бы сказать короче: всякая последовательность нулей и единиц сводится по Тьюрингу к некоторой последовательности, случайной по равномерной мере (мы уже упоминали этот результат на с. 205).

Применить для доказательства тот же приём (взять прообразы малого дефекта и их предельную точку) не удаётся, поскольку образ этой предельной точки теперь может оказаться конечным. Поэтому придётся воспользоваться другой конструкцией.

◄ Будем доказывать несколько более сильное утверждение и построим вычислимое непрерывное отображение f (одно для всех α), для которого образ $f(R)$ множества R случайных относительно равномерной меры последовательностей покрывает всё Ω.

Более того, мы для любого эффективно открытого множества U (объединения перечислимого семейства интервалов) с достаточно малой мерой построим вычислимое отображение f, для которого образ $f(\Omega \setminus U)$ покрывает всё Ω. Поскольку в качестве U можно взять покрытие максимального эффективно нулевого множества, отсюда следует требуемое утверждение.

Вот идея конструкции. Прежде всего, мы будем представлять себе область определения отображения f (канторовское пространство) не как двоичное дерево, а как дерево с ветвлением $2^{k_0}, 2^{k_1}, \ldots$ (где целые положительные числа k_i достаточно быстро растут, см. далее). Мы вкладываем в него двоичное дерево (при этом остаётся много свободного места), и отображаем это вложенное дерево на полное двоичное

дерево (естественным образом). Другими словами, среди сыновей корня в новом дереве (слов длины k_0) мы выбираем какие-то два слова s_0 и s_1, затем у каждого их них выбираем по два продолжения длины $k_0 + k_1$ (слова s_{00} и s_{01} у s_0 и слова s_{10} и s_{11} у s_1), и так далее; f отображает s_v в v.

После этого мы смотрим, какие интервалы появляются в множестве U. Если они не включают в себя точки, соответствующие путям двоичного поддерева, то они никому не мешают: образ дополнения к U по-прежнему содержит все двоичные последовательности. Если же какой-то путь попадёт в U, то есть один из интервалов, задающих U, попадёт в вершину двоичного поддерева, то эту вершину нужно заменить на запасную, найдя подходящего брата, и заново вложить соответствующий кусок двоичного поддерева, продолжив и отображение f. Это можно делать, пока испорченных вершин не станет слишком много: у корня надо испортить почти всех сыновей (=всех, кроме одного); сын корня будет испорчен, если у него самого почти все сыновья испорчены и так далее, пока мы не дойдём до первоисточников порчи, то есть интервалов множества U. Этих интервалов наберётся не менее чем на

$$\left(\frac{2^{k_0}-1}{2^{k_0}}\right)\cdot\left(\frac{2^{k_1}-1}{2^{k_1}}\right)\cdot\left(\frac{2^{k_2}-1}{2^{k_2}}\right)\cdot\ldots, \tag{$*$}$$

так что проблем с запасными вершинами не будет, если размер U меньше этой границы. Остаётся выбрать числа $k_0, k_1, \ldots$ достаточно большими и быстро растущими, чтобы это произведение было строго положительным (или даже близким к единице).

Изложим это рассуждение более подробно. Выбрав (вычислимую) последовательность положительных целых чисел $k_0, k_1, k_2, \ldots$, мы рассматриваем в качестве вершин нового дерева слова длины k_0 (получается ветвление 2^{k_0}), затем слова длины $k_0 + k_1$ (так что на следующем уровне ветвление 2^{k_1}) и так далее. Слова этих длин становятся вершинами некоторого дерева T; мы будем их называть T-вершинами, если нужно подчеркнуть различие с полным двоичным деревом.

Сначала мы выбираем (произвольным вычислимым образом) некоторое двоичное поддерево в T, и отображаем с помощью f его вершины в вершины полного двоичного дерева. Затем мы перечисляем интервалы эффективно открытого множества U. Можно считать без ограничения общности, что эти интервалы соответствуют T-вершинам (имеют нужную длину). Когда появляется новый интервал, мы делаем следующее:

- Объявляем соответствующую T-вершину плохой.
- Распространяем плохие T-вершины к корню: если у T-вершины все или почти все (=все, кроме одного) сыновья плохие, то её тоже считаем плохой. В итоге может добавиться некоторая цепочка плохих T-вершин в направлении корня.
- Если плохие T-вершины пересеклись с текущим двоичным поддеревом, то корректируем его: берём ближайшую к корню плохую T-вершину, и заменяем её в поддереве на хорошего брата. (Отец ближе к корню, поэтому хороший, поэтому есть два хороших сына — так что найдётся хороший брат, не входящий в поддерево.) Затем из заменённой T-вершины выращиваем поддерево

из хороших T-вершин (и снова это возможно, так как у хорошей T-вершины всегда есть два хороших сына в T).

- Доопределяем отображение f на вновь появившейся части двоичного поддерева.

Единственный случай, когда этого сделать не удастся — когда корень дерева станет плохой вершиной. В этом случае все его T-сыновья (кроме, возможно, одного) будут плохими, каждый из них либо окажется интервалом из U, либо будет иметь всех (кроме, возможно, одного) плохих сыновей, и так далее. В итоге получится дерево из плохих вершин, имеющее ветвление не меньше $2^{k_0} - 1$ (в корне), $2^{k_1} - 1$ (на следующем уровне) и так далее, а листья соответствуют интервалам U, так что размер U не меньше произведения $(*)$. Получаем положительную нижнюю оценку на размер U, если это произведение сходится — а оно сходится в том и только том случае, когда ряд $\sum 2^{-k_i}$ сходится. Поэтому, взяв (скажем) $k_i = \lceil 2\log i \rceil$ (при $i \geqslant 2$) и достаточно малое множество U, мы гарантируем, что корень никогда не станет плохим.

Почему эта конструкция даёт нужный результат? Надо заметить, что

- множество плохих T-вершин может только увеличиваться;
- на каждом шаге конструкции двоичное поддерево не содержит ни одной плохой T-вершины;
- если T-вершина была исключена из двоичного поддерева, то эта T-вершина никогда вновь не попадёт в двоичное поддерево (и переопределять f на ней не придётся).

Все эти свойства ясны по построению. (Последнее: если T-вершина была исключена из двоичного поддерева, то кто-то из её предков был в этот момент плохим, каковым он и остался, так что вновь в двоичное поддерево она попасть не сможет.)

Остаётся доказать, что при таком построении f для любой последовательности $\alpha \in \Omega$ найдётся f-прообраз, не входящий в выброшенное открытое множество U. По построению в каждый момент t найдётся прообраз ω_t, не входящий в уже построенную часть U. Кроме того, с ростом t последовательности ω_t сходятся к некоторому пределу ω (индукцией по уровню доказываем, что на этом уровне будет стабилизация, так как число возможных изменений ограничено). Надо проверить, что ω не принадлежит U и что $f(\omega) = \alpha$.

Если ω принадлежит U, то ω попадает в некоторый интервал U, который рано или поздно обнаружится. После этого момента ω_t не могут лежать в U, что противоречит сходимости. Покажем, что $f(\omega) = \alpha$. Пусть z — произвольное конечное начало α; покажем, что $f(\omega)$ начинается на z. Пусть k — длина z. На каждом шаге имеется некоторая ветвь в дереве (большого ветвления), состоящая из k блоков и отображающаяся в z. Эта ветвь стабилизируется с ростом t, и после стабилизации в f добавляется часть, гарантирующая, что $f(\omega)$ начинается на z. ▶

$\boxed{\mathbf{190}}$ Покажите, что построенная в доказательстве теоремы случайная последовательность будет вычислимой с двумя оракулами α и $\mathbf{0}'$ (проблема остановки). [Указание: предельное положение вложенного дерева вычислимо относительно $\mathbf{0}'$.]

$\boxed{\textbf{191}}$ Выведите из доказательства теоремы, что для всякой последовательности α найдётся случайная последовательность ω, к которой она сводится по Тьюрингу, причём для получения n битов последовательности α используется не более $n + o(n)$ первых битов последовательности ω. [Указание. Вместо двоичного дерева, которое мы вкладываем, можно рассматривать дерево ветвления 2^{m_i}; тогда нужна сходимость произведения $(1 - 2^{m_i}/2^{k_i})$, то есть ряда $\sum 2^{m_i - k_i}$. Можно положить, например, $m_i = i$ и $k_i = i + 2\log i$.]

В порядке философических спекуляций можно объяснить смысл утверждения теоремы Кучеры – Гача так. Для любой последовательности можно представить объяснение, почему её появление не удивительно: про случайные последовательности это как бы подразумевается определением случайности, и остаётся пристроить к датчику случайных чисел машину, реализующую вычислимое отображение f.

Утверждение теоремы можно интерпретировать как парадокс: с одной стороны, если некоторая последовательность α с положительной вероятностью вычислима относительно случайно взятой последовательности (в том смысле, что множество тех последовательностей, к которым α сводится, имеет положительную меру), то α вычислима. В самом деле, множество всех возможных оракулов разбивается на счётное число частей (соответствующих возможным машинам с оракулом), поэтому одна из частей имеет положительную меру. Поэтому априорная вероятность α положительна и α вычислима. С другой стороны, для каждой последовательности α есть случайная по Мартин-Лёфу последовательность ω, относительно которой α вычислима. Противоречия тут нет, потому что нулевое множество тех последовательностей, относительно которых α вычислима, не обязано быть эффективно нулевым.

6. Общая классификация сложностей

6.1. Сложность разрешения

Начав с простой колмогоровской сложности KS, мы рассматривали затем также префиксную сложность KP и монотонную сложность KM. Все три определялись с помощью кратчайших описаний, но способы описаний были из разных классов. Для простой сложности это были просто вычислимые функции, для префиксной сложности — вычислимые непрерывные отображения $\Sigma \to \mathbb{N}_\perp$, для монотонной — вычислимые непрерывные отображения $\Sigma \to \Sigma$.

Для единообразия можно считать, что простая колмогоровская сложность определяется с помощью непрерывных вычислимых отображений $\mathbb{N}_\perp \to \mathbb{N}_\perp$, должным образом определив это понятие. Топология на множестве $\mathbb{N}_\perp$ и само это множество были описаны в разделе 4.4.3 (с. 104). Несложно проверить, что непрерывные отображения $\mathbb{N}_\perp \to \mathbb{N}_\perp$ бывают двух видов: если $\perp$ отображается в некоторое натуральное число (а не в $\perp$), то отображение постоянно; если же $\perp$ отображается в $\perp$, то натуральные числа могут отображаться куда угодно независимо друг от друга: отображения второго типа находятся во взаимно однозначном соответствии с частичными функциями из $\mathbb{N}$ в $\mathbb{N}$, если значение $\perp$ ставить в соответствие точкам, на которых функция не определена. Как и раньше, естественно назвать отображение $f\colon \mathbb{N}_\perp \to \mathbb{N}_\perp$ *вычислимым*, если множество пар $\langle x, y\rangle$, для которых $y \preccurlyeq f(x)$, перечислимо. Ясно, что все постоянные отображения вычислимы, а среди непостоянных вычислимы как раз те, которым соответствуют вычислимые в обычном смысле функции (вычислимость функции равносильна перечислимости графика). Поэтому ничего нового мы не получаем: добавляются постоянные функции, отображающие элемент $\perp$ и все натуральные числа в некоторое c, но с точки зрения сложности они ничего не дают. (Формально говоря, такие функции следует отличать от функций, которые соответствуют постоянным всюду определённым функциям $\mathbb{N} \to \mathbb{N}$, поскольку первые отображают $\perp$ в натуральное число, а вторые — в $\perp$.)

Весь этот формализм, впрочем, нужен лишь в качестве мотивировки схемы, объясняющей происхождение рассмотренных нами сложностей (см. рис. 6.1). Каждая из трёх сложностей получается, если в качестве декомпрессоров (способов описания) рассматривать вычислимые непрерывные отображения соответствующих пространств (описаний в объекты).

Одна из клеток в этой таблице пока не заполнена; она соответствует способам описания, которые являются вычислимыми непрерывными отображениями $\mathbb{N}_\perp$ в Σ. Сейчас мы определим этот вид сложности, который называется *сложностью разрешения* и будет обозначаться KR (иногда его ещё обозначают KD, но в последнее

		объекты	
		$\mathbb{N}_\perp$	Σ
описания	$\mathbb{N}_\perp$	KS	?
	Σ	KP	KM

Рис. 6.1. Происхождение трёх видов сложности.

время это обозначение стало применяться в ином смысле, для так называемой distinguishing complexity, так что мы будем его избегать). Такой вид сложности был определён Д. Лавлэндом в [90].

Будем рассматривать в качестве способов описания машины, которые на вход получают двоичное слово (как завершённый объект, скажем, на ленте с маркером конца после этого слова) и на выходе печатают бит за битом. При этом они могут и не останавливаться, так что для каждого входного слова x получается конечная или бесконечная последовательность. (Если эта последовательность бесконечна, то она, очевидно, вычислима.)

Таким образом, машине такого типа соответствует отображение множества всех двоичных слов (которые можно отождествить с натуральными числами, входящими в $\mathbb{N}_\perp$) в множество Σ конечных и бесконечных последовательностей. Если M — такая машина, то сложностью $KR_M(x)$ слова x относительно способа описания M называется длина кратчайшего слова y, для которого $M(y)$ (выходная последовательность машины, получившей y на входе) начинается на x.

192 Проверьте, что среди всех машин указанного вида имеется оптимальная машина M, для которой KR_M минимально с точностью до $O(1)$.

193 Дайте определение вычислимых непрерывных отображений $\mathbb{N}_\perp \to \Sigma$. Чем они отличаются от рассмотренных нами машин? Почему это различие несущественно для определения сложности? [Указание: значение на $\perp$ может быть непустым.]

Таким образом, пустая клеточка в нашей таблице заполняется (см. рис. 6.2).

		объекты	
		$\mathbb{N}_\perp$	Σ
описания	$\mathbb{N}_\perp$	KS	KR
	Σ	KP	KM

Рис. 6.2. Четыре вида сложности.

В следующей теореме перечислены основные свойства сложности разрешения.

Теорема 127. (а) *Если слово x является началом слова y, то $KR(x) \leqslant KR(y)$.*

(б) *Сложность начальных отрезков последовательности $\omega \in \Omega$ (неубывающая с ростом длины отрезка) ограничена тогда и только тогда, когда последовательность ω вычислима. (Назвав предел сложности начальных отрезков сложностью разрешения последовательности ω, можно сказать, что сложность разрешения бесконечной последовательности конечна тогда и только тогда, когда последовательность вычислима.)*

(в) $KR(x) \leqslant KS(x) + O(1)$ *для любого слова* x.

(г) $KR(x) \leqslant KM(x) + O(1)$ *для любого слова* x.

(д) $KM(x) \leqslant KR(x) + O(\log KR(x))$ *для любого слова* x.

(е) $KS(x\,|\,l(x)) \leqslant KR(x) + O(1)$ *для любого слова* x.

(ё) *Если* $f\colon \Sigma \to \Sigma$ *— вычислимое непрерывное отображение, то* $KR(f(x)) \leqslant KR(x) + O(1)$ *(константа зависит от* f*, но не от* x*).*

(ж) *Если* $f\colon \Sigma \to \mathbb{N}_\perp$ *— вычислимое непрерывное отображение, то* $KS(f(x)) \leqslant KR(x) + O(1)$ *(константа зависит от* f*, но не от* x*).*

(з) *Если* $f\colon \mathbb{N}_\perp \to \Sigma$ *— вычислимое непрерывное отображение, то* $KR(f(x)) \leqslant KS(x) + O(1)$ *(константа зависит от* f*, но не от* x*).*

(и) *Любой набор попарно несравнимых слов (ни одно не является началом другого), все из которых имеют сложность разрешения меньше* n*, содержит менее* 2^n *слов.*

(й) *Функция* KR *перечислима сверху.*

(к) *Функция* KR *является минимальной (с точностью до константы) функцией, обладающей двумя предыдущими свойствами: если* (i) *функция* K *перечислима сверху и* (ii) *любое множество несравнимых слов, у которого* $K(x) < n$ *для всех элементов, содержит* $O(2^n)$ *слов, то* $KR(x) \leqslant K(x) + O(1)$.

(л) $KR(x) \leqslant KA(x) + O(1)$ *для всех слов* x.

◀ (а) Непосредственно следует из определения (описание слова есть описание любого его начала).

(б) Пусть последовательность ω вычислима. Возьмём в качестве способа описания машину, которая независимо от входа печатает ω на выходе бит за битом. Сложность начальных отрезков последовательности ω при таком способе равна нулю (пустой вход является их описанием), и потому при оптимальном способе описания ограничена. Обратно, если сложность начальных отрезков ограничена, то некоторое описание годится для бесконечного числа начальных отрезков, и потому последовательность вычислима.

(в) Всякая вычислимая частичная функция, аргументами и значениями которой являются двоичные слова, может рассматриваться как способ описания в нашем смысле (пока вычисление не закончилось, ничего не печатаем, как только закончилось — печатаем результат бит за битом).

(г) Всякое вычислимое непрерывное отображение $\Sigma \to \Sigma$ может рассматриваться как способ описания в нашем смысле (после ограничения на конечные аргументы; можно сказать, что мы с самого начала подаём описание на вход корректной машины и более никаких клавиш не нажимаем).

(д) Пусть $R: \mathbb{N} \to \Sigma$ — оптимальный способ описания при определении сложности разрешения. Рассмотрим вычислимое отображение $\hat{R}: \Sigma \to \Sigma$, для которого

$\hat{R}(\hat{x}u) = R(x)$, где $\hat{x}$ — самоограниченный код слова x (скажем, само x, предварённое двоичной записью его длины с удвоенными битами и разделителем 01), а u — любое слово (последнее необходимо для монотонности).

(е) Пусть снова $R\colon \mathbb{N} \to \Sigma$ — оптимальный способ описания для сложности разрешения. Определим способ условного описания S, положив $S(y, n)$ равным n первым битам последовательности $R(y)$ (если n больше длины последовательности, то $S(y, n)$ не определено).

(ё) Рассмотрим новый способ описания, являющийся композицией оптимального способа описания и отображения f, и сравним его с оптимальным.

(ж) Рассмотрим композицию оптимального способа описания для сложности разрешения и отображения f как способ описания для простой колмогоровской сложности.

(з) Рассмотрим композицию оптимального способа описания для простой колмогоровской сложности и отображения f как способ описания для сложности разрешения.

(и) Попарно несравнимые слова не могут иметь общего описания (в этом случае они были бы началами описываемой им последовательности). Если сложности попарно несравнимых слов меньше n, то их описания — различные слова длины меньше n, а таких слов меньше 2^n.

(й) Применяя параллельно оптимальный способ описания ко всем словам, мы получаем постепенно улучшающиеся верхние оценки для функции KR (постепенно обнаруживая новые описания).

(к) Это — первое содержательное утверждение теоремы (до сих пор были лишь простые вариации на знакомые темы).

Пусть дана функция K, обладающая двумя указанными свойствами. Увеличивая её на константу, можно считать без ограничения общности, что существует не более 2^n несравнимых слов x, для которых $K(x) < n$.

Мы построим способ описания, при котором всякое слово x с $K(x) < n$ будет иметь описание длины ровно n. Это делается параллельно и независимо для каждого n. А именно, наблюдая за уменьшающимися верхними оценками для K, мы пополняем список слов x, для которых $K(x) < n$. Можно считать, что в каждый момент этот список конечен и с течением времени растёт. Рассмотрим поддерево двоичного дерева, состоящее из слов этого списка и всех их начал. Со временем это поддерево растёт. В каждый момент у него не более 2^n листьев, поскольку листья являются несравнимыми словами x, для которых $K(x) < n$. Каждому листу мы присвоим метку, которая является словом длины n. При добавлении нового слова к дереву это новое слово либо продолжает какой-то из листьев (который перестаёт быть листом), либо ответвляется от дерева во внутренней вершине. В первом случае новое слово становится листом, который наследует метку прежнего листа. Во втором случае для нового листа мы выделяем новое слово в качестве метки (что всегда возможно, так как листьев меньше 2^n).

Фиксируем метку и посмотрим, что происходит с листьями с этой меткой. Вначале таких листьев нет (метка не использована). Возможно, так и останется навсегда (до этой метки дело не дойдёт), но если дойдёт, то метка будет сдвигаться по дереву вверх (каждое следующее её положение будет продолжением предыдуще-

го), и потому ей соответствует некоторая конечная или бесконечная ветвь в дереве (последовательность нулей и единиц). Возникает отображение слов длины n в Σ (при этом меткам, до которых дело не дойдёт, соответствуют последовательности нулевой длины).

Объединив эти отображения при всех n, получим способ описания, при котором (для любого n) сложность всех слов x с $K(x) < n$ не превосходит n, что и требовалось.

(л) Если x_i — несравнимые двоичные слова, то $\sum 2^{-KA(x_i)} \leqslant 1$ (поскольку $2^{-KA(x_i)}$ есть априорная вероятность множества Σ_{x_i}, а эти множества не пересекаются). Поэтому несравнимых слов, у которых $KA(x_i) < n$, не может быть больше 2^n, и осталось воспользоваться предыдущим утверждением теоремы. ▶

194 Покажите, что сложность разрешения можно определять так: для каждой вычислимой функции S двух аргументов (первый — слово, второй — натуральное число; значения — нули и единицы) определим $KR_S(x)$ для любого слова $x = x_0 \ldots x_{n-1}$ как наименьшую длину слова y, при котором $S(y, i) = x_i$ при всех $i = 0, 1, \ldots, n-1$, после чего среди всех функций S выберем оптимальную.

195 Покажите, что сложность разрешения слова x равна (с точностью до $O(1)$) минимуму $KS(p)$ по всем программам p (данного языка программирования, скажем, паскаля), которые не имеют входа и печатают на выходе слово x или его продолжение.

196 Покажите, что если в предыдущей задаче вместо KS взять KP, то получится верхняя оценка для монотонной сложности. Покажите, что она не является $O(1)$-точной. [Указание. Монотонная сложность всех слов длины n не превосходит n. Печатающие их (или их продолжения) программы должны быть все различны, и слов префиксной сложности не более n не хватит.]

6.2. Сравнение сложностей

Сложности, входящие в рассмотренную нами таблицу (с двумя вариантами для пространств описаний и двумя вариантами для описываемых объектов), можно изобразить в виде ромба, стороны которого соответствуют неравенствам между сложностями (с точностью до $O(1)$), как показано на рис. 6.3.

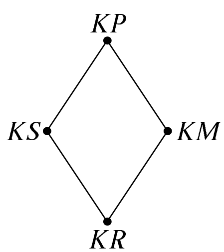


Рис. 6.3. Неравенства между сложностями.

Если мы хотим избежать упоминания топологических понятий и определения вычислимых непрерывных отображений для разных пространств (хотя они здесь по существу, как учит теория абстрактных типов данных в смысле Скотта и близкая к ней теория f_0-пространств Ершова, см. [152]), то для построения указанных четырёх сложностей по единой схеме можно обойтись следующей упрощённой конструкцией [168].

На множестве $\Xi = \mathbb{B}^*$ всех двоичных слов рассмотрим два бинарных отношения: равенство слов ($=$) и сравнимость слов (наличие общего продолжения, обозначение $\asymp$). Пусть α и β — любые из этих двух отношений (так что всего есть четыре варианта).

Будем говорить, что множество $S \subset \Xi \times \Xi$ является α-β-корректным, если следующее условие выполнено для всех слов x_1, x_2, y_1, y_2:

$$(x_1, y_1) \in S,\ (x_2, y_2) \in S,\ x_1 \alpha x_2 \ \Rightarrow\ y_1 \beta y_2.$$

Например, $=$-$=$-корректные отношения — это униформные множества, то есть графики функций.

$\boxed{\textbf{197}}$ (**а**) Покажите, как по $\asymp$-$=$-корректному отношению построить непрерывное отображение $\Sigma \to \mathbb{N}_\perp$.

(**б**) Покажите, как по $\asymp$-$\asymp$-корректному отношению построить непрерывное отображение $\Sigma \to \Sigma$.

(**в**) Покажите, как по $=$-$\asymp$-корректному отношению построить непрерывное отображение $\mathbb{N}_\perp \to \Sigma$.

Будем называть *α-β-способом описания* перечислимое α-β-корректное отношение на $\Xi \times \Xi$. Для всякого способа описания S определим сложность как функцию, сопоставляющую с каждым словом x длину его кратчайшего описания, то есть длину кратчайшего слова y, для которого $\langle y, x \rangle$ принадлежит S.

Теорема 128. *Для каждой из четырёх комбинаций $\alpha, \beta \in \{=, \asymp\}$ существует оптимальный α-β-способ описания, и соответствующая сложность совпадает (с точностью до $O(1)$) с одной из четырёх сложностей KS, KP, KM, KR.*

◀ В каждом из случаев по α-β-корректному способу описания легко строится вычислимое отображение соответствующих пространств (см. задачу 197), задающее ту же самую сложность, и наоборот. ▶

Таким образом, в таблице сложностей можно поменять названия строк и столбцов (см. рис. 6.4).

$\boxed{\textbf{198}}$ Покажите, как определить для пары слов:

(**а**) монотонную сложность, рассматривая в качестве декомпрессоров вычислимые непрерывные отображения $\Sigma \to \Sigma \times \Sigma$. (Такие отображения можно отождествить с парами вычислимых отображений $\Sigma \to \Sigma$.)

(**б**) априорную вероятность, рассматривая машины с датчиком случайных чисел и двумя выходными лентами, где можно печатать бит за битом.

(**в**) сложность разрешения, рассматривая вычислимые непрерывные отображения $\mathbb{N}_\perp \to \Sigma \times \Sigma$.

описания \ объекты	$=$	$\asymp$
$=$	KS	KR
$\asymp$	KP	KM

Рис. 6.4. α-β-сложности.

Аналогичные определения возможны и для троек, четвёрок и т. д.

199 Покажите, что сложность разрешения пары $\langle x, y\rangle$ не превышает $l(x) + l(y)$. [Указание. Слово z можно считать описанием пары $\langle z, z^R\rangle$, где z^R — слово z, прочитанное с конца.]

Удивительным образом такое же свойство верно для троек слов [57] и даже для кортежей большей длины (это следует из результатов Нидеррайтера [126]). А для монотонной сложности пар аналогичное свойство неверно, как показал Павел Карпович [57]: значение $KM(x, y)$ может превосходить $l(x) + l(y)$ на величину порядка $\log n$ для слов длины n. Отсюда следует, в частности, что для пар априорная сложность отличается от монотонной на величину того же порядка.

Другая схема (восходящая к Левину, см. [77]) классификации сложностей состоит в описании их как наименьших перечислимых сверху функций, удовлетворяющих некоторым ограничениям. Соберём соответствующие ограничения (на перечислимую сверху функцию K):

- число различных слов x, при которых $K(x) < n$, есть $O(2^n)$ (простая сложность KS, теорема 8, с. 29);
- ряд $\sum_x 2^{-K(x)}$ сходится (префиксная сложность KP, теорема 62, с. 116);
- число различных несравнимых слов x, при которых $K(x) < n$, есть $O(2^n)$ (сложность разрешения KR, теорема 127, с. 219);
- $\sum_x 2^{-K(x_i)} \leqslant 1$ для любого множества попарно несравнимых слов x_i (априорная сложность KA, теорема 80, с. 145).

В этой схеме фигурируют те же четыре сложности, за одним исключением: вместо монотонной сложности фигурирует априорная. (Для случая префиксной сложности такого расхождения нет, поскольку она совпадает с логарифмом максимальной перечислимой снизу полумеры на $\mathbb{N}$.)

Поэтому при объединении этих двух схем у нас получается уже не ромб, а пятиугольник (рис. 6.5).

Напомним основные результаты о сравнении сложностей в этом пятиугольнике. Прежде всего, все сложности отличаются не более чем на $O(\log n)$ для слов длины n. В самом деле, $KP(x) \leqslant KS(x) + O(\log KS(x))$ (теорема 65, с. 118). С другой стороны, $KS(x) \leqslant KS(x \,|\, l(x)) + KS(l(x)) + O(\log n) \leqslant KR(x) + O(\log n)$. По-

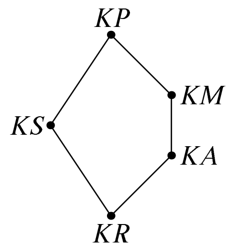


Рис. 6.5. Пять сложностей.

этому две наиболее различающиеся сложности в этом пятиугольнике (верхняя и нижняя) отличаются не более чем на $O(\log n)$ для слов длины n.

Более сложная ситуация возникает, если мы хотим ограничить разность между двумя сложностями константой, умноженной на логарифм одной из них. Это можно сделать для неравенств вдоль линий, идущих в «северо-восточном» направлении:

$$KP(x) \leqslant KS(x) + O(\log KS(x))$$

(см. теорему 65) и

$$KM(x) \leqslant KR(x) + O(\log KR(x))$$

(теорема 127). Тем более это верно, если заменить KM на KA (как мы уже упоминали в задаче 140, с, 165). Для «северо-западных» неравенств это сделать нельзя: KM и KR ограничены для начальных отрезков вычислимых последовательностей (в частности, для слов из одних нулей), а KS и KP — нет (для слова из n нулей сложность равна сложности n, и достигает $\log n$ при некоторых n). Мы уже обсуждали это в теореме 86, где отмечали, что разница между KP и KM может быть обоих знаков порядка $\log n$ для некоторых слов длины n (при бесконечно многих n). Теорема Гача – Дея показывает, что разница между KM и KA не ограничена и для некоторых слов достигает примерно $\log\log n$ или больше (теорема 87).

Отметим также, что из теоремы 73 (с. 130) следует, что разность между $KP(x)$ и $KS(x)$ не только не ограничена, но и стремится к бесконечности. Некоторые другие оценки разности между сложностями $KS(x)$ и $KP(x)$ (а также другими парами сложностей) приведены в работе [168].

6.3. Условные сложности

Мы уже упоминали несколько видов условной сложности (когда определяется сложность одного слова относительно другого). В разделе 2.2 мы определяли условную сложность $KS(x\,|\,y)$ как длину кратчайшего описания p слова x при известном y, то есть минимальную длину слова p, для которого $S(p,y) = x$. Здесь S — способ описания; способами описания были произвольные частичные вычислимые функции двух аргументов.

В разделе 4.7 определялась условная префиксная сложность $KP(x\,|\,y)$. При этом требовалось, чтобы способ описания был префиксно корректным *относительно p*: это значит, что если слово p является описанием данного x при данном y, то любое продолжение слова p также является таким описанием.

Наконец, при доказательстве теоремы 93 мы упоминали условную монотонную сложность $KM(x\,|\,y)$. При её определении в качестве способа описания берётся вычислимое семейство вычислимых непрерывных отображений $D_y\colon \Sigma \to \Sigma$, индексированное словами y. Вычислимость семейства понимается в том смысле, что множество троек $\langle y, u, v\rangle$, для которых $v \preccurlyeq D_y(u)$, перечислимо.

Аналогичным образом можно было бы определить и условную сложность разрешения.

Во всех этих определениях условия рассматриваются как законченные слова, никак не связанные друг с другом: если p является описанием x при известном y, для другого слова y то же самое p может быть описанием совершенно другого x (даже если новое y отличается от старого добавлением одного бита).

Другими словами, способ описания (скажем, для условной префиксной сложности) мы считаем вычислимым отображением $D\colon \Sigma \times \mathbb{N} \to \mathbb{N}_\perp$, при этом первый член пары $\langle p, y\rangle \in \Sigma \times \mathbb{N}$ считается описанием (требуется монотонность по p), а второй условием (монотонность по y не требуется).

Если рассматривать условия как вершины дерева и требовать монотонности по этому аргументу, получатся четыре других вида условной сложности, которые, однако, практически не рассматривались (работа [29] — одно из редких исключений).

Таким образом, всего возникает 8 видов условных сложностей (каждый из трёх компонентов — условия, описания и объекты — можно понимать двумя способами). Технически проще всего определить эти условные сложности так. Пусть $\alpha, \beta, \gamma \in \{=, \asymp\}$ (см. раздел 6.2). Определим $(\alpha, \beta)|\gamma$-способ описания как перечислимое множество S троек слов $\langle p, x, y\rangle$, удовлетворяющее следующему требованию:

$$\langle p_1, x_1, y_1\rangle \in S,\ \ \langle p_2, x_2, y_2\rangle \in S,\ \ p_1 \alpha p_2,\ \ y_1 \gamma y_2 \ \Rightarrow\ x_1 \beta x_2.$$

Далее сложность $K_S(x\,|\,y)$ определяется как наименьшая длина слова p, при котором $\langle p, x, y\rangle \in S$.

Теорема 129. *В каждом из восьми случаев среди способов описания существует оптимальный* (*для которого сложность минимальна в этом классе способов описания с точностью до* $O(1)$).

$\boxed{\textbf{200}}$ Проведите аккуратное доказательство этой теоремы (по существу ничем не отличающееся от случая обычной или префиксной условной сложности).

Соответствующую оптимальному $(\alpha, \beta)|\gamma$-способу описания сложность можно обозначить $K_{(\alpha,\beta)|\gamma}$. При таком обозначении $KP(x\,|\,y)$ есть $K_{(\asymp,=)|=}$, а $KS(x\,|\,y)$ есть $K_{(=,=)|=}$.

$\boxed{\textbf{201}}$ Покажите, что от замены $=$ на $\asymp$ в качестве γ условная сложность может только возрасти. [Указание: на способ описания накладываются дополнительные ограничения, поэтому их становится меньше. По этой же причине обычная сложность не превосходит префиксной.]

Было бы интересно исследовать эти сложности подробнее; видимо, этим никто не занимался. Насколько, скажем, велико возрастание, отмеченное в этой задаче?

Вот пример утверждения, включающего в себя рассмотренные только что модификации условных сложностей:

$\boxed{\textbf{202}}$ Докажите, что $KS(x) \leqslant K_{(=,=)|\asymp}(x|y) + KR(y) + O(\log KR(y))$.

Опишем ещё один подход к определению условной сложности, восходящий к колмогоровской интерпретации логических связок как операций над задачами [61]. Условную сложность x при известном y можно интерпретировать как сложность задачи «указать x по данному y»; в свою очередь, эту задачу можно считать множеством всех функций, переводящих y в x (любая такая функция считается «решением» этой задачи).

Это можно уточнить, например, так. Рассмотрим пространство $\mathbb{F}$, элементами которого являются все частичные функции с натуральными аргументами и значениями. На этом множестве определим порядок: $f_1 \preccurlyeq f_2$, если функция f_2 продолжает функцию f_1. *Конечными элементами* этого множества будем считать функции с конечной областью определения. Для каждого конечного элемента f рассмотрим конус над ним, то есть множество всех его продолжений (конечных и бесконечных). Непрерывное отображение $T\colon \mathbb{N}_\perp \to \mathbb{F}$ будем называть *вычислимым*, если перечислимо множество пар $\langle a, f\rangle$, где $a \in \mathbb{N}_\perp$, f — конечный элемент $\mathbb{F}$ и $f \preccurlyeq T(a)$. Такие вычислимые отображения будем считать способами описания (функций). Для каждой функции f определим сложность (относительно данного способа описания T) как минимальную длину слова (точнее, логарифм числа, ведь мы отождествляем числа и двоичные слова) a, для которого $f \preccurlyeq T(a)$.

$\boxed{\textbf{203}}$ Докажите, что при таком определении существует оптимальный способ описания и что сложность функции $y \mapsto x$ (область определения состоит из единственного слова y, значение на котором равно x) равна $KS(x\,|\,y) + O(1)$.

В этом же духе можно интерпретировать и все восемь указанных выше условных сложностей (для пространств $Y, X \in \{\mathbb{N}_\perp, \Sigma\}$ определив пространство функций $(Y \to X)$, а затем в качестве способов описания рассматривая вычислимые отображения пространства описаний $P \in \{\mathbb{N}_\perp, \Sigma\}$ в пространство $(Y \to X)$. При этом удобно использовать домены Скотта или f_0-пространства в смысле Ершова (подробности см. в [152]).

Несколько другая интерпретация условной сложности как сложности задачи «получить x из y», не использующая вычислимость в пространствах функций, рассмотрена в главе 13.

Родственное (хотя и несколько другое) понятие сложности функции было рассмотрено Шнорром [145, 147]. Напомним, что в теории рекурсии *нумерацией* называется некоторое отображение ν, сопоставляющее с каждым натуральным числом n некоторую функцию ν_n (частичную) из $\mathbb{N}$ в $\mathbb{N}$. Нумерация ν называется *вычислимой*, если частичная функция двух аргументов $\langle n, x\rangle \mapsto \nu_n(x)$ является вычислимой, и *гёделевой*, или (в русскоязычной литературе) *главной*, если для любой другой вычислимой нумерации μ существует вычислимая всюду определённая функция h, *сводящая* μ к ν в том смысле, что $\mu_n = \nu_{h(n)}$ при всех n. (В частности, среди ν_n должны встречаться все вычислимые функции.)

Следуя Шнорру, усилим это условие и дополнительно потребуем, чтобы $h(n) = O(n)$ (длина слова $h(n)$ превосходила бы длину n не более чем на константу, если отождествить натуральные числа с двоичными словами). Если для всякой вычислимой нумерации μ существует функция h с таким свойством, то нумерацию ν будем называть *оптимальной*.

Теорема 130. *Оптимальные нумерации существуют.*

◀ Выберем какой-нибудь естественный язык программирования двухместных функций и будем считать слово $\hat{u}v$ номером функции, которая получится, если в функции с программой u фиксировать первый аргумент равным v. (Здесь $\hat{u}$ — самоограниченное описание слова u, получаемое, например, удвоением каждого бита и приписыванием 01 в конец.) ▶

Шнорр (см. [145, 147]) предложил определить сложность вычислимой функции как логарифм её номера в оптимальной нумерации. (При этом минимальная сложность функции, отображающей x в y, также оказывается равной $KS(y|x)$.) Он показал также, что любые две оптимальные нумерации ν_1 и ν_2 отличаются вычислимой перестановкой номеров π, увеличивающей объём в ту и другую сторону не более чем на $O(1)$. (Это значит, что $\nu_1(n) = \nu_2(\pi(n))$ при всех n, причём $\pi(n) = O(n)$ и $\pi^{-1}(n) = O(n)$.)

6.4. Сложности и оракулы

6.4.1. Сложность с оракулом

В теории вычислимых функций хорошо известен следующий метод, называемый *релятивизацией*: берётся какое-то определение или утверждение, касающееся класса вычислимых функций, и всюду в нём вычислимые функции заменяются на функции, вычислимые относительно некоторого *оракула*. В качестве такого оракула берётся некоторая всюду определённая функция α, аргументами и значениями которой являются натуральные числа или двоичные слова (обычно характеристическая функция некоторого множества A). Алгоритмам разрешается в качестве элементарного действия обращаться к «внешней процедуре», возвращающей значение $\alpha(n)$ для переданного ей значения параметра n. (Для характеристической функции множества A это означает, что про любое натуральное число можно узнать, принадлежит ли оно множеству A или нет.) Если сама функция α не вычислима (множество A неразрешимо), то класс вычислимых функций расширяется; вычисляемые такими алгоритмами функции называются α-вычислимыми.

После этого практически все понятия и теоремы теории вычислимых функций переносятся на α-вычислимые функции. Скажем, можно говорить об α-перечислимых множествах, или об α-вычислимых действительных числах, или (подходя ближе к колмогоровской сложности) об α-перечислимых снизу полумерах. При этом (практически все) теоремы остаются верными и в таком «релятивизованном» варианте.

В частности, для любого множества A можно определить понятие колмогоровской сложности, релятивизованной относительно множества A, рассматривая

способы описания (декомпрессоры) с оракулом A. Это можно сделать и для обычной, и для префиксной, и для всех других видов рассмотренных нами сложностей (условных и безусловных). Обычно использование оракула отмечают верхним индексом, так что, скажем, $KP^A(x)$ есть префиксная сложность слова x с оракулом A.

На самом деле мы делаем даже несколько большее: мы не просто для фиксированного оракула A определяем релятивизованную сложность (с точностью до $O(1)$), а определяем (с этой точностью) некоторую функцию двух аргументов: $\langle x, A\rangle \mapsto K^A(x)$ (где K — один из четырёх видов сложности).

$\boxed{\mathbf{204}}$ Проверьте, что это действительно можно сделать, и что получающиеся сложности являются пределами условных сложностей, описанных в разделе 6.3:

$$KP^A(x) = K^A_{\asymp,=}(x) = \lim_{n\to\infty} K_{(\asymp,=|\asymp)}(x\,|\,A_n),$$

где A_n — начальный отрезок длины n характеристической функции множества A. (Аналогично и для других видов сложностей.)

Заметим, что релятивизованные сложности (с точностью $O(1)$) не больше обычных (алгоритм может не обращаться к оракулу, потому среди A-способов описания есть и обычные способы описания).

При некоторых A функция сложности с оракулом A может существенно уменьшаться (по сравнению со сложностью без оракула). Например, рассмотрим в качестве оракула A универсальное перечислимое неразрешимое множество. (Такой оракул обычно обозначают $\mathbf{0}'$.) Другими словами, $\mathbf{0}'$-оракул есть внешний оракул для проблемы остановки: ему можно послать для анализа программу (без оракула), и он скажет, завершится её работа или нет. Имея такой оракул, можно для каждого слова x найти его кратчайшее описание (поскольку мы можем с помощью оракула проверять, закончит декомпрессор работу на данном слове или нет). Поэтому функция KS является $\mathbf{0}'$-вычислимой (это же относится к функции KP, условным сложностям и т.п.), а список всех слов сложности меньше n, который без оракула имел сложность $n + O(1)$, а также числа $B(n)$ и $BB(n)$ (см. раздел 1.2) теперь имеют $\mathbf{0}'$-сложность $O(\log n)$.

С другой стороны, большинство слов длины n имеют $\mathbf{0}'$-сложность $n - O(1)$, и потому для них $\mathbf{0}'$-сложность близка к обычной (нерелятивизованной).

$\boxed{\mathbf{205}}$ Покажите, что если для некоторого множества A его использование в качестве оракула не меняет функции KS, то есть если $KS(x) = KS^A(x)$, то A разрешимо. Аналогичное утверждение верно для KM, KR, KA. [Указание. Вычислимость последовательности можно характеризовать в терминах скорости роста сложности начальных отрезков, см. задачу 49, с. 53.]

Как мы уже отмечали, для префиксной сложности аналогичное утверждение неверно: существуют KP-низкие множества (в английских текстах их называют K-low, поскольку префиксная сложность обычно обозначается буквой K), добавление которых в качестве оракула не меняет KP. Этот важный и сложный результат был получен сравнительно недавно (см. [127, 37]).

6.4.2. Сложность при условии больши́х чисел

Определим новый вид условной сложности: сложность числа (слова x) относительно множества A. Эта сложность, говоря неформально, есть сложность такой задачи: «получить x, если задан некоторый элемент множества A» (неизвестно какой). Такую сложность можно определить несколькими эквивалентными способами.

Пусть фиксирован некоторый разумный язык программирования. (Формально говоря, нужно, чтобы соответствующая ему нумерация была главной, то есть чтобы была возможна эффективная трансляция программ с других языков [175].) Определим условную сложность объекта (слова, натурального числа) x относительно множества объектов A как минимальную (простую колмогоровскую) сложность программы, переводящей *любой* из элементов множества A в x. (В более общем виде это определение рассматривается в главе 13.)

Возможность трансляции гарантирует, что это определение корректно, то есть что сложность (с точностью до $O(1)$) не зависит от выбора главной нумерации.

Важно отличать эту сложность от сложности x относительно конечного множества A, заданного списком своих элементов (известен не список элементов A, а некоторый его элемент — неважно какой). Мы будем обозначать только что определённую сложность $KS(x \,\|\, A)$, сохраняя обозначение $KS(x \,|\, A)$ для сложности x относительно конечного множества A, заданного списком своих элементов.

Другой (эквивалентный) способ определения $KS(x \,\|\, A)$ состоит в следующем. Пусть D — вычислимая частичная функция двух аргументов (декомпрессор), x — двоичное слово, а A — множество двоичных слов. Определим $KS_D(x \,\|\, A)$ как наименьшую длину такого слова p, что $D(p, y) = x$ для всех $y \in A$.

$\boxed{\textbf{206}}$ Докажите, что среди всех вычислимых частичных функций имеется оптимальная функция (для которой функция KS_D минимальна с точностью до $O(1)$). Докажите, что для оптимальной функции D величина $KS_D(x \,\|\, A)$ совпадает с ранее определённой сложностью $KS(x \,\|\, A)$ с точностью до $O(1)$.

Для случая одноэлементного множества $A = \{a\}$ обе величины $KS(x \,|\, A)$ и $KS(x \,\|\, A)$ совпадают (с точностью до $O(1)$) с обычной условной сложностью $KS(x \,|\, a)$; см. задачу 28.

Рассмотрим в качестве A множество всех чисел, больших некоторого числа n. (Как всегда, числа мы отождествляем со словами.) Сложность x относительно этого множества будем обозначать $KS(x \,\| \geqslant n)$. Очевидно, эта сложность не превосходит $KS(x)$ и убывает с возрастанием n (вообще $KS(x \,\|\, A)$ очевидным образом убывает при уменьшении множества A, превращаясь в $O(1)$ для пустого A). Поэтому у неё есть некоторый предел при $n \to \infty$.

Теорема 131.
$$\lim_{n\to\infty} KS(x \,\| \geqslant n) = KS^{\mathbf{0}'}(x) + O(1).$$

◀ Пусть указанный предел равен k. Тогда существует программа p сложности k, переводящая все достаточно большие числа в x. Имея дополнительно оракул $\mathbf{0}'$, можно использовать эту программу как описание объекта x. Именно, надо путём перебора найти такие N и y, что программа p *не переводит никакое* $n \geqslant N$ *ни*

в какой объект, отличный от y. Это можно сделать с помощью оракула, поскольку выделенное свойство пары N, y имеет перечислимое дополнение. При этом из нашего предположения следует, что $y = x$. Поэтому

$$KS^{\mathbf{0}'}(x) \leqslant \lim_{n\to\infty} KS(x \,\|\, \geqslant n) + O(1).$$

Напротив, пусть имеется описание y объекта x относительно $\mathbf{0}'$-оптимального способа описания, имеющее длину k. Рассмотрим следующую программу: получив некоторое N, сделать N шагов порождения универсального перечислимого множества $\mathbf{0}'$ и затем использовать полученное множество в качестве оракула, применяя оптимальный способ описания к y. Эта программа эффективно строится по y и потому её сложность не превосходит $KS(y) + O(1) \leqslant l(y) + O(1) = k + O(1)$. С другой стороны, для достаточно больших N эта программа порождает x (поскольку при вычислении значения оптимального способа описания на y используется лишь конечное число запросов к оракулу, при достаточно большом N все эти запросы получат правильные ответы даже и при замене настоящего оракула на его N-шаговое приближение). ►

Оказывается, что аналогичное утверждение верно, если заменить $KS(x \,\|\, \geqslant n)$ на величину $\sup_{m\geqslant n} KS(x \,|\, m)$. Сразу же ясно, что

$$\sup_{m\geqslant n} KS(x \,|\, m) \leqslant KS(x \,\|\, \geqslant n),$$

поскольку оптимальная программа в правой части годится и для любого m из левой. Удивительным образом оказывается, что это уменьшение не меняет предела при $n \to \infty$:

Теорема 132.

$$\limsup_{n\to\infty} KS(x \,|\, n) = KS^{\mathbf{0}'}(x) + O(1).$$

◄ Нужно доказать, что если для данного слова x выполнено свойство

сложность $KS(x \,|\, n)$ меньше k при всех достаточно больших n,

то $\mathbf{0}'$-сложность x не превосходит $k + O(1)$. Сложность здесь — в отличие от предыдущей теоремы — в том, что существующие по предположению программы (длины меньше k) для разных n могут быть разными.

Заметим, что слов x с указанным свойством (для данного k) меньше 2^k: если бы их было больше, то при достаточно бо́льших n на них не хватило бы программ длины меньше k.

Поэтому было бы достаточно доказать, что множество таких слов является $\mathbf{0}'$-перечислимым и соответствующий алгоритм перечисления эффективно строится по k (другими словами, что функция $x \mapsto \limsup KS(x \,|\, n)$ является $\mathbf{0}'$-перечислимой сверху). Однако естественное описание этого множества,

$$\exists N \;\; (\forall n \geqslant N) \;\; [KS(x \,|\, n) < k],$$

доказывает лишь его принадлежность к Σ_3 (условие в квадратных скобках перечислимо, и есть два дополнительных квантора), так что мы поступим иным образом.

Заметим, что нам не нужна $\mathbf{0}'$-перечислимость этого множества, нам достаточно включить его в некоторое $\mathbf{0}'$-перечислимое множество, содержащее менее 2^k элементов с данным k. Это делается следующим образом.

Рассмотрим двумерное перечислимое множество пар $\langle n, x\rangle$, для которых выполнено неравенство $KS(x\,|\,n) < k$. Это множество «тонкое» в том смысле, что все вертикальные сечения этого множества содержат менее 2^k элементов.

Пусть выбрана некоторая точка $\langle n, x\rangle$. Попробуем добавить горизонтальный луч, выходящий из этой точки, к нашему множеству (добавить в него все пары $\langle m, x\rangle$ при $m \geqslant n$). При этом множество может перестать быть тонким, а может и не перестать, и эти два случая можно различить с помощью $\mathbf{0}'$-оракула: нарушение тонкости есть перечислимое свойство (найдётся сечение, в котором обнаружатся 2^k различных элементов — вместе с добавленным).

Будем выполнять эту операцию по очереди для всех лучей (для всех пар $\langle n, x\rangle$) в некотором порядке. (При этом, если какой-то луч удалось добавить без нарушения тонкости, то в дальнейшем его элементы учитываются наравне с элементами исходного множества при добавлении следующих лучей.) Этот процесс является $\mathbf{0}'$-вычислимым, и потому x-координаты всех добавленных лучей образуют $\mathbf{0}'$-перечислимое множество.

Оно содержит менее 2^k элементов (поскольку добавление лучей не нарушает тонкости), а также содержит все x, у которых $\limsup K(x\,|\,n) < k$. В самом деле, для такого x найдётся луч, изначально лежавший в множестве, и потому его добавление заведомо возможно. ►

(Это доказательство является упрощённым вариантом рассуждения из [169]. См. также аналогичные рассуждения в [9].)

Теоремы, аналогичные теоремам 131 и 132, имеют место и для префиксной сложности. Определение префиксной сложности относительно множества требует специального обсуждения, которое мы отложим и начнём со второй из этих теорем.

Теорема 133.
$$\limsup_{n\to\infty} KP(x\,|\,n) = KP^{\mathbf{0}'}(x) + O(1).$$

◄ Переходя к априорным вероятностям (условным и безусловным), эту теорему можно переписать так:
$$\liminf_{n\to\infty} m(x\,|\,n) = m^{\mathbf{0}'}(x)$$
(с точностью до ограниченного и отделённого от нуля множителя).

Покажем сначала, что левая часть не меньше правой (точнее, меньше не более чем в $O(1)$ раз). В самом деле, рассмотрим машину с оракулом $\mathbf{0}'$, выход которой имеет распределение $m^{\mathbf{0}'}$. Теперь будем для каждого n запускать эту машину с конечной частью оракула, соответствующей n шагам его перечисления. При этом вероятность получить x на выходе может и увеличиться и уменьшиться, но нижний предел этих вероятностей не меньше соответствующей вероятности для настоящего оракула. В самом деле, вероятность получить x для настоящего оракула есть мера

открытого множества (объединения конусов), и каждый из этих конусов использует лишь конечное число вопросов к оракулу и на некотором шаге появится.

Докажем теперь обратное неравенство. Тут ситуация очень похожа на доказательство теоремы 132. Имеется перечислимое снизу семейство полумер; при каждом n функция $x \mapsto m(x\,|\,n)$ является полумерой (сумма ряда по x не больше 1). Отсюда легко следует, что для нижнего предела

$$m'(x) = \liminf_{n\to\infty} m(x\,|\,n)$$

сумма ряда $\sum_x m'(x)$ также не превышает единицы. Если бы функция m' оказалась $\mathbf{0}'$-перечислимой снизу, то всё было бы доказано — но запись этого свойства в виде

$$r < \liminf_{n\to\infty} m(x\,|\,n) \;\Leftrightarrow\; (\exists q > r)\; \exists N\; (\forall n > N)\; [q < m(x\,|\,n)]$$

содержит лишний квантор (свойство в квадратных скобках перечислимо, а не разрешимо). Но, как и в предыдущем доказательстве, нам достаточно построить $\mathbf{0}'$-перечислимую мажоранту.

Для этого будем рассматривать тройки $\langle N, x, \varepsilon\rangle$ (где ε — положительное рациональное число) и будем пытаться увеличивать функцию $m(\cdot\,|\,\cdot)$ до ε на луче, состоящем из пар $\langle n, x\rangle$ при данном x и при всех $n \geqslant N$, если это не нарушит свойство полумеры (для каждого n сумма по всем x не превышает 1).

Как и раньше, возможность такого добавления можно проверить с помощью $\mathbf{0}'$-оракула (поскольку невозможность есть перечислимое свойство). Будем производить такие добавления для всех троек (учитывая уже сделанные увеличения при определении возможности следующих). Оставляя от каждой добавленной тройки x и ε (то есть рассматривая для каждого x точную верхнюю грань всех ε, с которыми это x добавлено), получим $\mathbf{0}'$-перечислимую снизу полумеру, мажорирующую m' (ведь если m' где-то больше ε, то на некотором луче функция уже больше ε и тем самым добавление заведомо возможно). ►

Чтобы сформулировать аналогичное утверждение для $KP(x\,\|\geqslant n)$, надо прежде всего определить эту сложность (префиксную сложность с условием, являющимся множеством). Тут есть несколько возможностей, и не вполне ясно, какую из них следует считать «правильной».

Можно пытаться определить $KP(x\,\|\,A)$ как минимальную префиксную сложность программы, которая даёт x на любом входе из множества A. Однако, как показывает задача 109 (с. 121) тогда для одноэлементных множеств $A = \{a\}$ не получится $KP(x\,|\,a)$ и потому, видимо, такое определение надо считать неудачным.

Можно действовать по аналогии с задачей 206. Рассмотрим произвольную вычислимую функцию $\langle p, x\rangle \mapsto D(p, x)$ префиксно корректную по первому аргументу (при фиксированном втором). Для любого x и любого множества A определим $KP_D(k\,\|\,A)$ как наименьшую длину такого p, что $f(p, n) = k$ для всех $n \in A$. Разница с определением для обычной сложности состоит в том, что мы требуем от декомпрессора префиксной корректности по первому аргументу. Среди таких декомпрессоров есть оптимальный (в этом классе), и соответствующую функцию KP_D можно принять за определение префиксной сложности с множеством в качестве условия.

207 Покажите, что та же самая сложность (с точностью до $O(1)$) получится, если в качестве декомпрессоров рассматривать вычислимые непрерывные отображения $\Sigma \to \mathbb{F}$ (пространства конечных и бесконечных последовательностей нулей и единиц в пространство частичных функций из $\mathbb{N}$ в $\mathbb{N}$) и рассматривать кратчайшее слово p, образ которого есть некоторая частичная функция, равная x на всех элементах A.

Можно определять префиксную сложность относительно множества, используя беспрефиксные функции вместо префиксно корректных. В классе вычислимых беспрефиксных по первому аргументу функций существует функция, для которой относительная сложность $KP_f(x \,\|\, A)$ минимальна. Таким образом мы получаем определение $KP'(x \,\|\, A)$, аналогичное условной сложности $KP'(k \,|\, n)$ и совпадающее с ней при $A = \{n\}$ (с точностью до аддитивной константы).

Наконец, можно определить априорную вероятность $m(x \,\|\, A)$, взяв вероятностную машину с входом y и изучая меру множества всех последовательностей ω, которые (будучи использованы в качестве случайных битов) заставляют машину напечатать x при любом входе $y \in A$. Здесь также существует оптимальная машина (при которой эта вероятность наибольшая) и для одноэлементных множеств A это определение сводится к ранее известному.

Как и раньше, для всех k, A выполнены неравенства

$$-\log m(k \,\|\, A) \leqslant KP(k \,\|\, A) + O(1) \leqslant KP'(k \,\|\, A) + O(1).$$

Однако равенств здесь уже не получается, по крайней мере во втором неравенстве; это установила Елена Калинина [56]; про первое неравенство авторы не знают. Тем не менее можно отметить, что все эти величины превышают

$$-\log \inf_{x\in A} m(k \,|\, x) = \sup_{x\in A} KP(k \,|\, x),$$

и потому любая из них может быть использована в теореме, аналогичной теореме 131. В частности, для (видимо, наиболее естественной) величины $KP(k \,\|\, A)$, определённой выше с помощью префиксно корректных функций, получаем такое утверждение:

Теорема 134.

$$\lim_{n\to\infty} KP(k \,\|\, \geqslant n) = KP^{\mathbf{0}'}(k) + O(1).$$

208 Докажите, что величины $KP(k \,\|\, A)$, $KP'(k \,\|\, A)$ и $KS(k \,\|\, A)$ отличаются друг от друга не более, чем на линейную функцию от логарифма наименьшей из них, то есть, $O(\log KS(k \,\|\, A))$.

Авторам неизвестно, ограничена ли сверху $KS(k \,\|\, A)$ линейной или хотя бы вычислимой функцией от $-\log m(k \,\|\, A)$ (хотя бы для конечных или даже двухэлементных множеств).

Здесь уместно заметить, что существует и другой тип задач, для которого логарифм априорной вероятности может сильно отличаться от сложности. Речь идет о задачах перечисления множеств, рассмотренных в работе Соловея [162]. Будем рассматривать не останавливающиеся алгоритмы со входом, являющимся конечным

двоичным словом, и печатающие на выходе через запятую конечный или бесконечный список слов (между печатью двух соседних слов списка может быть сколь большой временной интервал, и для любого слова списка следующее за ним слово может быть вообще не напечатано никогда, однако, начав печатать какое-то слово, алгоритм обязательно должен его закончить, поставив запятую в конце). Для каждого такого алгоритма A и всякого множества слов S можно определить сложность S относительно A, как наименьшую длину входа, на котором A печатает список элементов S в произвольном порядке:

$$KE_A(S) = \min\{l(p) \mid M(p) \text{ печатает список } S\}.$$

Обычными рассуждениями нетрудно показать, что среди всех таких алгоритмов существует оптимальный. Сложность относительно него называется простой *сложностью перечисления* множества S и обозначается через $KE(S)$. Сложность перечисления конечна только для перечислимых множеств.

С другой стороны можно рассмотреть вероятностные алгоритмы перечисления множеств. Так называется любой вероятностный неостанавливающийся алгоритм без входа, который на своей выходной ленте печатает через запятую некоторый конечный или бесконечный список слов. Для любого такого алгоритма A можно рассмотреть вероятность того, что алгоритм напечатает данное множество S. Эту вероятность мы будем обозначать через $m_A(S)$. Обычными рассуждениями нетрудно показать, что среди всех алгоритмов A существует такой, для которого функция $m_A(S)$ максимальна с точностью до мультипликативной константы. Для этого алгоритма величина $m_A(S)$ называется *априорной вероятностью перечисления* множества S и обозначается через $m(S)$. Априорная вероятность положительна только для перечислимых множеств (это утверждение называется теоремой Де Леу – Мура – Шеннона – Шапиро [74]).

$\boxed{\textbf{209}}$ Докажите теорему Де Леу – Мура – Шеннона – Шапиро. [Указание. Если подмножество Ω имеет положительную меру, то в некотором интервале доля этого множества больше половины.]

Из определений следует, что $KE(S) \leqslant \log m(S) + O(1)$. Однако неизвестно, верно ли обратное неравенство, хотя бы с логарифмической точностью: $-\log m(S) \leqslant KE(S) + O(\log KE(S))$? Известно всего лишь [162], что обратное неравенство верно с мультипликативной константой 3 перед сложностью:

$$-\log m(S) \leqslant 3 \cdot KE(S) + O(\log KE(S)),$$

и для конечных множеств S константу 3 можно заменить на 2 (см. [170]).

6.4.3. Пределы частот и априорная вероятность, релятивизованная $\mathbf{0}'$

В заключение приведём результат из статьи [114], связывающий частоты появления в вычислимых последовательностях с релятивизованной относительно $\mathbf{0}'$ префиксной сложностью. (См. также упрощённое изложение в [9].)

Пусть $f(0), f(1), \ldots$ — вычислимая последовательность натуральных чисел. Для данных натуральных n и k подсчитаем, сколько раз встречается k среди

$f(0), \ldots, f(n-1)$, и поделим результат на n. Полученное частное можно назвать *частотой k среди первых n членов* последовательности.

Теперь при фиксированном k рассмотрим *нижний предел* частоты k среди первых n членов при $n \to \infty$, который будем называть *нижней частотой* числа k в последовательности f.

Пусть p_k — нижняя частота числа k в данной последовательности. Легко видеть, что $\sum_k p_k \leqslant 1$. В самом деле, если какая-то конечная частичная сумма этого ряда была бы больше 1, то сумма соответствующих нижних пределов была бы больше 1, и потому сумма достаточно далёких допредельных значений превзошла бы 1, что невозможно (сумма частот в любом начальном отрезке не больше 1).

Кроме того, выполнено следующее утверждение (для любой вычислимой последовательности f и нижних частот p_k появления в ней различных чисел k):

Теорема 135. *Функция $k \mapsto p_k$ $\mathbf{0}'$-перечислима снизу.*

Определение перечислимой снизу функции мы давали в разделе 4.1; сейчас мы рассматриваем релятивизованный относительно $\mathbf{0}'$ вариант этого понятия.

◀ В самом деле, утверждение $r < p_k$ (где r — рациональное число) можно записать так:

> существует такое рациональное $p > r$ и такое N, что *частота появления k в любом начальном отрезке последовательности длины больше N превосходит p.*

Набранное курсивом свойство является коперечислимым (имеет перечислимое дополнение): если оно неверно, то это можно обнаружить, предъявив соответствующий начальный отрезок. Поэтому это свойство $\mathbf{0}'$-разрешимо (применим оракул к алгоритму поиска этого отрезка). А потому свойство $r < p_k$ в целом $\mathbf{0}'$-перечислимо. ▶

По существу мы тут используем такой общий факт:

$\boxed{\mathbf{210}}$ Пусть r_n — вычислимая последовательность рациональных чисел. Покажите, что число $\liminf r_n$ является $\mathbf{0}'$-перечислимым снизу и соответствующий $\mathbf{0}'$-алгоритм можно построить, зная алгоритм для r_n.

Отметим кстати, что верно и обратное утверждение:

$\boxed{\mathbf{211}}$ Всякое $\mathbf{0}'$-перечислимое снизу действительное число является нижним пределом вычислимой последовательности рациональных чисел. [Указание. Это число является точной верхней гранью $\mathbf{0}'$-вычислимой последовательности рациональных чисел r_n, каждое из которых является предельным значением стабилизирующейся последовательности $r_{n,k}$. Положим s_k равным максимуму из чисел $r_{0,k}, \ldots, r_{t-1,k}$, где t — минимальное число, для которого $r_{t,k} \neq r_{t,k-1}$.]

Оказывается, что для подходящей последовательности f функция $k \mapsto p_k$ является максимальной $\mathbf{0}'$-перечислимой снизу полумерой. Это вытекает из следующего утверждения:

Теорема 136. *Для всякой $\mathbf{0}'$-перечислимой снизу последовательности неотрицательных действительных чисел $q_0, q_1, \ldots$ с $\sum_i q_i \leqslant 1$ найдётся вычислимая*

последовательность $f(0), f(1), \ldots$, нижняя частота появления любого натурального k в которой не меньше q_k.

Это позволяет дать эквивалентное определение $\mathbf{0}'$-релятивизованной префиксной сложности числа k как минус логарифма нижней частоты числа k в оптимальной последовательности f (той, в которой нижние частоты максимальны с точностью до $O(1)$-множителя).

◀ Перечислимость снизу означает, что множество пар $\langle r, k\rangle$, где r — рациональное число, меньшее q_k, перечислимо (с оракулом $\mathbf{0}'$). Как известно из теории вычислимых функций (см., например, [175]), $\mathbf{0}'$-перечислимые множества составляют класс Σ_2, и потому можно найти разрешимое свойство R, для которого

$$r < q_k \ \Leftrightarrow\ \exists u\ \forall v\ \ R(r, k, u, v).$$

Нам будет удобно иметь дело с несколько другим представлением, а именно, мы выберем вычислимую всюду определённую функцию $\langle r, k, n\rangle \mapsto S(r, k, n)$ со значениями 0 и 1, для которой $r < q_k$ тогда и только тогда, когда в последовательности $S(r, k, 0)$, $S(r, k, 1)\ldots$ конечное число нулей. Последовательность $S(r, k, 0)$, $S(r, k, 1)\ldots$ можно построить так: ищем последовательно для каждого из $u = 0, 1, 2, \ldots$ значение v, при котором $R(r, k, u, v)$ ложно, одновременно дописывая в последовательность единицы; как только такое значение обнаруживается, перемежаем единицы нулём. Конечность числа нулей означает, что для некоторого u такого v не найдётся, то есть что $r < q_k$.

Удобна такая метафора: время от времени для некоторых пар $\langle r, k\rangle$ появляется запрос «хочу, чтобы q_k было больше r», которые время от времени могут «сбрасываться», одновременно возобновляясь. При этом для пар $\langle r, k\rangle$, у которых $r < q_k$, число сбросов конечно (и возобновлённый в какой-то момент запрос уже не будет сброшен), а для остальных пар ($r \geqslant q_k$) запросы сбрасываются бесконечно много раз. (Моменты сбросов запросов соответствуют нулям в последовательности S.) Весь этот процесс вычислим и в каждый момент имеется лишь конечное число запросов (так нам удобно считать, а роли это не играет, поскольку конечность числа нулей в последовательности не зависит от её начального отрезка).

Напомним, что нам нужна вычислимая последовательность $f(0), f(1), \ldots$, в которой нижняя частота появления числа k не меньше q_k.

Для этого достаточно представить исходную $\mathbf{0}'$-перечислимую снизу полумеру как нижний предел вычислимой последовательности мер с рациональными значениями, то есть построить вычислимую двумерную таблицу из рациональных чисел

$$\begin{array}{cccc}
p_0^0 & p_1^0 & p_2^0 & \ldots \\
p_0^1 & p_1^1 & p_2^1 & \ldots \\
p_0^2 & p_1^2 & p_2^2 & \ldots \\
\ldots & \ldots & \ldots & \ldots
\end{array}$$

с такими свойствами: в каждой строке лишь конечное число ненулевых элементов, в сумме равных единице, а нижний предел в k-столбце не меньше q_k. В самом деле, пусть дана такая таблица. Можно считать, что все числа в i-й строке кратны $1/i$

(заменив их на приближения, что не повлияет на предел). Теперь строим последовательность f так: вначале некоторое время руководствуемся первой строкой как таблицей частот, затем переходим ко второй строке и пользуемся ей гораздо большее время (так, чтобы вклад первой строки в частоты стал мал), затем переходим к третьей строке и следуем ей ещё дольше (чтобы забить вклад первой и второй строки) и так далее.

Итак, осталось построить таблицу с таким свойством: если запрос «сделать q_k больше r» с некоторого момента не сбрасывается, то в k-м столбце нижний предел не меньше q_k. Это делается так: строя n-ю строку таблицы в момент времени n, мы удовлетворяем все имеющиеся к данному моменту запросы (для всех k) в порядке их «стажа» (сколько времени этот запрос уже не сброшен), увеличивая соответствующие p_k до соответствующего r пока это возможно (пока сумма не превзошла единицу; можно считать, что запросов достаточно много и это рано или поздно случится, в этот момент мы обрезаем последний запрос, делая сумму равной единице, и завершаем построение n-й строки).

Почему это гарантирует выполнение условия на нижний предел? Предположим, что $r < q_k$ на самом деле. Тогда запрос «сделать q_k больше r» в некоторый момент появится и уже более никогда не будет сброшен. Посмотрим на запросы, которые появились до него. Некоторые из них будут сброшены, а некоторые — так и не будут. Дождёмся такого момента, когда все эти сбросы произойдут. Тогда среди запросов с большим стажем останутся лишь такие пары r', k', которые никогда не будут сброшены, и потому для них $r' < q_{k'}$ и потому сумма всех таких r' вместе с нашим r меньше единицы. Поэтому запросы с бóльшим стажем не помешают удовлетворить наш запрос, что и требовалось доказать. ▶

212 Покажите, что можно построить вычислимую последовательность, в которой нижние пределы частот в точности равны q_k. [Указание. Надо скомбинировать приведённую конструкцию с решением задачи 211.]

Ещё один результат из упомянутой статьи Мучника [114]:

213 Докажите, что теорема 136 останется верной, если разрешить рассматривать частичные вычислимые функции f из $\mathbb{N}$ в $\mathbb{N}$ как «последовательности с пробелами»: для любой вычислимой частичной функции f из $\mathbb{N}$ в $\mathbb{N}$ существует вычислимая последовательность $g(0), g(1), \ldots$ (без пробелов) с теми же или бóльшими нижними частотами: нижняя частота любого числа k в g не меньше его нижней частоты в $f(0), f(1), \ldots$ (определяемой как нижний предел количества появлений числа k среди $f(0), \ldots, f(N-1)$, делённого на N). [Указание [9]. Для каждого N частоты появления в начальном отрезке длины N образуют перечислимую снизу полумеру (вместо вычислимой меры для всюду определённых последовательностей); конструкция из доказательства теоремы 133 позволяет мажорировать нижний предел $\mathbf{0}'$-перечислимой снизу полумерой, от которой уже можно перейти ко всюду определённой функции.]

7. Шенноновская энтропия и колмогоровская сложность

7.1. Шенноновская энтропия

Пусть нам нужно закодировать буквы некоторого алфавита A, состоящего из k букв $a_1, \ldots, a_k$, двоичными словами (наподобие азбуки Морзе, только вместо точки и тире используются нуль и единица). Пусть буква a_i кодируется некоторым словом c_i. Естественно требовать, чтобы все слова c_i были различны. Но этого мало, если записывать коды подряд. Скажем, если алфавит состоит из букв А, Б и В, имеющих коды 0, 1 и 01, то мы не сможем отличить слово АБАБ от слова АБВ: в обоих случаях будет последовательность 0101. (В азбуке Морзе, кстати, такой проблемы нет, поскольку между отдельными буквами делается перерыв — больший, чем между точками и тире внутри буквы.) Поэтому надо отдельно позаботиться об однозначности декодирования.

Другой предмет заботы при построении кода — его экономность. Полезно выбирать кодовые слова c_i по возможности более короткими (насколько это возможно при сохранении однозначности декодирования). Более того, если не удаётся сделать короткими все кодовые слова, разумно в первую очередь позаботиться о наиболее часто встречающихся буквах. (Это обстоятельство учитывалось и при составлении азбуки Морзе.)

7.1.1. Коды

Перейдём к формальным определениям. *Кодом* для алфавита A, состоящего из k букв $a_1, \ldots, a_k$, называется набор из k двоичных слов $c_1, \ldots, c_k$. Они называются *кодовыми словами* данного кода; слово c_i называется *кодом* буквы a_i; всякое слово в алфавите A кодируется двоичным словом, получаемым соединением кодов соответствующих букв.

Будем называть код *инъективным*, если коды различных букв различны, и *однозначно декодируемым*, если коды любых двух различных слов различны. Код называется *префиксным*, если ни одно из кодовых слов (соответствующих буквам алфавита A) не является началом (префиксом) другого. (Это название стало традиционным, хотя более логичное — *беспрефиксный* код — также используется.)

Теорема 137. *Всякий префиксный код является однозначно декодируемым.*

◄ Первое кодовое слово (код первой буквы) отщепляется однозначно (в силу префиксности), затем отщепляется код второй буквы и т. п. ►

214 Покажите, что не всякий однозначно декодируемый код является префиксным. [Указание. Он может быть, например, «суффиксным».]

215 Укажите явно взаимно однозначное соответствие между множеством бесконечных последовательностей цифр $0, 1, 2$ и множеством бесконечных последовательностей нулей и единиц. [Указание. Используйте префиксный код $0 \mapsto 00$, $1 \mapsto 01$, $2 \mapsto 1$.]

216 Пусть слова $c_1, \ldots, c_k$ и $d_1, \ldots, d_l$ образуют префиксный код (по отдельности). Покажите, что kl слов $c_i d_j$ (приписываем одно слово к другому без разделителя) также образуют префиксный код.

Чтобы сравнивать коды по их экономности, нужно фиксировать частоты букв. Пусть даны неотрицательные числа $p_1, \ldots, p_k$, в сумме равные единице; число p_i будем называть *частотой* (или *вероятностью*) буквы a_i. Для каждого кода $c_1, \ldots, c_k$ (для букв $a_1, \ldots, a_k$) определим *среднюю длину* кода как сумму

$$\sum_i p_i l(c_i).$$

Возникает задача: для данных $p_1, \ldots, p_k$ найти код по возможности меньшей средней длины (в том или ином классе кодов).

217 Как найти минимальный (с точки зрения средней длины) инъективный код для данного набора $p_1, \ldots, p_k$? [Указание: все буквы надо упорядочить по убыванию частот, а все двоичные слова — в порядке возрастания длин, начиная с пустого.]

7.1.2. Определение шенноновской энтропии

Что можно сказать о минимально возможной длине префиксного кода для данных частот $p_1, \ldots, p_k$? Для ответа на этот вопрос полезна *шенноновская энтропия*. Она определяется для данных $p_1, \ldots, p_k$ (неотрицательных и в сумме равных единице) как величина

$$H = p_1(-\log p_1) + p_2(-\log p_2) + \ldots + p_k(-\log p_k)$$

(при $p = 0$ мы полагаем $p \log p = 0$, доопределяя тем самым функцию $p \mapsto p \log p$ по непрерывности).

Мотивировка этой формулы такова: буква a_i появляется с частотой p_i, а каждое её появление несёт $-\log p_i$ битов информации; в среднем получается H битов на букву. Нужно только объяснить, почему мы считаем, что появление буквы с вероятностью p несёт $-\log p$ битов информации. Это можно сделать так. Пусть задумано одно из 2^n равновозможных чисел. Чтобы его отгадать, надо задать n вопросов (типа да/нет), каждый из которых даёт 1 бит информации. Значит, вероятность $1/2^n$ соответствует n битам информации в событии.

Конечно, последний абзац — это всего лишь метафорический комментарий, позволяющий легче запомнить формулу. Зато следующее утверждение является вполне точным.

Пусть фиксированы неотрицательные числа $p_1, \ldots, p_k$, в сумме равные единице.

Теорема 138. (а) *Для любого префиксного кода с кодовыми словами $c_1, \ldots, c_k$ выполняется неравенство*

$$\sum_i p_i l(c_i) \geqslant H$$

(*средняя длина любого кода не меньше энтропии*).

(б) *Существует префиксный код, для которого*

$$\sum_i p_i l(c_i) < H + 1.$$

◄ Заметим, что в этой теореме реально фигурируют не сами кодовые слова, а их длины. Поэтому важно знать, какие наборы чисел могут быть длинами кодовых слов префиксного кода. Ответ даёт такая лемма:

Лемма (неравенство Крафта). *Пусть фиксированы целые неотрицательные числа $n_1, \ldots, n_k$ и требуется найти двоичные слова $c_1, \ldots, c_k$ указанных длин ($l(c_i) = n_i$), причём так, чтобы ни одно из этих слов не было началом другого. Это возможно тогда и только тогда, когда $\sum_i 2^{-n_i} \leqslant 1$.*

Это утверждение нам уже встречалось, см. леммы в доказательстве теорем 56 (с. 107) и 58 (с. 108). В одну сторону: если ни одно из слов c_i не является началом другого, то соответствующие интервалы длин 2^{-n_i} не пересекаются, и потому сумма их длин не превосходит единицы. (В других терминах: случайная последовательность нулей и единиц начинается на c_i с вероятностью 2^{-n_i}; эти события несовместны, так как слова несравнимы, и потому сумма вероятностей не превосходит единицы.)

В обратную сторону можно воспользоваться даже более простым способом, чем при доказательстве теоремы 58, поскольку число слов конечно их длины заранее известны. Нужно просто выкладывать соответствующие интервалы длин 2^{-n_i} слева направо на отрезке $[0, 1]$, причём делать это в порядке убывания длин. Тогда каждый интервал будет правильно «выровнен» и ему будет соответствовать двоичное слово длины n_i.

Вернёмся к доказательству теоремы. Можно считать, что все p_i положительны, поскольку нулевые p_i не дают вклада ни в среднюю длину кода, ни в энтропию. В пункте (а) нам надо доказать, что если n_i — неотрицательные целые числа и $\sum_i 2^{-n_i} \leqslant 1$, то сумма $\sum p_i n_i$ не меньше шенноновской энтропии H. Это удобнее доказывать сразу для произвольных n_i (не обязательно целых) и перейдя к другим координатам. Обозначим через q_i величину 2^{-n_i}. В этих координатах утверждение таково: если $q_i > 0$ и $\sum q_i \leqslant 1$, то

$$\sum p_i(-\log q_i) \geqslant \sum p_i(-\log p_i).$$

Это неравенство иногда называют *неравенством Гиббса*. Чтобы доказать его, заметим, что разница между правой и левой частью равна

$$\sum_i p_i \log \frac{q_i}{p_i}$$

и в силу выпуклости логарифма (взвешенная сумма логарифмов не превосходит логарифма взвешенной суммы: $\sum p_i \log u_i \leqslant \log\bigl(\sum_i p_i u_i\bigr)$; это верно для любых положительных u_i) не превосходит

$$\log\Bigl(\sum_i p_i \frac{q_i}{p_i}\Bigr) = \log\Bigl(\sum q_i\Bigr) \leqslant \log 1 = 0.$$

Утверждение (а) доказано.

Отметим кстати, что неотрицательную величину

$$\sum_i p_i \log \frac{p_i}{q_i}$$

называют *расстоянием Кульбака – Лейблера* (Kullback – Leibler distance) между распределениями вероятностей p_i и q_i (при этом предполагается, что $\sum q_i = 1$), хотя это «расстояние» и не симметрично. Выпуклость логарифма (отрицательность второй производной) гарантирует, что расстояние Кульбака – Лейблера неотрицательно и обращается в нуль, лишь если $p_i = q_i$ при всех i.

Чтобы доказать утверждение (б), рассмотрим числа $n_i = \lceil -\log_2 p_i \rceil$ (где $\lceil u \rceil$ обозначает наименьшее целое число, большее или равное u). Тогда

$$\frac{p_i}{2} < 2^{-n_i} \leqslant p_i.$$

Неравенство $2^{-n_i} \leqslant p_i$ гарантирует, что выполнены условия леммы (и потому можно найти кодовые слова соответствующих длин). Неравенство $p_i/2 < 2^{-n_i}$ означает, что n_i превосходит $(-\log p_i)$ менее чем на 1, что сохраняется и после усреднения: средняя длина кода $(\sum p_i n_i)$ превосходит $H = \sum p_i(-\log p_i)$ менее чем на 1. ►

Кратко доказательство теоремы можно резюмировать так: если забыть, что длины кодовых слов должны быть целыми, и разрешать любые числа n_i, только бы сумма 2^{-n_i} не превышала единицы, то выгоднее всего взять $n_i = -\log p_i$ (следует из выпуклости логарифма). Требование же целочисленности приводит к увеличению n_i, но не более чем на единицу.

Теорема 139. *Энтропия распределения $p_1, \ldots, p_n$ с n значениями не превосходит $\log n$ и равна $\log n$ в единственном случае, когда все p_i равны.*

◄ Если n есть степень двойки, то неравенство $H \leqslant \log n$ прямо следует из теоремы 138, поскольку можно рассмотреть префиксный код, в котором n кодовых слов имеют длину $\log n$. В общем случае надо применить неравенство Гиббса с $q_i = 1/n$ при всех i и вспомнить, что это неравенство обращается в равенство при $p_i = q_i$. ►

7.1.3. Код Хаффмана

Мы доказали, что средняя длина оптимального префиксного кода (для данных $p_1, \ldots, p_k$) заключена между H и $H + 1$. Попробуем разобраться, как найти этот код.

Пусть $n_1, \ldots, n_k$ — длины кодовых слов оптимального кода для данных $p_1, \ldots$ $\ldots, p_k$. Будем предполагать (переставив буквы), что

$$p_1 \leqslant p_2 \leqslant \ldots \leqslant p_k.$$

Можно считать также, что

$$n_1 \geqslant n_2 \geqslant \ldots \geqslant n_k$$

(если бы более частая буква кодировалась длиннее, чем более редкая, то обмен кодов уменьшил бы среднюю длину; буквы с одинаковой частотой можно переставлять, не меняя средней длины кода).

Заметим, что для оптимального кода $n_1 = n_2$ (две наиболее редкие буквы всегда имеют одну и ту же длину кода). В самом деле, если $n_1 > n_2$, то n_1 больше всех остальных n_i. Поэтому в сумме $\sum_i 2^{-n_i}$ первое слагаемое меньше всех других, неравенство $\sum_i 2^{-n_i} \leqslant 1$ не может обратиться в равенство по соображениям чётности, и левая часть его меньше правой по крайней мере на 2^{-n_1}. А значит, n_1 можно уменьшить на единицу, не нарушив неравенства $\sum_i 2^{-n_i} \leqslant 1$, и код не является оптимальным.

Поэтому при выборе оптимального кода достаточно ограничиться кодами с $n_1 = n_2$, и оптимальный среди них соответствует минимуму выражения

$$p_1 n_1 + p_2 n_2 + p_3 n_3 + \ldots + p_k n_k = (p_1 + p_2)n + p_3 n_3 + \ldots + p_k n_k$$

(если через n обозначить общее значение n_1 и n_2) по всем $n, n_3, \ldots, n_k$, для которых

$$2^{-n} + 2^{-n} + 2^{-n_3} + \ldots + 2^{-n_k} \leqslant 1.$$

Перепишем это неравенство как

$$2^{-(n-1)} + 2^{-n_3} + \ldots + 2^{-n_k} \leqslant 1,$$

а выражение, подлежащее минимизации, как

$$(p_1 + p_2) + (p_1 + p_2)(n - 1) + p_3 n_3 + \ldots + p_k n_k.$$

Член $(p_1 + p_2)$ постоянен и не влияет на поиск минимума, а остаток выражения в точности соответствует задаче поиска оптимального префиксного кода для $k - 1$ букв с вероятностями $p_1 + p_2, p_3, \ldots, p_k$.

Получаем рекурсивный алгоритм: соединить две наиболее редкие буквы в одну (сложив вероятности), найти оптимальный префиксный код для этого случая (рекурсивный вызов), а потом вместо одного кодового слова x для соединённой буквы взять два кодовых слова на единицу длиннее ($x0$ и $x1$); ясно, что префиксность кода при этом не нарушится.

Построенный с помощью такого алгоритма оптимальный префиксный код называется *кодом Хаффмана* для данного набора вероятностей p_i.

7.1.4. Неравенство Крафта – Макмиллана

До сих пор мы изучали в основном префиксные коды. Оказывается, что переход к произвольным однозначно декодируемым кодам ничего не даёт (с точки зрения сокращения кода):

Теорема 140 (неравенство Макмиллана). *Пусть $c_1, \ldots, c_k$ — кодовые слова однозначно декодируемого кода, а $n_i = l(c_i)$ — их длины. Тогда*

$$\sum_i 2^{-n_i} \leqslant 1.$$

Тем самым (лемма Крафта) можно построить и префиксный код с теми же длинами слов.

◀ Будем считать, что в кодовых словах вместо цифр 0 и 1 используются буквы u и v. (Скажем, кодовые слова 0, 01 и 11 мы запишем как u, uv и vv.) Напишем формальную сумму $(c_1 + \ldots + c_k)$ всех кодовых слов, возведём её в N-ю степень (число N мы потом выберем) и раскроем скобки, не переставляя u и v (как если бы они не коммутировали). Например, для $N = 2$ и для приведённого выше примера получится

$$(u + uv + vv)(u + uv + vv) = \\ = uu + uuv + uvv + uvu + uvuv + uvvv + vvu + vvuv + vvvv.$$

Каждое слагаемое в правой части есть соединение некоторых кодовых слов, причём все слагаемые различны (свойство однозначности декодирования). Теперь подставим вместо u и v число $1/2$. В левой части $(c_1 + \ldots + c_k)^N$ превратится при этом в $(2^{-n_1} + \ldots + 2^{-n_k})^N$. Правую часть оценим сверху: если бы в неё входили все возможные слова данной длины t, то получилось бы 2^t членов, каждый из которых равен 2^{-t}, и сумма равнялась бы 1 (для каждой длины). Поэтому сумма в правой части не превосходит максимальной длины слагаемых, то есть, не больше $N \max(n_i)$.

Теперь видно, что если $\sum 2^{-n_i} > 1$, то при больших N левая часть (растущая экспоненциально) становится больше правой (растущей линейно). ▶

Это доказательство производит впечатление искусственного (хотя и красивого) трюка. Более естественное доказательство (или, если угодно, более естественное изложение того же доказательства) будет приведено ниже (с. 249).

7.2. Энтропия пары и условная энтропия

7.2.1. Энтропия пары случайных величин

При обсуждении энтропии удобно использовать стандартную для теории вероятностей терминологию. Пусть ξ — случайная величина, принимающая конечное число значений $\xi_1, \ldots, \xi_k$ с вероятностями $p_1, \ldots, p_k$. Тогда её *шенноновская энтропия* определяется формулой

$$H(\xi) = p_1(-\log p_1) + \ldots + p_k(-\log p_k).$$

Это определение позволяет говорить об энтропии пары случайных величин ξ и η (определённых на одном и том же вероятностном пространстве), поскольку такая пара сама образует случайную величину. Следующая теорема утверждает, что энтропия пары не превосходит суммы энтропий:

Теорема 141.
$$H(\langle \xi, \eta \rangle) \leqslant H(\xi) + H(\eta).$$

Мы предполагаем, что величины ξ и η принимают конечное число значений, поэтому эта теорема представляет собой некоторое неравенство с суммами логарифмов. Пусть ξ принимает k значений $\xi_1, \ldots, \xi_k$, а η принимает l значений $\eta_1, \ldots, \eta_l$. Тогда величина $\langle \xi, \eta \rangle$ может принимать, вообще говоря, kl значений $\langle \xi_i, \eta_j \rangle$ (некоторые из значений могут не встречаться или встречаться с вероятностью нуль). Распределение вероятностей для пары $\langle \xi, \eta \rangle$, таким образом, задаётся таблицей из k строк и l столбцов: число p_{ij}, стоящее в i-й строке и j-м столбце, представляет собой вероятность события «$(\xi = \xi_i)$ и $(\eta = \eta_j)$» (здесь $i = 1, \ldots, k$ и $j = 1, \ldots, l$). Все числа p_{ij} неотрицательны и в сумме равны единице (некоторые из них могут равняться нулю).

Сложив числа в строках, мы получим распределение вероятностей для величины ξ: она принимает значение ξ_i с вероятностью $\sum_j p_{ij}$; эту сумму удобно обозначить p_{i*}; аналогичным образом η принимает значение η_j с вероятностью p_{*j}, которая есть сумма чисел в j-м столбце.

Таким образом, сформулированная теорема представляет собой неравенство, справедливое для любой прямоугольной таблицы с неотрицательными числами, в сумме равными единице:
$$\sum_{i,j} p_{ij}(-\log p_{ij}) \leqslant \sum_i p_{i*}(-\log p_{i*}) + \sum_j p_{*j}(-\log p_{*j})$$
(где p_{i*} и p_{*j} определяются как суммы по строкам и столбцам).

Это неравенство в конечном счёте сводится к выпуклости логарифма, но полезно понимать его интуитивный смысл. Если отождествить (забыв про разницу порядка единицы) энтропию с длиной кратчайшего префиксного кода, то теорему можно доказать так: пусть имеются короткие префиксные коды для ξ и η (с кодовыми словами $c_1, \ldots, c_k$ и $d_1, \ldots, d_l$). Тогда можно рассмотреть код для пары $\langle \xi, \eta \rangle$, кодируя значения $\langle \xi_i, \eta_j \rangle$ словом $c_i d_j$ (приписываем d_j справа к c_i без разделителя). Это будет, как легко проверить, префиксный код (чтобы отщепить кодовое слово от бесконечной последовательности, надо сначала отщепить c_i, а потом d_j; в обоих случаях это делается однозначно). Средняя длина этого кода будет равна сумме средних длин кодов для ξ и η. Он не обязан быть оптимальным (ведь и неравенство может быть строгим), но даёт оценку сверху для оптимального кода.

◄ Это рассуждение можно превратить в строгое доказательство, если вспомнить, что при доказательстве теоремы 138 (с. 241) мы установили, что энтропия равна минимуму величины $\sum_i p_i(-\log_2 q_i)$ по всем наборам неотрицательных чисел q_i с единичной суммой. В частности, энтропия пары (левая часть неравенства) есть минимум сумм
$$\sum_{i,j} p_{ij}(-\log q_{ij})$$

по всем наборам q_{ij} неотрицательных чисел с суммой единица. Будем рассматривать не все наборы, а лишь наборы «ранга 1», которые получаются как произведения

$$q_{ij} = q_{i*} \cdot q_{*j}$$

для некоторых наборов неотрицательных чисел q_{i*} и q_{*j}, каждый из которых имеет сумму 1. Тогда $(-\log q_{ij})$ распадётся в сумму $(-\log q_{i*}) + (-\log q_{*j})$, а вся сумма — в две суммы, которые (после суммирования по одному из индексов) окажутся равными

$$\sum_i p_{i*}(-\log q_{i*})$$

и

$$\sum_j p_{*j}(-\log q_{*j})$$

соответственно. Минимумы этих сумм равны $H(\xi)$ и $H(\eta)$.

Таким образом, левая часть неравенства есть минимум некоторой величины по всем наборам, а правая — по наборам ранга 1, откуда и вытекает требуемое неравенство. ▶

7.2.2. Условная энтропия

Условной вероятностью некоторого события B при условии события A называют отношение вероятности события «A и B» к вероятности события A. Это определение имеет смысл, если вероятность события A отлична от нуля. Мотивировка понятна: мы рассматриваем долю исходов, где произошло B, не среди всех исходов, а только среди тех, где произошло A.

Если A — событие, а ξ — случайная величина с конечным числом значений $\xi_1, \ldots, \xi_k$, то можно рассмотреть (помимо вероятностей $\Pr[\xi = \xi_i]$) и условные вероятности $\Pr[(\xi = \xi_i) \mid A]$. Их сумма тоже равна единице, и получается некоторое новое распределение вероятностей. Его энтропия называется *условной энтропией величины ξ при условии A* и обозначается $H(\xi \mid A)$, а само это распределение вероятностей можно обозначить $(\xi \mid A)$.

218 Покажите, что величина $H(\xi \mid A)$ может быть и больше, и меньше величины $H(\xi)$. [Указание: распределение $(\xi \mid A)$ (особенно при малой вероятности события A) мало связано с распределением для ξ.]

Неформально говоря, $H(\xi \mid A)$ — это минимально возможная средняя длина кода, если нас интересуют лишь случаи, когда произошло событие A.

Пусть теперь (как и в прошлом разделе) даны две случайные величины ξ и η. Будем предполагать, что для каждой из них все значения имеют ненулевую вероятность (нулевые можно выбросить). Для каждого значения η_j величины η рассмотрим событие $\eta = \eta_j$ (его вероятность мы обозначали p_{*j}). Рассмотрим условную энтропию величины ξ при условии этого события. Она соответствует распределению вероятностей $i \mapsto p_{ij}/p_{*j}$. Далее усредним эти энтропии с весами, равными

вероятностям событий $\eta = \eta_j$. Полученное среднее называют *условной энтропией ξ при известном η* и обозначают $H(\xi \mid \eta)$. Формально говоря,

$$H(\xi \mid \eta) = \sum_j \Pr[\eta = \eta_j] H(\xi \mid \eta = \eta_j)$$

или, в наших обозначениях,

$$H(\xi \mid \eta) = \sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}} \left(-\log \frac{p_{ij}}{p_{*j}}\right).$$

Основные свойства условной энтропии перечислены в следующей теореме, справедливой для любых случайных величин ξ, η:

Теорема 142. (а) $H(\xi \mid \eta) \geqslant 0$;

(б) $H(\xi \mid \eta) = 0$ *тогда и только тогда, когда* $\xi = f(\eta)$ *с вероятностью* 1 *для некоторой функции* f (*оговорка про «вероятность* 1*» означает, что мы пренебрегаем значениями, которые имеют нулевую вероятность*);

(в) $H(\xi \mid \eta) \leqslant H(\xi)$;

(г) $H(\langle \xi, \eta \rangle) = H(\eta) + H(\xi \mid \eta)$.

◀ Первое из этих утверждений очевидно: все $H(\xi \mid \eta = \eta_j)$ неотрицательны, потому неотрицательна и их взвешенная сумма.

(б) Если взвешенная сумма равна нулю, то все слагаемые с ненулевыми коэффициентами равны нулю, то есть при каждом значении η_j величина $(\xi \mid \eta = \eta_j)$ имеет нулевую энтропию (и потому принимает лишь одно значение с точностью до событий нулевой вероятности).

Утверждение (в) можно объяснить так: $H(\xi \mid \eta)$ будет средней длиной оптимального кода для ξ, если разрешить кодировать значения ξ по-разному, в зависимости от значения величины η (в каждом случае свой код, который оптимизируется с учётом условных вероятностей). Ясно, что это облегчает построение оптимального кода, поэтому средняя длина получается меньше $H(\xi)$.

Более формально: при каждом j величина $H(\xi \mid \eta = \eta_j)$ равна минимуму суммы

$$\sum_i \frac{p_{ij}}{p_{*j}} (-\log q_{ij})$$

по всем неотрицательным $q_{1j} + q_{2j} + \ldots + q_{kj} = 1$ (мы используем свой набор переменных для каждого j) и потому $H(\xi \mid \eta)$ равна минимуму суммы

$$\sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}} (-\log q_{ij})$$

по всем таблицам, составленным из неотрицательных чисел q_{ij}, у которых сумма каждого столбца равна единице. Если мы теперь ограничимся таблицами, у которых все столбцы одинаковы, $q_{ij} = q_i$, то сумма превратится в

$$\sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}} (-\log q_i) = \sum_j \sum_i p_{ij} (-\log q_i) = \sum_i p_{i*} (-\log q_i)$$

и её минимум станет равным $H(\xi)$. Поэтому $H(\xi \mid \eta) \leqslant H(\xi)$.

Наконец, пункт (г) представляет собой равенство, которое непосредственно следует из определений:

$$\sum_{i,j} p_{ij}(-\log p_{ij}) = \sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}} \Bigl(-\log \frac{p_{ij}}{p_{*j}} - \log p_{*j}\Bigr) =$$
$$= \sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}} \Bigl(-\log \frac{p_{ij}}{p_{*j}}\Bigr) + \sum_j p_{*j} \sum_i \frac{p_{ij}}{p_{*j}} (-\log p_{*j}) =$$
$$= \sum_j p_{*j} H(\xi \,|\, \eta = \eta_j) + \sum_j p_{*j}(-\log p_{*j}) = H(\xi \,|\, \eta) + H(\eta).$$

Теорема доказана. ►

Из этой теоремы немедленно вытекает теорема 141 (с. 245). Кроме того, из неё видно, что энтропия пары не меньше энтропии любого из её членов (поскольку условная энтропия неотрицательна). Отсюда легко вытекает такое утверждение:

Теорема 143. *Пусть ξ — случайная величина с конечным числом значений, а f — функция, определённая на множестве значений величины ξ. Тогда*

$$H(f(\xi)) \leqslant H(\xi),$$

где $f(\xi)$ — случайная величина, получающаяся применением f к ξ (формально говоря, композиция f и ξ).

С точки зрения наборов чисел переход от ξ к $f(\xi)$ означает, что мы объединяем некоторые значения (складывая соответствующие вероятности).

◄ В самом деле, величина $\langle \xi, f(\xi) \rangle$ имеет в точности то же распределение, что и ξ, поэтому её энтропия не меньше энтропии второго члена пары. ►

219 Укажите прямое доказательство в терминах кодирования и поиска минимума.

220 В каких случаях неравенство теоремы 143 обращается в равенство?

7.2.3. Независимость и энтропия

Понятие независимости случайных величин легко выражается в терминах энтропии. Напомним, что величины ξ и η *независимы*, если вероятность события «$\xi = \xi_i$ и $\eta = \eta_j$» равна произведению вероятностей событий $\xi = \xi_i$ и $\eta = \eta_j$ по отдельности. (Переформулировка: если распределение вероятностей ξ относительно любого из условий $\eta = \eta_j$ совпадает с исходным; аналогично и для η относительно $\xi = \xi_i$.) В наших обозначениях независимость означает, что $p_{ij} = p_{i*}p_{*j}$ (матрица вероятностей имеет ранг 1).

Теорема 144. *Величины ξ и η независимы тогда и только тогда, когда*

$$H(\langle \xi, \eta \rangle) = H(\xi) + H(\eta).$$

Другими словами, критерий независимости состоит в том, что неравенство теоремы 141 обращается в равенство. Используя теорему 142, можно переписать это равенство в виде $H(\xi) = H(\xi \,|\, \eta)$ или $H(\eta) = H(\eta \,|\, \xi)$.

◄ Логарифм является строго выпуклой функцией: неравенство

$$\log\Bigl(\sum p_i x_i\Bigr) \geqslant \sum p_i \log x_i ,$$

справедливое для неотрицательных p_i с единичной суммой и произвольных положительных x_i, обращается в равенство, лишь если все x_i равны (за исключением тех, которые входят с нулевыми коэффициентами p_i).

Отсюда следует, что для любых неотрицательных p_i, в сумме равных единице, минимум выражения

$$\sum p_i(-\log q_i),$$

который берётся по всем неотрицательным q_i, в сумме равным 1, достигается в единственной точке, когда $q_i = p_i$. (Надо уточнить, что при $p_i = 0$ мы полагаем $p_i(-\log q_i) = 0$ и разрешаем q_i быть нулевым.)

Теперь вспомним, что делалось при доказательстве теоремы 141. Минимум по матрицам ранга 1 (при котором правая часть равна сумме энтропий) достигался при

$$q_{ij} = p_{i*} \cdot p_{*j}.$$

Если он совпадает с минимумом по всем наборам q_{ij}, который достигается при $q_{ij} = p_{ij}$, то это значит, что имеет место равенство

$$p_{ij} = p_{i*} \cdot p_{*j}$$

и величины ξ и η независимы. ►

221 Проведите аналогичное рассуждение, основываясь на теореме 142.

222 Докажите, что величины α, β, γ *независимы в совокупности* (это значит, что вероятность события $(\alpha = \alpha_i, \beta = \beta_j, \gamma = \gamma_k)$ равна произведению трёх отдельных вероятностей) тогда и только тогда, когда

$$H(\langle \alpha, \beta, \gamma \rangle) = H(\alpha) + H(\beta) + H(\gamma).$$

Теоремы 141 и 144 показывают, что разность $H(\xi) + H(\eta) - H(\langle \xi, \eta \rangle)$ всегда неотрицательна и обращается в нуль тогда и только тогда, когда величины ξ и η независимы. Тем самым логично считать её числовой мерой «степени зависимости» между ξ и η. Эту разность обозначают $I(\xi:\eta)$ и называют *взаимной информацией* случайных величин ξ и η. Теорема 142 позволяет переписать выражение для $I(\xi:\eta)$ так:

$$I(\xi:\eta) = H(\eta) - H(\eta\,|\,\xi) = H(\xi) - H(\xi\,|\,\eta)$$

(взаимная информация показывает, «насколько знание одной величины уменьшает энтропию другой»).

В качестве применения рассмотренных понятий снова докажем неравенство Макмиллана. Мы будем действовать немного в другом порядке и сначала докажем, что для однозначно декодируемого кода случайной величины ξ средняя длина кодового слова не меньше $H(\xi)$.

Чтобы сделать это, заметим, что для инъективного (не обязательно однозначно декодируемого) кода, у которого все кодовые слова имеют длину меньше c, средняя длина не меньше $H(\xi) - \log c$. В самом деле, если n_i — длины кодовых слов такого кода, то сумма величин 2^{-n_i} не превосходит c (для каждой длины сумма по словам этой длины не превосходит единицы), и потому неравенство теоремы 138 ухудшается не более чем на $\log c$.

Теперь рассмотрим N независимых одинаково распределённых копий случайной величины ξ. Возникающая случайная величина, которую мы обозначим ξ^N, имеет энтропию в N раз больше. Если использовать для кодирования каждой из копий наш однозначно декодируемый код, а потом все слова соединить, то получится код для ξ^N, средняя длина которого в N раз больше средней длины исходного кода для ξ. Этот код будет инъективным (по определению однозначного декодирования), и максимальная длина его не превосходит cN, где c — максимальная длина кодовых слов рассматриваемого нами однозначно декодируемого кода. Применяя утверждение предыдущего абзаца, получаем, что

$$N \cdot (\text{средняя длина однозначно декодируемого кода}) \geqslant NH(\xi) - \log(cN).$$

Теперь, поделив на N и устремив N к бесконечности, получаем искомое утверждение (поскольку $\log(cN)/N$ стремится к 0).

Теперь уже легко доказать собственно неравенство Макмиллана. Пусть однозначно декодируемый код состоит из слов длиной $n_1, \ldots, n_k$, и при этом $\sum 2^{-n_i} > 1$. Положим сначала $p_i = 2^{-n_i}$, а затем пропорционально уменьшим все p_i, сделав их сумму равной единице. Рассмотрим случайную величину с распределением вероятностей p_i и её кодирование с помощью нашего кода. Средняя длина кода будет $\sum p_i n_i$ и будет меньше $H = \sum p_i(-\log p_i)$, поскольку $n_i < -\log p_i$ за счёт уменьшения величин p_i.

$\boxed{\mathbf{223}}$ Проведите подробно это доказательство неравенства Макмиллана и установите его соответствие с ранее приведённым.

7.2.4. «Релятивизация» и базисные неравенства

Доказанные нами утверждения имеют свои «условные» варианты. Например, неравенство

$$H(\langle\xi,\eta\rangle) \leqslant H(\xi) + H(\eta)$$

при добавлении случайной величины α в качестве условия превращается в

$$H(\langle\xi,\eta\rangle \,|\, \alpha) \leqslant H(\xi \,|\, \alpha) + H(\eta \,|\, \alpha).$$

Нового доказательства по существу не требуется, так как при каждом значении α_i случайной величины α выполнено неравенство

$$H(\langle\xi,\eta\rangle \,|\, \alpha = \alpha_i) \leqslant H(\xi \,|\, \alpha = \alpha_i) + H(\eta \,|\, \alpha = \alpha_i)$$

(применяем теорему 141 к условным распределениям вероятностей случайных величин ξ и η), и остаётся сложить эти неравенства с весами $\Pr[\alpha = \alpha_i]$.

Теперь можно выразить условные энтропии через безусловные, используя формулу $H(\beta\,|\,\gamma) = H(\langle\beta,\gamma\rangle) - H(\gamma)$, и привести подобные члены. Получится такое утверждение:

Теорема 145 (базисное неравенство).

$$H(\xi,\eta,\alpha) + H(\alpha) \leqslant H(\xi,\alpha) + H(\eta,\alpha).$$

Для краткости мы здесь опускаем угловые скобки и пишем $H(\xi,\eta,\alpha)$ вместо $H(\langle\xi,\eta,\alpha\rangle)$ или ещё более подробного $H(\langle\langle\xi,\eta\rangle,\alpha\rangle)$.

Аналогичную «релятивизацию» (добавление случайных величин как условий) можно применить и к понятию взаимной информации. Например, $I(\alpha:\beta\,|\,\gamma)$ определяется как

$$H(\alpha\,|\,\gamma) + H(\beta\,|\,\gamma) - H(\langle\alpha,\beta\rangle\,|\,\gamma).$$

Базисное неравенство (теорема 145) утверждает, что $I(\alpha:\beta\,|\,\gamma) \geqslant 0$ для любых случайных величин α,β,γ.

$\boxed{\textbf{224}}$ Докажите, что $I(\langle\alpha,\beta\rangle:\gamma) \geqslant I(\alpha:\gamma)$.

$\boxed{\textbf{225}}$ Докажите, что

$$I(\langle\alpha,\beta\rangle:\gamma) = I(\alpha:\gamma) + I(\beta:\gamma\,|\,\alpha).$$

Если $I(\alpha:\gamma\,|\,\beta) = 0$, говорят, что α и γ ***независимы при известном*** β, а также что α,β,γ образуют ***марковскую цепь*** («прошлое» α связано с «будущим» γ лишь через «настоящее» β).

$\boxed{\textbf{226}}$ Докажите, что в этом случае $I(\alpha:\gamma) \leqslant I(\alpha:\beta)$, и потому выполняется неравенство $I(\alpha:\gamma) \leqslant H(\beta)$.

При решении этих задач полезно использовать диаграммы, аналогичные диаграммам для колмогоровской сложности в главе 2. Диаграмма для двух величин состоит из трёх областей, каждой из которых соответствует неотрицательное значение; суммы значений в двух областях слева равна $H(\alpha)$, а в двух областях справа — $H(\beta)$ (рис. 7.1).

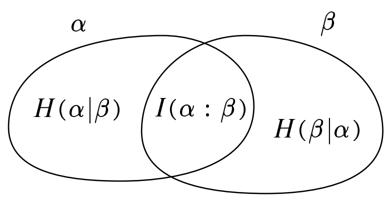


Рис. 7.1. Энтропии двух случайных величин.

Диаграмма для трёх величин α,β,γ показана на рис. 7.2. Центральной области соответствует значение, которое мы обозначили через $I(\alpha:\beta:\gamma)$; его можно определить как $I(\alpha:\beta) - I(\alpha:\beta\,|\,\gamma)$, а также как $I(\alpha:\gamma) - I(\alpha:\gamma\,|\,\beta)$ и т.п. — при переходе к безусловным энтропиям получается одно и то же выражение

$$I(\alpha:\beta:\gamma) = H(\alpha) + H(\beta) + H(\gamma) - H(\alpha,\beta) - H(\alpha,\gamma) - H(\beta,\gamma) + H(\alpha,\beta,\gamma).$$

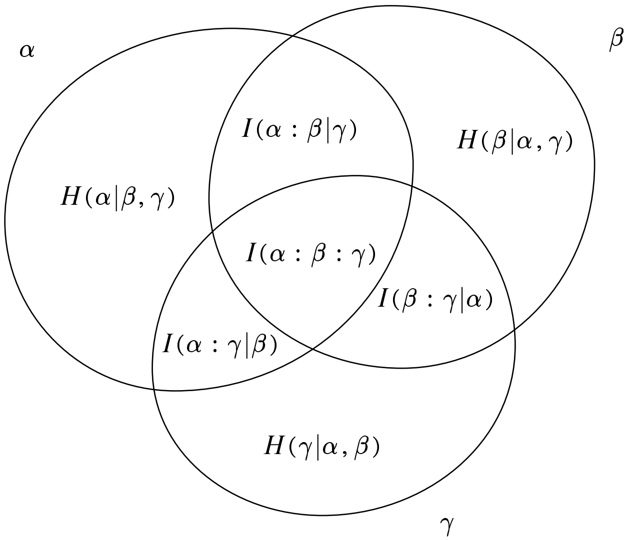


Рис. 7.2. Энтропии трёх случайных величин.

В отличие от шести других величин на рисунке, величина $I(\alpha:\beta:\gamma)$ может быть отрицательной. Так будет, например, если величины α и β независимы, но зависимы при известном γ.

227 Приведите пример таких величин α, β, γ. [Указание. По аналогии с примером на с. 61 можно рассмотреть независимые величины α и β, равномерно распределённые на $\{0, 1\}$, и положить $\gamma = (\alpha + \beta) \bmod 2$.]

228 Докажите, что если величины α и β различаются с вероятностью $\varepsilon < 1/2$, и число различных значений величины α равно a, то выполняется *неравенство Фано*:

$$H(\alpha\,|\,\beta) \leqslant \varepsilon \log a + h(\varepsilon),$$

где $h(\varepsilon)$ — энтропия случайной величины с двумя значениями, имеющими вероятности ε и $1 - \varepsilon$. [Указание. Введём величину γ, которая принимает два значения 0 и 1 при $\alpha \neq \beta$ и $\alpha = \beta$ соответственно. Тогда $H(\alpha\,|\,\beta) \leqslant H(\alpha, \gamma\,|\,\beta) = H(\gamma\,|\,\beta) + H(\alpha\,|\,\beta, \gamma) \leqslant H(\gamma) + H(\alpha\,|\,\beta, \gamma)$. Первое слагаемое равно $h(\varepsilon)$, а второе надо записать как

$$\Pr[\gamma = 0] H((\alpha\,|\,\beta)\,|\,\gamma = 0) + \Pr[\gamma = 1] H((\alpha\,|\,\beta)\,|\,\gamma = 1),$$

то есть

$$\Pr[\alpha \neq \beta] H((\alpha\,|\,\beta)\,|\,\alpha \neq \beta) + \Pr[\alpha = \beta] H((\alpha\,|\,\beta)\,|\,\alpha = \beta),$$

что не превосходит $\varepsilon \log a + 0$.]

229 Пусть $H(\alpha\,|\,\beta, \gamma) = 0$ и $I(\beta:\alpha) = 0$. Докажите, что $H(\gamma) \geqslant H(\alpha)$. (Если агент хочет передать в Центр секретное сообщение α в виде открытого текста β

с помощью заранее согласованного с Центром ключа γ, причём так, чтобы враги, не знающие γ, не получили никакой информации об α, то энтропия ключа должна быть не меньше энтропии сообщения. Это утверждение называют иногда *теоремой Шеннона об идеальном шифре.*)

230 Докажите, что для любых трёх случайных величин α, β, γ выполнено неравенство

$$2H(\alpha, \beta, \gamma) \leqslant H(\alpha, \beta) + H(\beta, \gamma) + H(\alpha, \gamma).$$

[Указание: см. доказательство аналогичного утверждения для колмогоровской сложности в теореме 26 (с. 59).]

231 Докажите аналогичное неравенство для любого числа случайных величин:

$$(n-1)H(\alpha_1, \ldots, \alpha_n) \leqslant H(\alpha_2, \ldots, \alpha_n) + \ldots + H(\alpha_1, \ldots, \alpha_{n-1}).$$

(В правой части неравенства складываются энтропии кортежей, состоящих из всех случайных величин, кроме одной.)

232 (неравенство Ширера [31]) Докажите следующее обобщение предыдущего неравенства. Пусть $T_1, \ldots, T_k$ — произвольные кортежи, составленные из случайных величин $\alpha_1, \ldots, \alpha_n$, причем каждая случайная величина входит ровно в r кортежей из числа $T_1, \ldots, T_k$. Тогда

$$rH(\alpha_1, \alpha_2, \ldots, \alpha_n) \leqslant H(T_1) + \ldots + H(T_k).$$

233 Покажите, что из неравенства задачи 231 следует оценка для объёма n-мерного тела через объемы его $n-1$-мерных проекций (обобщение неравенства, упомянутого на с. 22): объем тела, возведенный в $(n-1)$-ю степень, не превышает произведения $n-1$-мерных объемов его n проекций на координатные гиперплоскости. [Указание. Сначала докажите неравенство для тел, составленных из единичных кубиков. Для этого рассмотрите случайную величину, равномерно распределенную среди кубиков, составляющих тело. Для общего случая надо перейти к пределу, разбивая тело на кубики.]

Это неравенство является частным случаем более общего геометрического неравенства Лумиса – Уитни [87].

7.3. Сложность и энтропия

Рассмотренные нами свойства шенноновской энтропии, условной энтропии и взаимной информации (для случайных величин) сходны с аналогичными утверждениями о колмогоровской сложности (см. главу 2). Можно ли каким-либо образом уточнить и формализовать эти аналогии?

Вопрос этот можно понимать двояко. Во-первых, можно доказывать, что колмогоровская сложности и шенноновская энтропия похожи по своим свойствам (скажем, для них верны одни и те же неравенства, см. раздел 10.6, с. 364). С другой стороны, можно сравнивать численные значения колмогоровской сложности и шенноновской энтропии, и этим мы сейчас займёмся. Правда, тут сразу же возникает

проблема: шенноновская энтропия определена для случайных величин, а колмогоровская сложность — для двоичных слов, как же их сравнивать? Тем не менее иногда такое сравнение возможно, и мы сейчас разберём несколько утверждений такого вида. Общее и весьма расплывчатое описание соответствующих результатов звучит так: шенноновская энтропия учитывает лишь частотные закономерности, а колмогоровская сложность — все алгоритмические закономерности, поэтому в общем случае колмогоровская сложность меньше. Однако в тех случаях, если объект порождается случайно по вероятностной мере, с большой вероятностью других (не частотных) закономерностей нет, и сложность близка к энтропии.

Перейдём к конкретным уточнениям этого общего тезиса.

7.3.1. Колмогоровская сложность и энтропия частот

Будем рассматривать произвольный (не обязательно двухбуквенный) конечный алфавит A и слова в этом алфавите. Колмогоровская сложность для них определяется естественным образом, и мы будем говорить о ней без особых оговорок. (Мы нигде не использовали, что объекты, сложность которых определяется, являются именно двоичными словами. Но важно, что мы рассматриваем двоичные слова в качестве *описаний*: измеренная в байтах сложность будет в восемь раз меньше битовой!)

Пусть фиксирован алфавит A, содержащий k букв. Рассмотрим произвольное слово x некоторой длины N в этом алфавите. Пусть $p_1, \ldots, p_k$ — частоты появления букв в слове x. Все они являются дробями со знаменателем N; сумма частот равна единице. Пусть $h(p_1, \ldots, p_k)$ — шенноновская энтропия соответствующего распределения.

Теорема 146.
$$\frac{KS(x)}{N} \leqslant h(p_1, \ldots, p_k) + \frac{O(\log N)}{N}.$$
Неравенство обращается в равенство для большинства слов x (среди всех слов длины N с заданными частотами букв).

Имеется в виду, что $O(\log N)$ не больше $c \log N$, причём константа c не зависит ни от N, ни от слова x, ни от частот $p_1, \ldots, p_k$, но может зависеть от k, так что в этой теореме размер алфавита считается фиксированным.

◄ По существу это чисто комбинаторное утверждение. В самом деле, условния сложность $KS(x \mid N, p_1, \ldots, p_k)$ не превосходит $\log C(N, p_1, \ldots, p_k) + O(1)$, где
$$C(N, p_1, \ldots, p_k) = \frac{N!}{(p_1 N)!\,(p_2 N)! \ldots (p_k N)!}$$
есть количество слов длины N в алфавите A, имеющих данные частоты $p_1, \ldots, p_k$. Это следует из того, что каждое слово с такими частотами может быть (при известных параметрах) задано своим порядковым номером в (скажем, алфавитном) списке таких слов, для чего достаточно $\log C(N, p_1, \ldots, p_k)$ битов.

Оценка числа $C(N, p_1, \ldots, p_k)$ выполняется по формуле Стирлинга. Если отбросить полиномиальные по N множители (возникающие из-за $\sqrt{2\pi k}$ в формуле для $k!$), то останется как раз $2^{Nh(p_1, \ldots, p_k)}$. Подробно это вычисление для случая

$k = 2$ проводилось при доказательстве усиленного закона больших чисел (теорема 27, с. 66); общий случай аналогичен.

Остаётся заметить, что для задания $N, p_1, \ldots, p_k$ достаточно примерно $k \log N$ битов (нужно указать k целых чисел $p_1 N, \ldots, p_k N$, в сумме дающих N), поэтому при удалении условия в выражении $KS(x \mid N, p_1, \ldots, p_k)$ сложность возрастёт не более чем на $O(\log N)$ (с константой, примерно равной k).

Отметим, что верхняя оценка $\log C(N, p_1, \ldots, p_k)$ должна достигаться (с точностью $O(\log N)$) для большинства слов из рассматриваемого списка, поскольку слов с более короткими описаниями меньше, чем $C(N, p_1, \ldots, p_k)$). В частности, верхняя оценка достигается для слова x с максимальной (среди слов в данном списке) колмогоровской сложностью. ▶

Другой способ доказательства того же утверждения использует верхнюю оценку для монотонной сложности (теорема 89, с. 165). Рассмотрим распределение вероятностей на последовательностях букв алфавита A, соответствующее независимым испытаниям с частотами $p_1, \ldots, p_k$. Вероятность появления последовательности, начинающейся на слово z длины N, в котором частоты букв равны $q_1, \ldots, q_k$, равна

$$p_1^{q_1 N} \ldots p_k^{q_k N}$$

(буква a_i имеет вероятность p_i и появляется $q_i N$ раз), а её двоичный логарифм равен

$$-N \cdot (q_1(-\log p_1) + \ldots + q_k(-\log p_k)).$$

В частности, если $q_i = p_i$, то логарифм вероятности равен $-Nh(p_1, \ldots, p_k)$ и потому монотонная сложность оценивается сверху числом $Nh(p_1, \ldots, p_k)$. Остаётся заметить, что монотонная сложность отличается от обычной на величину порядка $O(\log N)$ для слов длины N.

Это рассуждение, однако, не вполне корректно. Дело в том, что в оценке для монотонной сложности предполагалось, что мера фиксирована и оцениваются сложности различных слов. Здесь же, интересуясь оценкой для сложности слова x, мы рассматриваем меру, которая соответствует частотам появления букв в само́м слове x, поэтому формально утверждение теоремы 89 применить нельзя. Если вспомнить её доказательство, то видно, что получается оценка для «условной» монотонной сложности при известных $p_1, \ldots, p_k$; переход от неё к безусловной сложности может привести к увеличению оценки на $O(\log N)$, так что таким способом мы получаем другое доказательство теоремы 146.

234 Оцените зависимость константы в $O(\log N)$ от k. [Указание: оба доказательства дают $k(1 + o(1)) \log N$.]

235 Покажите, что при отделённых от нуля частотах $p_1, \ldots, p_k$ оценку предыдущей задачи можно улучшить до $(k/2 + O(1)) \log N$. Точнее, для любых $\varepsilon > 0$ и k найдется константа C такая, что для всех слов x длины N в k-буквенном алфавите истинно неравенство $KS(x) \leqslant h(p_1, \ldots, p_k)N + (k/2 + 1) \log N + C$, при условии, что частоты $p_1, \ldots, p_k$ вхождений всех букв в x не меньше ε. [Указание. В первом доказательстве надо вспомнить об опущенных квадратных корнях в формуле Стирлинга, которые в основном попадают в знаменатель. Второе доказательство также

можно модифицировать, рассмотрев меру, в которой вместо точных значений p_k берутся их приближения с точностью до $1/\sqrt{N}$. При этом оценка ухудшится, но поскольку в окрестности минимума приращение гладкой функции пропорционально квадрату аргумента, то это ухудшение будет не более чем на $O(1)$; на этом мы сэкономим половину битов при задании чисел $p_1, \ldots, p_k$.]

Заметим, что неравенство теоремы 146 может быть как угодно далеко от равенства: если, скажем, алфавит состоит из двух букв, и в слове x они чередуются, то правая часть равна 1, а левая имеет порядок $(\log N)/N$. Что и не удивительно — правая часть учитывает только частоты, а не другие (не-частотные) закономерности. В следующем разделе мы покажем, что при случайном порождении слова сложность близка к энтропии.

7.3.2. Математическое ожидание сложности

Пусть фиксирован k-буквенный алфавит A и k чисел $p_1, \ldots, p_k \geqslant 0$ с суммой 1 (которые мы для простоты считаем рациональными).

Рассмотрим случайную величину ξ, значениями которой являются буквы алфавита A, принимаемые с вероятностями $p_1, \ldots, p_k$. Для каждого N рассмотрим случайную величину ξ^N, соответствующую N независимым копиям случайной величины ξ. Её значениями являются слова длины N в алфавите A (каждая буква порождается независимо, причём вероятность появления i-й буквы равна p_i). Нас будет интересовать математическое ожидание сложности значений случайной величины ξ^N (взвешенное среднее с весами, равными вероятностям).

Теорема 147. *Математическое ожидание $KP(\xi^N\,|\,N)$ равно $NH(\xi) + O(1)$ (константа в $O(1)$ может зависеть от ξ, но не зависит от N).*

Заметим, что (при положительных p_i) среди значений величины ξ^N будут встречаться все слова длины N в алфавите A; некоторые из них имеют сложность много больше NH (если только распределение вероятностей не равномерное, поскольку в этом случае таких слов нет), а другие имеют сложность много меньше NH.

◀ Для каждого слова длины N (то есть для каждого значения случайной величины ξ^N) рассмотрим его кратчайшее описание при известном N (относительно оптимального префиксно корректного способа описания). Полученные слова образуют префиксный код в смысле раздела 7.1.1. Средняя длина этого кода как раз равна математическому ожиданию сложности значений величины ξ^N. Поэтому теорема 138 (с. 241) гарантирует, что это математическое ожидание не меньше $H(\xi^N) = NH(\xi)$. Нижняя оценка (и даже без константы $O(1)$) доказана.

Та же самая теорема позволяет получить и верхнюю оценку. В самом деле, она утверждает существование префиксного кода, имеющего среднюю длину кодового слова не выше $H(\xi^N) + 1$. Такой код можно алгоритмически построить, зная число N (и числа p_i, предполагаемые фиксированными). Например, можно использовать конструкцию из доказательства теоремы 138, или применить код Хаффмана, или даже просто перебирать все коды, пока не найдётся подходящий.

Так или иначе построенный код можно рассматривать как некоторый способ условного префиксного описания (при известном N), для которого средняя пре-

фиксная сложность значения величины ξ^N не превосходит $H(\xi^N) + 1 = NH(\xi) + 1$. Переход к оптимальному префиксно корректному способу описания увеличит среднюю сложность не более чем на константу. ▶

236 Покажите, что можно немного усилить верхнюю оценку и показать, что среднее значение монотонной сложности $KM(\xi^N)$ не превосходит $NH(\xi) + O(1)$. [Указание. Примените теорему 89 к распределению вероятностей, соответствующему независимым испытаниями величины с распределением ξ.]

Мы предполагали, что величины $p_1, \ldots, p_k$ рациональны и фиксированы заранее. Если стремиться получить «равномерную» оценку, имеющую место для всех наборов $p_1, \ldots, p_n$, то следует добавить их в условие и оценивать среднее значение величины $KP(\xi^N \mid N, p_1, \ldots, p_k)$. На нижнюю оценку это вообще не повлияет, поскольку она верна для любого префиксного кода, а для получения верхней оценки (построения кода) этой информации достаточно. (Сказанное относится к случаю рациональных значений p_i; для произвольных действительных значений p_i можно взять их достаточно точные приближения.)

237 Сформулируйте и докажите соответствующие точные утверждения.

Доказанная теорема показывает, что *в среднем* сложность равна энтропии, хотя она бывает и больше, и меньше её. На самом деле верно и более сильное утверждение: почти что с единичной вероятностью сложность близка к энтропии, и вероятность того, что случайно выбранное значение величины ξ^N будет иметь сложность, сильно отличающуюся от энтропии, мала. Это утверждение является алгоритмическим аналогом теоремы Шеннона о пропускной способности канала без шума, и мы вернёмся к этому вопросу в разделе 7.3.4.

7.3.3. Сложность начальных отрезков случайных последовательностей

Покажем, как связана сложность с энтропией для начальных отрезков случайных (в смысле Мартин-Лёфа) последовательностей. Пусть по-прежнему фиксирован некоторый алфавит A из k букв и распределение вероятностей $p_1, \ldots, p_k$ на буквах этого алфавита. Будем считать, что $p_1, \ldots, p_k$ являются вычислимыми действительными числами.

Рассмотрим пространство A^∞ бесконечных последовательностей букв алфавита A и распределение вероятностей на нём, соответствующее независимым появлениям каждой буквы с распределением $p_1, \ldots, p_k$. Получается вычислимое распределение вероятностей на A^∞, и можно говорить о случайных по Мартин-Лёфу элементах A^∞. (Соответствующая теория, которая была у нас для двухбуквенного алфавита, переносится безо всяких изменений.)

Теорема 148. *Пусть ω — случайная по Мартин-Лёфу последовательность (относительно указанного распределения), $(\omega)_N$ — её начальный отрезок длины N. Тогда*

$$\lim \frac{KS((\omega)_N)}{N} = H,$$

где H — шенноновская энтропия, то есть $H = \sum p_i(-\log p_i)$.

238 Покажите, что для равномерного распределения это утверждение является непосредственным следствием критерия случайности (теорема 90, с. 167).

(Редкий случай, когда равномерность распределения действительно существенна.)

В формулировке теоремы мы использовали простую колмогоровскую сложность KS, но это не имеет значения, поскольку разные варианты отличаются на $O(\log N) = o(N)$. В доказательстве будет удобнее использовать монотонную сложность.

◀ По критерию случайности Левина – Шнорра (теорема 90, с. 167) сложность близка к минус логарифму вероятности появления начального отрезка $(\omega)_N$. Вероятность берётся относительно рассматриваемого распределения вероятностей на A^∞, и упомянутый минус логарифм равен $N \sum q_i(-\log p_i)$, где q_i — частота появления i-й буквы в слове $(\omega)_N$. Остаётся воспользоваться тем, что для случайных последовательностей справедлив усиленный закон больших чисел, который гарантирует, что q_i стремится к p_i при $N \to \infty$. ▶

Из доказательства видно, что отклонение удельной сложности (в расчёте на букву) от энтропии складывается из трёх частей: дефект случайности (который имеет порядок $O(1)/N$), различие между простой сложностью и монотонной (порядок $O(\log N)/N$) и отклонение частот от вероятностей (оно даёт наибольший вклад; закон повторного логарифма в теории вероятностей говорит, что это отклонение чуть больше $O(\sqrt{N})/N$.

Мы предполагали, что числа p_i вычислимы, поскольку в противном случае нельзя говорить о случайности по Мартин-Лёфу. Тем не менее и в этом случае можно утверждать, что множество тех последовательностей, у которых удельная сложность начальных отрезков не стремится к шенноновской энтропии соответствующего распределения, имеет меру 0 (относительно данного распределении вероятностей).

239 Докажите это. [Указание. При получении верхней оценки для сложности можно использовать не сами числа p_i, а приближения к ним (с некоторым запасом точности, скажем, с точностью до $1/N^2$ для слов длины N); при этом для задания этих приближений потребуется дополнительная информация объёма $O(\log N)$. Нижняя же оценка вообще не использует алгоритмической природы p_i (скажем, можно получить оценку для сложности с оракулом, при котором p_i вычислимы).]

7.3.4. Вероятность отклонения сложности от энтропии

Теорема 148 носит «предельный» характер; интересно, однако, получить и оценки вероятностей больших отклонений энтропии от сложности для конечных последовательностей. (Поясним это аналогией: помимо усиленного закона больших чисел, говорящего о пределах частот, в теории вероятностей есть оценки вероятностей больших уклонений на конечных отрезках.)

Пусть фиксировано некоторое распределение вероятностей $p_1, \ldots, p_k$ на буквах алфавита A. Будем снова для простоты предполагать, что числа p_i рациональны (или хотя бы вычислимы). Рассмотрим распределение вероятностей на A^N, считая

буквы во всех позициях независимыми. Для каждого слова, порождаемого таким процессом, рассмотрим его сложность. Мы уже знаем (теорема 147), что среднее значение этой сложности равно энтропии NH, где $H = \sum p_i(-\log p_i)$. Но интересно знать также, как именно колеблется сложность вокруг этого среднего значения.

В самом простом (но, как мы увидим, совсем не типичном) случае, когда имеются две равновероятные буквы, распределение вероятностей равномерно на всех словах длины N. Мы знаем, что в этом случае все слова имеют сложность не больше $N + O(1)$, а подавляющее большинство слов (за вычетом доли 2^{-c}) имеют сложность не меньше $N - c$, то есть сколько-нибудь большие отклонения сложности от энтропии весьма маловероятны.

Примерно так же обстоит дело и в случае равновероятных букв для алфавита любого размера. Однако если не все буквы равновероятны, ситуация меняется.

Ключевое наблюдение здесь состоит в следующем. Для каждого слова x, помимо фиксированных вероятностей p_i, рассмотрим также фактические частоты $q_i(x)$ появления соответствующих букв в x. Оказывается, что с большой вероятностью сложность случайно выбранного (в соответствии с рассматриваемым распределением p_i) слова близка к $k(x) = N\sum_i q_i(x)(-\log p_i)$. В самом деле, монотонная сложность по теореме 89 (с. 165) превосходит $k(x)$ не более чем на константу. С другой стороны, рассуждение при доказательстве теоремы Левина – Шнорра (с. 167, лемма 1) показывает, что вероятность события $KM(x) < k(x) - c$ (в соответствии с рассматриваемым нами распределением вероятности на словах длины N) не превосходит 2^{-c} для любого c.

Таким образом, вопрос о распределении сложностей сводится к хорошо известному в теории вероятностей вопросу о распределении частот. Известно (теорема Муавра – Лапласа), что оно близко к нормальному (гауссовскому); математическое ожидание частоты есть вероятность, а среднеквадратическое отклонение убывает пропорционально $1/\sqrt{N}$. По сравнению с таким отклонением разницей порядка $O(\log N)$ (различные виды сложности, использование N в качестве условия и т.д.) можно пренебречь. Сказанное при некотором уточнении превращается в доказательство следующей теоремы:

Теорема 149. *Пусть фиксирована случайная величина ξ с k значениями. Для любого положительного $\varepsilon > 0$ существует такое число c, что при всех N вероятность события $NH(\xi) - c\sqrt{N} < KS(x) < NH(\xi) + c\sqrt{N}$, если все буквы слова x независимы и распределены как ξ, не меньше $1 - \varepsilon$.*

Строго говоря, наши рассуждения использовали вычислимость чисел p_i. Но от этого предположения можно избавиться, если заменить числа p_i достаточно точными приближениями (скажем, с точностью до $1/N^2$, что требует $O(\log N)$ дополнительных битов).

7.3.5. Теорема Шеннона о кодировании

Доказанное нами утверждение об отклонении сложности от энтропии является естественным переводом на язык теории сложности известных результатов Шеннона о длине кода, который позволяет передавать блоки из N независимых случайных символов с данным распределением с малой вероятностью ошибки.

Пусть вновь ξ — случайная величина с k значениями (буквами алфавита A) и некоторым распределением. Пусть N — некоторое число, а ξ^N — случайная величина со значениями в A^N, у которой буквы в каждой позиции распределены как ξ и независимы. Мы хотим кодировать значения случайной величины m-битовыми словами по схеме рис. 7.3.

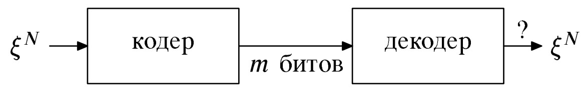


Рис. 7.3. Кодирование величины ξ^N с помощью m битов.

Здесь в качестве кодера рассматривается произвольное отображение множества A^N в множество $\mathbb{B}^m$ двоичных слов длины m, а в качестве декодера — произвольное отображение $\mathbb{B}^m$ в A^N. Говорят, что происходит ошибка, если выходное слово в алфавите A не равно входному (значению величины ξ^N). Вероятность ошибки измеряется относительно описанного распределения случайной величины ξ^N. Нас интересует, при каких m и ε существует код длины m с вероятностью ошибки не более ε. Начнём со следующего очевидного замечания:

Теорема 150. *Такой код существует тогда и только тогда, когда 2^m наиболее вероятных значений случайной величины ξ^N имеют суммарную вероятность не меньше $1-\varepsilon$.*

◀ В самом деле, при кодировании m битами можно передать без ошибок 2^m значений, но не больше. Ясно, что вероятность ошибки будет минимальной, если в качестве этих 2^m значений выбрать 2^m наиболее вероятных значений. ▶

В следующей теореме алфавит A и величина ξ фиксированы.

Теорема 151. *Для всякого $\varepsilon > 0$ существует такое число c, что*:

(а) *для ξ^N существует код длины не больше $NH(\xi) + c\sqrt{N}$ с вероятностью ошибки не больше ε*;

(б) *любой код длины не больше $NH(\xi) - c\sqrt{N}$ для ξ^N имеет вероятность ошибки не меньше $1-\varepsilon$ (вероятность правильной передачи не больше ε).*

◀ (а) Мы знаем, что при подходящем c значение случайной величины ξ^N имеет сложность менее $m = NH(\xi) + c\sqrt{N}$ с вероятностью $1-\varepsilon$. Поэтому для таких значений можно использовать кратчайшие описания (в смысле колмогоровской сложности) в качестве кодов. (Заметим, что в теореме ничего не говорится про алгоритм кодирования, нам важно лишь существование кода. Формально говоря, в качестве закодированных сообщений у нас выступают слова длины меньше m, но их менее 2^m штук и можно их заменить на слова длины m.)

(б) Здесь нам понадобится небольшая хитрость. Если существует код указанной длины с указанной вероятностью ошибки, то можно построить такой код эффективно (воспользовавшись предыдущей теоремой, или просто полным перебором) и рассматривать декодирующую функцию этого кода как способ условного описания (при известных параметрах p_i и N). Поэтому все значения величины ξ^N, которые раскодируются правильно, имеют (условную — относительно указанных

параметров) сложность не более $NH(\xi) - c\sqrt{N} + O(\log N)$ (последний член соответствует сложности параметров и поглощается при увеличении константы c). По предыдущей теореме (при подходящем c) вероятность этого не превосходит ε. ▶

240 Как и раньше, в этом рассуждении предполагается, что мы знаем числа p_i точно, и если они невычислимы, возникают дополнительные трудности. Покажите, как можно исправить это рассуждение, заменив p_i на их достаточно точные приближения.

241 Сформулируйте и докажите аналогичным способом теорему об условном кодировании и условной энтропии. [Указание. Пусть имеются две зависимые случайные величины ξ и η, проводится N независимых испытаний, значение величины η^N известно и отправителю, и получателю, и отправитель хочет передать m битов так, чтобы получатель смог восстановить значение величины ξ^N. При каких m это возможно, а при каких нет?]

8. Некоторые приложения

8.1. Бесконечность множества простых чисел

Начнём с совсем «игрушечного» применения: докажем, что множество простых чисел бесконечно.

Пусть это не так и существует всего m различных простых чисел $p_1, \ldots, p_m$. Тогда любое целое число x, разлагаясь на простые множители, представляется в виде

$$x = p_1^{k_1} p_2^{k_2} \ldots p_m^{k_m},$$

и тем самым может быть задано набором степеней $k_1, \ldots, k_m$. Каждое из чисел k_i не превосходит $\log x$ (основания степеней не меньше 2) и потому имеет сложность не более $O(\log\log x)$ (двоичная запись содержит $O(\log\log x)$ битов). Поскольку m фиксировано, то сложность набора $\langle k_1, k_2, \ldots, k_m\rangle$ есть $O(\log\log x)$ и потому сложность любого числа x (которое алгоритмически получается из этого набора) есть $O(\log\log x)$. А для «случайного» n-битового числа x сложность примерно равна n, а не $O(\log n)$, как получается по этой формуле (логарифм n-битового числа не превосходит n).

Можно ли считать это «честным» применением колмогоровской сложности? Скептик скажет, что здесь всего лишь производится обычный подсчёт числа возможностей (counting argument, как говорят): если существует лишь m различных простых чисел, то различных чисел от 1 до x существует не более $(\log x)^m$, поскольку каждое такое число задаётся показателями степеней при простых множителях, и каждый показатель меньше $\log x$. После чего мы приходим к противоречию, поскольку $x > (\log x)^m$ при больших x.

Возразить на это нечего — именно это рассуждение и пересказано с помощью колмогоровской сложности (и стало немного более громоздким за счёт различных асимптотических обозначений). Тем не менее в подобном переводе может быть некоторый смысл, поскольку новый язык формирует новую интуицию, и она может быть полезна, даже если потом можно всё пересказать на старом языке.

К этому обсуждению мы ещё вернёмся, разобрав другие примеры.

8.2. Перенос информации по ленте

Следующий пример тоже модельный — мы докажем, что копирование слова длины n на одноленточной машине Тьюринга требует (в худшем случае) не менее εn^2 шагов. Этот классический результат был получен в 1960-е годы с помощью так

называемого «метода следов»; наше доказательство является переводом классического на язык колмогоровской сложности. (Мы предполагаем известными основные понятия, связанные с машинами Тьюринга; см, например, [175]).

Пусть на ленте машины Тьюринга выделена «буферная зона» некоторого размера b; нас интересует скорость переноса информации через эту буферную зону, скажем, слева направо, из области L в область R (рис. 8.1).

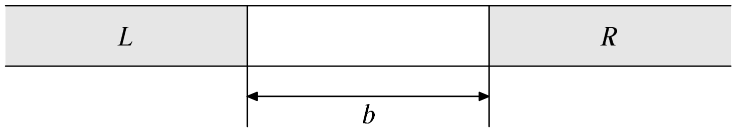


Рис. 8.1. Буферная зона размера b.

Предположим, что в некоторый момент ленты буферная область и R пусты, а область L содержит произвольную информацию. Нас интересует, какова может быть сложность слова R через t шагов после этого. Мы докажем, что она не больше $(t\log m)/b + O(\log t)$, где m — число состояний машины Тьюринга, а b — ширина буферной зоны. В самом деле, каждое из m состояний машины несёт $\log m$ битов информации, и за шаг информация переносится на одну клетку, а перенос её на b клеток требует в b раз больше времени.

Осталось лишь уточнить формулировку и доказательство.

Теорема 152. *Пусть фиксирована машина Тьюринга с m состояниями. Тогда существует такая константа c, что для любого b и для любого вычисления с буферной зоной b (вначале эта зона и лента справа от неё пусты, головка машины слева от зоны) сложность правой части ленты $R(t)$ после t шагов вычисления не превосходит*

$$\frac{t\log m}{b} + 4\log t + c.$$

◀ Проведём границу где-нибудь посередине буферной зоны, и при каждом пересечении границы головкой машины Тьюринга слева направо (при «выезде за границу») будем записывать, в каком состоянии она её пересекла (мечта ЧК-ГБ). Записанная последовательность состояний называется *следом* машины. Зная след, можно восстановить работу машины справа от границы (без использования информации о ленте слева от границы). В самом деле, надо искусственно поместить машину в первое состояние, указанное в следе, и выпустить её за границу; по возвращении перевести её во второе состояние следа и снова выпустить и так далее. При этом за границей машина будет вести себя так же, как и раньше (ведь ничего, кроме состояния, при пересечении границы у неё нет). В частности, в некоторый момент t' правая часть ленты будет содержать $R(t)$. При этом t' не превосходит t, хотя может быть и меньше (поскольку время, проведённое машиной слева от границы, теперь не учитывается). Таким образом, можно алгоритмически восстановить $R(t)$, зная след, t', а также расстояние b' от границы до начала зоны R. Поэтому найдётся такая константа c (зависящая от машины, но не от b и t), что для любого

следа S

$$KS(R(t)) \leqslant l(S)\log m + 4\log t + c$$

(мы умножили длину $l(S)$ следа S на $\log m$, поскольку след является словом в m-буквенном алфавите и на каждую букву приходится $\log m$ битов; приписывание к нему чисел b' и t' (в самоограниченной записи) требует не более $2\log b + 2\log t$ битов, причём можно считать, что $t > b$ (иначе R пуста, головка туда не дошла); константа c возникает при переходе к оптимальному способу описания).

Это неравенство верно для любого начального содержимого части L и для любого положения границы: если теперь для данного L взять самый короткий из следов, то его длина меньше t/b (всего разных положений границы $b+1$, и в каждый момент пересечение происходит лишь в одном месте, так что сумма длин следов не больше t). Получаем требуемое неравенство. ►

$\boxed{\textbf{242}}$ Покажите, что оценку теоремы можно улучшить, заменив b в знаменателе на $2b$. [Указание. Въезд суммарно требует почти столько же шагов, сколько и выезд.]

Теперь сразу же получается квадратичная оценка для задачи копирования.

Пусть одноленточная машина M копирует любое входное слово: если вначале на ленте написано слово x из нулей и единиц, а дальше лента пуста, то в конце работы справа от x появляется его копия, и на ленте написано xx.

Теорема 153. *Существует такая константа $\varepsilon > 0$, что для любого n существует слово длины n, копирование которого с помощью M занимает не менее εn^2 шагов.*

◄ Будем для простоты считать, что n чётно, и возьмём в качестве x слово, у которого вторая половина u имеет сложность не меньше длины (то есть $n/2$). Применим теорему о скорости переноса информации, считая буферной зоной участок длины $n/2$ справа от x (рис. 8.2).

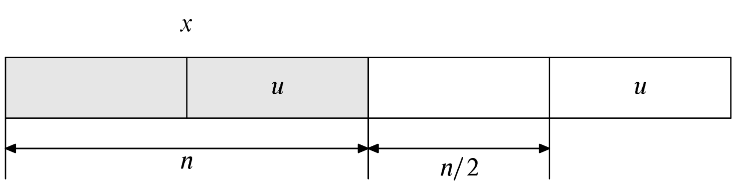


Рис. 8.2. Буферная зона при копировании.

Пусть копирование заняло t шагов, тогда сложность зоны R после этого не меньше $n/2$; с другой стороны, она не превосходит $t\log m/b + 4\log t + c$ по доказанной теореме, где b — ширина буферной зоны, то есть $n/2$. Получаем, что

$$\frac{n}{2} \leqslant \frac{t\log m}{n/2} + 4\log t + c.$$

Без ограничения общности считаем, что $t < n^2$ (иначе всё и так хорошо) и заменяем $4\log t$ на $8\log n$. Получаем, что

$$t \geqslant \frac{n^2}{4\log m} - O(n\log n).$$

Второй член мал по сравнению с первым при бо́льших n (и формально можно распространить неравенство на все n за счёт уменьшения коэффициента при n^2). ►

Насколько существенна в этом доказательстве колмогоровская сложность? Скептик может вновь сказать, что по существу мы просто оценивали число слов, которые можно скопировать за данное время, используя тот факт, что разные слова должны давать разные следы (иначе справа от границы поведение машины было бы одно и то же). Действительно, именно так и выглядело первоначальное доказательство этой оценки (уточним, что в первоначальных работах рассматривалось не копирование, а распознавание симметрии, но вся техника та же самая). Насколько идея этого доказательства становится понятнее при переводе его на язык колмогоровской сложности — вопрос вкуса.

Многие оценки в теории сложности вычислений могут быть доказаны аналогичным образом, когда в качестве «трудного случая» рассматривается слово максимальной сложности в некотором классе и затем показывается, что если бы оценка не соблюдалась, то это слово имело бы меньшую сложность. Много таких приложений (со ссылками на оригинальные работы) приведено в книге Ли и Витаньи [85], которые сами сыграли большую роль в развитии этого метода (называемого *Incompressibility method*). Отметим, что во многих случаях оценка впервые была доказана как раз с использованием колмогоровской сложности.

В следующем разделе мы приведём ещё один пример использования этого метода, а затем перейдём к некоторым другим приложениям. Они интересны не сами по себе, а как примеры разнообразных способов, с помощью которых можно применять теорию колмогоровской сложности для доказательства утверждений, с ней не связанных.

8.3. Конечные автоматы с несколькими головками

Конечный автомат с k головками отличается от обычного (с которым, как мы предполагаем в этом разделе, читатель знаком) тем, что вместо одной читающей головки он имеет k односторонних (движущихся слева направо) читающих головок. В начальный момент k читающих головок видят первый (крайний слева) символ входного слова. На каждом шаге, в зависимости от состояния автомата и от k символов, читаемых k головками, автомат меняет своё состояние и продвигает некоторые из k головок (не менее одной) вправо, после чего процесс повторяется. Вслед за словом на ленте записан специальный символ конца слова; работа автомата заканчивается, когда все головки попадают на этот символ. Автомат *допускает* слово, если по окончании работы он находится в заключительном состоянии. Множество всех слов, допускаемых автоматом, называется *языком, распознаваемым этим автоматом.*

Пример. Рассмотрим множество всех слов вида $x\#x$, где x — всевозможные двоичные слова. Этот язык не распознаётся конечным автоматом с одной головкой (классический результат теории конечных автоматов), но распознаётся автоматом с двумя головками. В самом деле, одну головку надо отправить на поиски символа #, после чего синхронно сдвигать обе головки и проверять, что под ними проходят одинаковые символы.

Тем самым две головки лучше, чем одна (можно распознавать больше языков). Оказывается, что вообще $k+1$ головок лучше, чем k:

Теорема 154. *Для всякого k существует язык, который распознаётся некоторым автоматом с $k+1$ головками, но не распознаётся никаким автоматом с k головками.*

◀ Для каждого $m \geqslant 1$ рассмотрим язык L_m, состоящий из всех слов

$$w_1\#\ldots\#w_m\#w_m\#\ldots\#w_1$$

(для всех двоичных слов $w_1,\ldots,w_m$). Каждое из слов w_i повторяется дважды, причём в правой половине слова идут в обратном порядке (это очень существенно).

Автомат с k головками может распознавать язык L_m, если m не слишком велико по сравнению с k (как мы увидим, при $m \leqslant C_k^2$). Это делается так: одна из головок отправляется в правую половину слова, а остальные $k-1$ головок встают на начала слов $w_1,\ldots,w_{k-1}$. Каждая из этих $k-1$ головок сверяет порученное ей слово с его копией, когда первая головка проходит мимо этой копии. После этого первые $k-1$ слов проверены, первая головка полностью использована (она дошла до конца слова и больше ни на что не пригодна), а остальные головки стоят левее ещё не проверенных слов ($w_k, w_{k+1},\ldots$). Процесс повторяется: ещё одна головка отправляется через всё слово, а $k-2$ головок сверяют каждая своё слово и так далее. Всего таким образом можно сверить

$$(k-1)+(k-2)+\ldots+1=\frac{k(k-1)}{2}=C_k^2$$

слов. (Заметим, что число m у нас фиксировано, поэтому поиск слова с заданным номером требует ограниченного числа состояний автомата.)

Осталось показать, что *при $m > C_k^2$ язык L_m не распознаётся никаким k-головочным автоматом.* Пусть это не так и автомат M его распознаёт. Чтобы придти к противоречию, рассмотрим независимые случайные слова $w_1,\ldots,w_m$ достаточно большой длины N. Более формально, рассмотрим слово длины mN и сложности не меньше mN, и разрежем его на m слов длины N, которые обозначим $w_1,\ldots,w_m$. По предположению слово

$$W=w_1\#\ldots\#w_m\#w_m\#\ldots\#w_1$$

допускается автоматом M; мы придём к противоречию, показав, что либо слово $w_1\ldots w_m$ имеет меньшую сложность, чем мы предположили, либо автомат M работает неверно (не распознаёт L_m).

Будем говорить, что пара головок автомата M *проверила* слово w_i, если при работе автомата M на данном слове был момент, когда эти головки находились внутри левой и правой копии слова w_i. (Говоря о парах, мы имеем в виду неупорядоченные пары с различными элементами.) Ключевое наблюдение: *никакая пара головок не может проверить два разных слова*. В самом деле, если она проверила слово w_i, то до этого левая головка читала слова с номерами, меньшими i, а правая — с бóльшими, а после этого наоборот.

По предположению $m > C_k^2$; поэтому существует некоторое слово w_i, не проверенное ни одной парой головок. Покажем, что либо это слово проще, чем нужно, либо одну из его копий можно изменить незаметно для автомата M (тем самым дискредитировав этот автомат).

Будем наблюдать за работой автомата M на слове W, обращая особое внимание на те моменты, когда одна из головок автомата входит в слово w_i (в любую из копий) или выходит из него, и записывать состояние автомата в эти моменты, а также передвижение головок (откуда и куда они переместились). Полученный протокол P имеет сложность $O(\log N)$ (константа зависит от k, m и от числа состояний автомата, но не от N), поскольку моментов входа и выхода не более $4k$ (по четыре для каждой головки), и в каждый момент надо записать состояние автомата и положение всех головок, что требует $O(\log N)$ битов.

Покажем, что (если автомат M распознаёт L_m) слово w_i восстанавливается однозначно по всем остальным словам w_j с $j \neq i$ и протоколу P. В этом случае сложность слова $w_1 \ldots w_m$ не превосходит $(m-1)N$ (число битов в остальных словах) плюс $O(\log N)$ (число битов в протоколе) плюс $O(1)$, что меньше mN при больших N, и полученное противоречие завершает доказательство.

Восстанавливать будем так: подставляя по очереди всевозможные слова w данной длины на место w_i (при тех же w_j с $j \neq i$), мы запускаем автомат M и смотрим, не получился ли протокол P. При этом возможны разные случаи.

(1) Если хотя бы при одном w автомат M не допустит соответствующего слова, он заведомо неправильный.

(2) Автомат M допускает все такие слова, при этом протокол P встречается только однажды, при $w = w_i$. (В этом случае w_i восстанавливается по P и остальным словам.)

(3) Автомат M допускает все такие слова, при этом протокол P встречается и при некотором $w \neq w_i$. Покажем, что в этом случае автомат M допускает неправильное слово (не из L_m), а именно слово W', в котором в левой половине записано w_i, а в правой на месте w_i стоит слово w.

В самом деле, у нас есть два допускающих поведения автомата M: когда с обеих сторон w_i и когда с обеих сторон w. Разрежем каждое из них на отрезки по тем моментам, когда одна из головок входит или выходит из w_i (или w). Положения всех головок в эти моменты, а также состояния автомата, зафиксированы в протоколе P, и потому одни и те же в обоих случаях. (Сами моменты времени могут быть различны, поскольку их мы — хотя могли бы — не включили в протокол.) Поэтому можно склеивать отрезки поведения автомата для двух разных случаев; покажем, что это можно сделать так, чтобы в левой половине было w_i, а в правой w.

По предположению в ходе обработки слова W никакая пара головок одновременно не попадает в левую и правую копии слова w_i; поскольку прохождение границы фиксируется в протоколе, это же верно и после замены w_i на w. Поэтому для каждого отрезка протокола есть три возможности: (а) в левой копии есть головка; (б) в правой копии есть головка; (в) ни то, ни другое. Если теперь для случая (а) брать кусок поведения автомата M на слове W, для случая (б) брать кусок поведения на изменённом слове (когда w_i заменено на w), а для случая (в) брать любой из двух вариантов (они совпадают), то получится работа автомата M на смешанном слове W', и потому автомат допускает W' в противоречии с предположением. ▶

8.4. Усиленный закон больших чисел

Усиленный закон больших чисел был доказан нами в разделе 3.2 (теорема 27, с. 66) без использования колмогоровской сложности (прямым подсчётом). Подробно был разобран случай равномерной меры: в этом случае закон утверждает, что множество тех последовательностей $\omega = \omega_0\omega_1 \ldots$, у которых предел частот

$$p_n = \frac{\omega_0 + \omega_1 + \ldots + \omega_{n-1}}{n}$$

при $n \to \infty$ существует и равен $1/2$, имеет меру 1 относительно равномерной меры на пространстве Ω. Другими словами, дополнение этого множества (те последовательности, где предел не существует или не равен $1/2$), является нулевым. Впоследствии, в теореме 32 (с. 78), мы установили, что это нулевое множество является и эффективно нулевым, откуда заключили, что для всех случайных по Мартин-Лёфу последовательностей предел частот равен $1/2$ (теорема 33, с. 78).

Однако можно действовать и в другом направлении. А именно, можно сначала доказать, что для всех случайных по Мартин-Лёфу последовательностей предел частот равен $1/2$, используя критерий случайности в терминах сложности (теорема 90, с. 167). Этот критерий говорит, что для случайной по равномерной мере последовательности ω монотонная сложность её начального отрезка $(\omega)_n$ длины n равна $n + O(1)$. Из этого можно заключить, что частота единиц p_n единиц в $(\omega)_n$ стремится к $1/2$. В самом деле, по теореме 146, сложность (обычная или монотонная, разница между ними логарифмическая) слова $(\omega)_n$ не превосходит $nh(p_n, 1-p_n) + O(\log n)$, и потому величина $h(p_n, 1-p_n)$ стремится к 1 при $n \to \infty$ (разница между ними не превосходит $O(\log n)/n$). Отсюда и из графика рис. 3.1 (с. 67) можно заключить, что $p_n \to 1/2$ при $n \to \infty$ для любой случайной последовательности, а такие последовательности образуют множество меры 1.

Возникает вопрос, является ли это рассуждение действительно новым или просто мы повторили по существу те же аргументы на несколько другом языке? Скорее второе: если вспомнить, как доказывалась теорема 146, то видно, что там использовалась та же оценка с формулой Стирлинга, что и в прямом доказательстве закона больших чисел. (Другое рассуждение, с оценкой через логарифм меры по теореме 89, также имеет прямой перевод в терминах оценок вероятностей, который приводился нами в разделе 3.2, после доказательства теоремы 27.)

Что же мы выигрываем от перехода к колмогоровской сложности? Можно указать несколько преимуществ. Во-первых, можно расширить класс последовательностей, для которых справедлив усиленный закон больших чисел:

Теорема 155. *Пусть последовательность ω такова, что $KS((\omega)_n) = n + o(n)$. Тогда последовательность частот единиц в её начальных отрезках стремится к $1/2$.*

◀ Доказательство аналогично: $h(p_n, 1-p_n)$ по прежнему есть $1 + o(1)$. ▶

Во-вторых, мы можем не ограничиваться утверждением $p_n \to 1/2$, а оценивать скорость этой сходимости. В теории вероятностей одна из таких оценок даётся *законом повторного логарифма*; этот закон выполняется для последовательностей, случайных по Мартин-Лёфу (как доказал Вовк [182]). Следуя Вовку, приведём простое доказательство верхней оценки в этом законе, использующее колмогоровскую сложность.

Теорема 156. *Пусть ω — последовательность, случайная по Мартин-Лёфу относительно равномерной меры, p_n — частота единиц в её начальном отрезке длины n. Тогда, каково бы ни было $\varepsilon > 0$, неравенство*

$$|p_n - 1/2| \leqslant (1+\varepsilon)\sqrt{\frac{\ln\ln n}{2n}}$$

выполнено для всех достаточно больши́х n.

◀ Посмотрим, какую оценку даёт нам (описанное выше) использование колмогоровской сложности. Мы знаем, что

$$n - O(1) \leqslant KM((\omega)_n) \leqslant nh(p_n, 1-p_n) + O(\log n),$$

откуда

$$h(p_n, 1-p_n) \geqslant 1 - O(\log n/n).$$

Функция

$$p \mapsto h(p, 1-p) = p(-\log p) + (1-p)(-\log(1-p))$$

имеет максимум в точке $p = 1/2$, при этом вторая производная в точке максимума отлична от нуля (и равна $-4/\ln 2$). Поэтому по формуле Тейлора получаем, что

$$h(1/2+\delta, 1/2-\delta) = 1 - \frac{2}{\ln 2}\delta^2 + o(\delta^2)$$

при $\delta \to 0$. Следовательно, для $\delta_n = p_n - 1/2$ имеем

$$\delta_n^2 = O(\log n/n),$$

то есть

$$|p_n - 1/2| = O\left(\sqrt{\frac{\log n}{n}}\right).$$

Это уже кое-что, хотя до нужной нам оценки ещё далеко. (Заметим в скобках, что и в теории вероятностей оценка с повторным логарифмом была получена не

сразу: сначала Хаусдорф (1913) доказал оценку $O(n^{\varepsilon}/\sqrt{n})$, затем Харди и Литлвуд (1914) её улучшили, заменив n^{ε} на $\sqrt{\log n}$, затем Штейнгауз (1922) доказал оценку $(1+\varepsilon)\sqrt{(2\ln n)/n}$, и лишь потом Хинчин (1924) получил более сильную оценку доказываемой нами теоремы 156, заменив логарифм на повторный логарифм. Так что пока что мы находимся на уровне Харди и Литлвуда.)

Посмотрим, как можно улучшить верхнюю оценку для $KM((\omega)_n)$. (Нижнюю оценку улучшать особенно некуда.) Она получалась сравнением $KM((\omega)_n)$ с логарифмом вероятности последовательности $(\omega)_n$ относительно бернуллиевой меры с параметром p_n. Этот логарифм был в точности равен $nh(p_n, 1-p_n)$, но сама бернуллиева мера зависела от n, и потому конструкция из доказательства теоремы 89 давала оценку с дополнительным членом, равным $KP(p_n)$ (мы должны начать с самоограниченного кода для p_n). При этом $KP(p_n)$ можно оценить как $(2+\varepsilon)\log n$, поскольку числитель и знаменатель дроби p_n не превосходят n. Всё это вместе даёт оценку
$$\frac{2}{\ln 2}(p_n - 1/2)^2 \approx 1 - h(p_n, 1-p_n) \leqslant (2+\varepsilon)\log n/n,$$
что по-прежнему недостаточно.

Что же ещё можно сэкономить? Заметим, что мы уже знаем, что p_n достаточно близко к $1/2$: если знаменателем считать n, то числитель отклоняется от $n/2$ на величину чуть больше $\sqrt{n}$. Поэтому (при известном знаменателе) количество битов для описания числителя можно сократить примерно вдвое, и тогда вместо 2 в правой части будет 1,5. Но мы хотим большего.

Решающая идея состоит в следующем. Пусть для определённости $p_n = 1/2 + \delta_n > 1/2$. Будем вместо p_n использовать для оценки монотонной сложности величину $1/2 + \delta'_n$, где δ'_n — приближение к δ_n снизу с небольшой (фиксированной) относительной погрешностью. Пусть, например, $0{,}9\delta_n < \delta'_n \leqslant \delta_n$. Такое число δ'_n можно найти среди членов геометрической прогрессии $(0{,}9)^k$, и его сложность равна примерно $\log k$, то есть $\log(-\log\delta_n/\log 0{,}9) = \log(-\log\delta_n) + c$. При этом, если $\delta_n < 1/\sqrt{n}$, то доказывать нечего, поэтому сложность δ_n можно оценить как $(1+\varepsilon)\log\log n$ (при сколь угодно малом ε и достаточно больших n).

Это хорошо, но есть и ухудшение. А именно, вместо $h(p_n, 1-p_n)$ получается более сложное выражение
$$p_n(-\log p'_n) + (1-p_n)(-\log(1-p'_n)),$$
где $p'_n = 1/2 + \delta'_n$; можно сказать, что фактические частоты нулей и единиц в $(\omega)_n$ равны p_n и $1-p_n$, но при «кодировании» используются не они, а приближённые (и упрощённые) p'_n и $1-p'_n$. Ясно, что это выражение лишь увеличится, если заменить в нём p_n на p'_n (поскольку $p_n > p'_n > 1/2$, вторая скобка с логарифмом больше первой, и увеличение её веса за счёт первой увеличивает всю сумму).

В итоге получаем оценку
$$n - O(1) \leqslant nh(p'_n, 1-p'_n) + (1+\varepsilon)\log\log n$$
для любого $\varepsilon > 0$ (при достаточно больши́х n) и отсюда, как и раньше, получаем
$$\delta'_n \leqslant (1+\varepsilon)\sqrt{\ln 2\cdot\log\log n/2n}.$$

Оценка для «истинного» δ_n лишь ненамного больше (в $1/0{,}9$ раза в нашем примере); поскольку вместо $0{,}9$ можно было взять сколь угодно близкое к единице число, получаем искомое утверждение (множитель $\ln 2$ превращает первый логарифм в натуральный, а второй можно заменить на натуральный, поскольку это поглощается множителем $(1+\varepsilon)$). ▶

$\boxed{\textbf{243}}$ Приведённое доказательство теоремы 156 не полностью использует случайность последовательности ω: покажите, что оно проходит для последовательностей, у которых $n - KM((\omega)_n) = o(\log\log n)$.

8.5. Последовательности без запрещённых подслов

8.5.1. Запрещённые и простые слова

Доказываемое в этом разделе утверждение интересно не столько само по себе, сколько как неожиданное применение колмогоровской сложности.

Теорема 157. *Пусть $\alpha < 1$ — некоторое положительное действительное число. Пусть для каждого n некоторые двоичные слова, общим числом не более $2^{\alpha n}$, объявлены запрещёнными. Тогда существуют число c и бесконечная последовательность нулей и единиц, не содержащая запрещённых подслов длиннее c.*

Например, можно считать запрещёнными слова, у которых (простая) колмогоровская сложность меньше αn, их как раз не больше $2^{\alpha n}$. Получаем такое следствие («лемма Левина», [39]):

Теорема 158. *Пусть $\alpha < 1$ — некоторое положительное действительное число. Тогда существует бесконечная последовательность нулей и единиц, в которой все подслова достаточно большой длины n имеют сложность не меньше αn.*

Поучительно сравнить это утверждение с критерием случайности для равномерной меры (теорема 94, с. 172). Там речь шла лишь о начальных отрезках, а не всех подсловах, зато монотонная сложность была не меньше $n - O(1)$ (для обычной сложности это даёт оценку $n - O(\log n)$). Как показывает следующая задача, для всех подслов (а не только для начал) такую более сильную оценку получить нельзя. (Это и не удивительно: в настоящей случайной последовательности должны появляться все подслова, в том числе и малой сложности.)

$\boxed{\textbf{244}}$ Для всякой последовательности ω нулей и единиц найдётся $\alpha < 1$ и бесконечно много различных подслов, у которых отношение сложности к длине не больше α. [Указание. Рассмотрим два случая. Если последовательность содержит в качестве подслов все двоичные слова, то доказывать нечего. Если же в неё не входит некоторое слово u длины k, то разбивая длинные слова на блоки длины k и кодируя их оптимальным образом (блок u никогда не встречается и код ему не нужен), мы получаем отношение сложности к длине не больше $(\log(2^k - 1))/k$.]

Доказательство теоремы 157 происходит в два этапа. Сначала мы докажем её частный случай, теорему 158, а затем удивительным образом окажется, что из него следует и общий случай.

◄ Для доказательства теоремы 158 рассмотрим произвольное β, для которого $\alpha < \beta < 1$. Пользуясь теоремой 71 (с. 129), найдём число N с таким свойством: ко всякому слову x можно приписать N битов так, чтобы префиксная сложность возросла от этого не менее чем на βN. Будем использовать это свойство многократно, начав с пустого слова: получится бесконечная последовательность блоков длины N, от добавления каждого блока префиксная сложность возрастает как минимум на βN. Отсюда следует, что сложность любой группы из подряд идущих k блоков не меньше $\beta kN - O(1)$. В самом деле, от добавления этой группы сложность растёт не менее чем на βkN, а $KP(xy) \leqslant KP(x) + KP(y) + O(1)$. Отсюда следует, что и для любого подслова u (не обязательно начинающегося и кончающегося на границе блока) сложность не меньше $\beta l(u) - O(1)$, поскольку изменение сложности и длины из-за граничных эффектов (обрезание по границе блока) есть $O(1)$. Остаётся вспомнить, что β мы взяли (по сравнению с α) с некоторым запасом, который покрывает и граничные эффекты, и разницу между обычной и префиксной сложностями. ►

245 Проведите аналогичное рассуждение, используя обычную сложность вместо префиксной. [Указание. См. задачу 46 на с. 52.]

246 Докажите утверждение задачи 47 (с. 52), заменив в нём обычную сложность начальных отрезков на префиксную.

◄ Теперь докажем теорему 157; проще всего это сделать, если воспользоваться релятивизованным вариантом сложности. Будем считать, что множество F запрещённых слов дано нам в качестве оракула, то есть всем рассматриваемым алгоритмам разрешается (в качестве внешней процедуры) узнавать про любое слово, запрещено оно или нет. Как обычно в теории алгоритмов, такая релятивизация ничему не мешает, и мы будем считать, что теорема 158 верна и для релятивизованной относительно F сложности.

Заметим, что при таком оракуле все запрещённые слова длины n имеют сложность (при известном n) не более $\alpha n + O(1)$, поскольку каждое слово может быть задано своим порядковым номером в списке запрещённых слов. Поэтому безусловная сложность запрещённого слова длины n не больше $\alpha n + O(\log n)$, и достаточно взять последовательность, у которой нет длинных подслов с удельной сложностью меньше β для некоторого $\beta > \alpha$. ►

Любопытно, что при желании можно вывести теорему 157 из теоремы 158, не используя релятивизации (см. [137]). А именно, справедливо следующее утверждение:

Теорема 159. *Если для некоторого рационального α и некоторого множества F запрещённых слов утверждение теоремы 157 не выполняется, то оно не выполняется и для некоторого разрешимого множества F.*

Ясно, что для разрешимого множества F релятивизация ничего не меняет и не нужна; ограничение рациональными α тоже, очевидно, несущественно.

◄ Пусть для некоторого $\alpha < 1$ и некоторого множества F утверждение теоремы 157 неверно. Тогда для всякого c можно указать множество F_c с такими

свойствами:

(а) F_c содержит только слова длиннее c;

(б) F_c содержит не более $2^{\alpha k}$ слов длины k (при любом k);

(в) всякая бесконечная последовательность содержит хотя бы одно подслово из F_c.

(В самом деле, можно в качестве F_c взять множество всех слов из F, имеющих длину больше c.)

Обычное рассуждение (лемма Кёнига, компактность) показывает, что тогда и всякая достаточно длинная последовательность содержит хотя бы одно слово из F_c, и потому множество F_c без ограничения общности можно считать конечным. Более того, его (точнее, какое-либо множество с указанными свойствами) можно найти перебором, так что оно вычислимо строится по c. (Переход к конечному множеству нужен, чтобы воспользоваться перебором.)

Построим теперь последовательность c_i, беря c_{i+1} больше длин всех слов в множестве F_{c_i}. Объединение соответствующих множеств F_{c_i} и будет разрешимым множеством, для которого не выполнено утверждение теоремы 157. ▶

Обратите внимание на схему рассуждения: мы предполагали, что объект с некоторым свойством существует, а затем замечали, что тогда можно перебором построить (возможно, другой) объект с таким свойством. Эта схема в той или иной форме ещё не раз нам встретится.

247 Докажите, что если существует бесконечная вправо последовательность, не содержащая подслов из некоторого множества F, то существует и бесконечная в обе стороны последовательность, не содержащая подслов из F. [Указание. В силу компактности оба эти свойства равносильны тому, что существуют сколь угодно длинные конечные последовательности, не содержащие подслов из F.]

Дж. Миллер [106] предложил прямое доказательство теоремы 157, в котором искомая последовательность строится слева направо, при этом мы следим за тем, чтобы некоторая явно определяемая величина («уровень тревоги») оставалась небольшой.

Фиксируем множество запрещённых слов F. Пусть дано некоторое слово x (уже построенное начало последовательности). Определим величину $w_c(x)$; чем она больше, тем мы ближе к появлению запрещённого слова. Здесь c — некоторый числовой параметр, который впоследствии мы выберем чуть больше $1/2$. Для каждого возможного вхождения запрещённого слова $z \in F$, которое пересекается с x (и совпадает на пересечении), мы считаем число k недостающих (в x) битов и добавляем к $w_c(x)$ число c^k. (Таким образом, $w_c(x)$ есть сумма по всем $z \in F$ и всем возможным местам их вхождения.)

При $c = 1/2$ величина $w_c(x)$ имеет простой вероятностный смысл: математическое ожидание числа вхождений запрещенных слов, пересекающих x, если к x справа добавить последовательность случайных битов. Из этой интерпретации ясно соотношение

$$w_{1/2}(x) = \frac{1}{2}w_{1/2}(x0) + \frac{1}{2}w_{1/2}(x1) - \sum_{z\in F}(1/2)^{l(z)}.$$

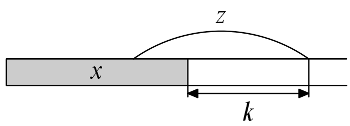


Рис. 8.3. Вхождение запрещённого слова z, часть битов которого уже есть в x, но недостаёт k битов, дает вклад c^k в $w_c(x)$.

В самом деле, с вероятностью $1/2$ мы начинаем с добавления бита 0, и с вероятностью $1/2$ добавляем бит 1. Надо учесть еще, что в правой части мы захватим и вхождения, примыкающие к x, но не пересекающие x, и их надо вычесть. Это равенство — чисто комбинаторный подсчёт, который можно выполнить и при произвольном c, и получится

$$w_c(x) = cw_c(x0) + cw_c(x1) - \sum_{z\in F} c^{l(z)}.$$

Изначально (на пустом слове) значение w_c равно нулю, а если в x уже встретилось запрещённое слово, то значение равно как минимум 1. Поэтому достаточно доказать, что можно поддерживать неравенство $w_c(x) < 1$, добавляя очередной бит. Для этого достаточно, чтобы

$$w_c(x0) + w_c(x1) = \frac{1}{c}\Bigl(w_c(x) + \sum_{z\in F} c^{l(z)}\Bigr)$$

было меньше 2 при $w_c(x) < 1$, что будет гарантировано, если $1 + \sum_{z\in F} c^{l(z)} < 2c$.

Для доказательства теоремы 157 осталось заметить, что сумму можно сделать сначала конечной, выбрав подходящее c (тем ближе к $1/2$, чем ближе α к единице), а затем сколь угодно малой (выбросив из F слова малых длин).

248 Покажите, что это рассуждение позволяет доказать эффективный вариант теоремы 157: если множество F (в дополнение к условию теоремы) разрешимо, то можно построить вычислимую последовательность, не содержащую коротких слов из F.

Предыдущую задачу можно обобщить (как сообщил авторам К. Макарычев) на двусторонние последовательности и найти вычислимую двустороннюю последовательность с тем же свойством. (Рассуждение следует идее Миллера, но надо ввести два параметра — левый и правый, и следить за их поведением, когда добавляется несколько букв: надо, чтобы один сильно уменьшался, а другой не сильно увеличивался.) Другой способ, описанный в [136], использует эффективный вариант леммы Ловаса и обобщается на многомерный случай.

8.5.2. Лемма Ловаса

Мы показали, как от сложностного утверждения перейти к комбинаторному аналогу. В этом разделе мы будем двигаться и в обратную сторону, начав с комбинаторного утверждения (локальной леммы Ловаса) и выведя из него следствия, касающиеся колмогоровской сложности. Но вначале несколько общих слов.

Вероятностные доказательства существования. Часто существование объекта, удовлетворяющего некоторому набору условий, доказывают так: на множестве всех объектов вводят распределение вероятностей, и для каждого условия подсчитывают вероятность того, что оно будет нарушено. Если эти вероятности малы — настолько малы, что их сумма меньше 1, — то случайный объект удовлетворяет всем условиям с положительной вероятностью и потому искомый объект существует.

Свойство вероятностей, используемое в этом рассуждении: если вероятность события A_i не больше ε_i, то вероятность объединения событий $A_1, \ldots, A_n$ не больше $\sum \varepsilon_i$, а вероятность того, что не произойдёт ни одного из них, не меньше

$$1 - \varepsilon_1 - \varepsilon_2 - \ldots - \varepsilon_n.$$

Вычисление вероятностей часто сводится к подсчёту количества «плохих» элементов в некотором множестве: если общее число плохих элементов меньше числа всех элементов, то есть и хорошие элементы. Как мы уже говорили, это можно перевести и на сложностной язык: если плохих элементов мало, то они имеют малую сложность, а потому несжимаемый элемент (элемент максимальной сложности) обязательно будет хорошим.

Если попытаться доказать теорему 157 с помощью такого рода подсчётов, то ничего не получится: вероятность обнаружить плохую строку в данной позиции действительно мала ($2^{(\alpha-1)n}$), но если позиций достаточно много, то сумма этих вероятностей уже станет больше единицы — а нам надо доказать существование последовательностей любой длины без плохих подслов. Спасительное наблюдение: если два участка случайной последовательности не пересекаются, то появления в них запрещённых подслов — независимые события.

Случай независимых событий. Для начала рассмотрим случай полностью независимых событий. Если события A_i независимы, и вероятность A_i равна ε_i, то вероятность того, что ни одно из A_i не произойдёт, равна

$$(1 - \varepsilon_1) \cdot (1 - \varepsilon_2) \cdot \ldots \cdot (1 - \varepsilon_n).$$

Отсюда, кстати, можно заключить, что это произведение больше $1 - \varepsilon_1 - \varepsilon_2 - \ldots$ (вариант неравенства Бернулли).

Видно, что для независимых событий эта вероятность будет положительной, даже если сумма всех ε_i больше единицы; достаточно, чтобы каждое из ε_i было меньше единицы.

Лемма Ловаса относится к промежуточной ситуации, когда событий слишком много (и потому не годится первая оценка), и они не все независимы (и потому не годится вторая оценка).

Пусть имеется n событий $A_1, \ldots, A_n$ и для каждого i от 1 до n фиксировано некоторое множество $N(i) \subset \{1, \ldots, n\}$, не содержащее i. Его элементы будем называть *соседями* i. (Никаких условий симметрии не предполагается: вершина может не быть соседом своего соседа.)

Предположим, что событие A_i независимо со всеми событиями, кроме себя и своих соседей (мы допускаем вольность речи и отождествляем событие A_i с

индексом i). Здесь имеется в виду именно независимость от набора всех несоседних событий, а не от каждого из них в отдельности. Тогда справедлива такая оценка:

Теорема 160 (локальная лемма Ловаса). *Если для каждого i выбрано положительное число $\varepsilon_i < 1$ и*

$$\Pr[A_i] \leqslant \varepsilon_i \prod_{j \in N(i)} (1 - \varepsilon_j),$$

то вероятность того, что не произойдёт ни одного из событий A_i, не меньше

$$(1 - \varepsilon_1) \cdot (1 - \varepsilon_2) \cdot \ldots \cdot (1 - \varepsilon_n).$$

Таким образом, мы получаем в точности ту же оценку, что и для независимых событий, но условие сильнее: за каждое событие-соседа j, с которым независимости нет, мы расплачиваемся множителем $(1 - \varepsilon_j)$ в правой части.

◀ Доказательство леммы Ловаса довольно странно — в нём вроде бы ни в какой момент ничего нетривиального не происходит, но придумать или даже запомнить его не так просто. Поэтому начнём для тренировки с простых замечаний.

(а) Если A и B — два события, то

$$\Pr[A \mid B] \leqslant \frac{\Pr[A]}{\Pr[B]}.$$

В самом деле, по определению условная вероятность $\Pr[A \mid B]$ есть $\Pr[A \wedge B] / \Pr[B]$, а $\Pr[A \wedge B] \leqslant \Pr[A]$. (Как обычно в логике, $\wedge$ означает «и».)

(б) Добавим всюду в предыдущем утверждение условие C («релятивизация»). Получим

$$\Pr[A \mid B \wedge C] \leqslant \frac{\Pr[A \mid C]}{\Pr[B \mid C]}.$$

В доказательстве леммы Ловаса это будет использоваться в ситуации, когда A и C независимы и потому числитель есть $\Pr[A]$, а знаменатель не слишком мал.

Теперь можно уже доказать лемму Ловаса по индукции. Как обычно, для индуктивного рассуждения полезно усилить утверждение. Будем доказывать такие факты ($\neg$ означает отрицание, то есть переход к дополнению):

(1) для любого i и для любых $p, q, \ldots$ (отличных от i и различных между собой) выполняется неравенство

$$\Pr[A_i \mid \neg A_p \wedge \neg A_q \wedge \ldots] \leqslant \varepsilon_i;$$

(2) для любых непересекающихся множеств индексов $i, j, \ldots$ и $p, q, \ldots$ выполняется неравенство

$$\Pr[\neg A_i \wedge \neg A_j \wedge \ldots \mid \neg A_p \wedge \neg A_q \wedge \ldots] \geqslant (1 - \varepsilon_i) \cdot (1 - \varepsilon_j) \cdot \ldots$$

Заметим, что из первого утверждения следует второе, если слева от черты (среди $i, j, \ldots$) только одно утверждение: если вероятность события при некотором

условии не больше ε_i, то вероятность его отрицания при том же условии не меньше $(1-\varepsilon_i)$.

Более того, это рассуждение можно продолжить и для большего числа событий в левой части:

$$\Pr[\neg A_i \wedge \neg A_j \,|\, \neg A_p \wedge \neg A_q \wedge \ldots] = \\ = \Pr[\neg A_i \,|\, \neg A_j \wedge \neg A_p \wedge \neg A_q \wedge \ldots] \cdot \Pr[\neg A_j \,|\, \neg A_p \wedge \neg A_q \wedge \ldots];$$

остаётся оценить каждый из сомножителей по (1).

С другой стороны, из второго утверждения можно вывести первое. Для этого разделим условия в (1) на две группы: входящие в $N(i)$ и не входящие в $N(i)$. (Здесь i — номер события, стоящего в левой части.) Пусть N и F — конъюнкции отрицаний соответствующих условий (соседних и остальных). Тогда, следуя ранее описанной схеме, можно оценить вероятность так:

$$\Pr[A_i \,|\, N \wedge F] \leqslant \frac{\Pr[A_i \,|\, F]}{\Pr[N \,|\, F]} = \frac{\Pr[A_i]}{\Pr[N \,|\, F]}.$$

Знаменатель в предположении (2) не меньше, чем произведение скобок $(1-\varepsilon_t)$ по всем $t \in N$, и остаётся посмотреть на предположение леммы, где при ε_i как раз стоит множитель, не больший этого произведения. (Если первая группа условий пуста, то под N следует понимать тривиально истинное условие: знаменатель в этом случае равен 1 и неравенство следует из условия леммы, а предположение (2) при этом не используется.)

Остаётся лишь объяснить, почему взаимное сведение (1) к (2) и обратно не приводит к порочному кругу. Когда мы сводим (1) к (2) с помощью только что рассмотренной оценки, общее количество событий в используемом неравенстве (2) по крайней мере на единицу меньше общего количества событий в доказываемом неравенстве (1) — исчезает событие A_i. Обратное сведение (2) к (1), с которого мы начали, не увеличивает общего количества событий: общее количество событий в каждом из используемых неравенств (1) не больше общего количества событий в доказываемом неравенстве (2). ▶

Вот пример комбинаторной задачи, где может быть полезна лемма Ловаса.

249 В каждой клетке конечной ленты мы хотим написать число от 1 до N. При этом для каждой границы между клетками некоторые пары чисел (l, r) запрещены, то есть нельзя, чтобы слева от границы стояло l, а справа r. Докажите, что если (для каждой границы) доля запрещённых пар среди всех пар (которых N^2) не больше $4/27$, то заполнение возможно. [Указание. Сопоставим каждой границе событие (запрещение пары для случайного набора чисел). Соседей у каждого события два, при $\varepsilon_i = \frac{1}{3}$ условие леммы выполняется.]

250 Докажите аналогичный результат (даже с лучшими параметрами) без ссылки на лемму Ловаса: если множество плохих пар имеет меру меньше $1/4$, то существует заполнение без плохих пар. [Указание. В каждой позиции имеются более половины чисел, для которых допустимы более половины правых соседей, и более

половины чисел, для которых допустимы более половины левых соседей. Значит, для хотя бы одного числа допустимы большинство соседей слева и большинство соседей справа. Взяв это число, можно дописывать к нему справа и слева числа из того же большинства (поскольку два множества, больших половины, обязательно пересекаются).]

8.5.3. Лемма Ловаса и запрещённые слова

Применим лемму Ловаса для доказательства теоремы 157. По соображениям компактности достаточно доказать существование сколь угодно длинных конечных последовательностей без запрещённых подслов.

Будем считать, что биты последовательности равновероятны и независимы. Появление запрещённой последовательности длины n в данной позиции (на данном интервале I) имеет вероятность $2^{(\alpha-1)n}$, где n — длина интервала. В качестве оценки в лемме Ловаса для этого события возьмём $2^{(\beta-1)n}$ для некоторого $\beta \in (\alpha, 1)$. Надо подобрать β так, чтобы выполнялось условие леммы Ловаса.

Соседями события на интервале I являются события на интервалах J, которые перекрываются с I. Поскольку вероятности событий зависят от длины, при подсчётах удобно группировать интервалы по длинам. Имеется $n+k-1$ интервалов J длины k, перекрывающихся с данным интервалом I длины n. Для каждого из них в правой части оценки леммы Ловаса появляется множитель $(1-2^{(\beta-1)k})$, и всего получается

$$(1-2^{(\beta-1)k})^{n+k-1}.$$

Теперь надо это перемножить по всем k, начиная с некоторого N (если мы доказываем существование последовательности без плохих подслов короче N). Таким образом, для применения леммы Ловаса нам нужно, чтобы

$$2^{(\alpha-1)n} \leqslant 2^{(\beta-1)n} \cdot \prod_{k \geqslant N} (1-2^{(\beta-1)k})^{n+k-1}.$$

(На самом деле в нашем подсчёте есть погрешность — при $n=k$ мы учитываем сам интервал в качестве соседа, но это в нашу пользу.) Нам достаточно очень грубой оценки: заменим $n+k-1$ на nk, извлечём корень n-степени и используем неравенство Бернулли в правой части. Останется доказать, что

$$2^{\alpha-\beta} \leqslant 1 - \sum_{k \geqslant N} k 2^{(\beta-1)k}.$$

Бесконечный ряд $\sum_k k 2^{(\beta-1)k}$ сходится при $\beta < 1$, а левая часть меньше 1 при $\alpha < \beta$, так что при достаточно большом N неравенство выполняется.

(Таким образом, порядок выбора параметров такой: берём любое $\beta \in (\alpha, 1)$, а затем подбираем N. Затем применяем лемму Ловаса к словам длины не меньше N и получаем, что существуют сколь угодно длинные конечные последовательности без запрещённых слов длины N и более (оценки выполнены при любой длине последовательности). Наконец, по компактности получаем бесконечную последовательность.)

$\boxed{\mathbf{251}}$ Докажите двумерный аналог теоремы 157: можно заполнить бесконечную клетчатую бумагу нулями и единицами так, чтобы любой прямоугольник достаточно большой площади не был запрещённым. (Предполагаем, что для каждого прямоугольника площади k выбрано не более $2^{\alpha k}$ запрещённых заполнений, где $\alpha < 1$ — некоторая константа.) [Указание: несложно провести аналогичные оценки.]

8.5.4. Запрещённые подпоследовательности

До сих пор мы говорили о запрещённых *подсловах*, то есть комбинациях подряд идущих битов. Это ограничение выглядит несколько искусственно. Более общая постановка вопроса, согласно [135], может быть такой: есть счётное число битовых переменных и некоторые ограничения; каждое ограничение запрещает конечному числу переменных принимать (одновременно) некоторые значения. Нас интересуют результаты такого типа: если ограничений «не слишком много», то они заведомо совместны (независимо от того, что конкретно запрещается).

На другом языке можно сказать, что мы хотим доказать выполнимость некоторой конъюнктивной нормальной формы (КНФ), то есть конъюнкции нескольких дизъюнктов (клауз). Например, в формуле

$$(\neg a \vee b \vee c) \wedge (a \vee c \vee \neg d) \wedge \ldots$$

первый дизъюнкт $(\neg a \vee b \vee c)$ запрещает комбинацию значений $a = 1$, $b = 0$, $c = 0$. Найти выполняющий набор значений (при котором формула истинна) означает удовлетворить всем ограничениям.

Два ограничения независимы, если не содержат общих переменных. Поэтому, имея в виду лемму Ловаса, имеет смысл ограничить число ограничений, включающих данную переменную.

Начнём с обозначений. Пусть $\omega = \omega_0\omega_1\omega_2\ldots$ — некоторая последовательность. Для произвольного конечного множества $F \subset \mathbb{N}$ индексов через $\omega(F)$ обозначим двоичное слово, составленное из членов последовательности с индексами в F (в порядке возрастания индексов). Рассмотрим пару (F, X), где F — конечное множество индексов, а X — двоичное слово, длина которого равна числу элементов в F. Последовательность ω *запрещается* парой (F, X), если $\omega(F) = X$. Пары такого вида будем называть *запрещениями*, а число элементов в F (длину слова X) — *размером* запрещения. Запрещение *покрывает* индексы, входящие в F. Теперь мы можем сформулировать и доказать такое утверждение [135]:

Теорема 161. *Пусть дано действительное число $\alpha \in (0,1)$ и множество запрещений (F, X), при этом для любого индекса i и числа n имеется не более $2^{\alpha n}$ запрещений размера n, которые покрывают i. Тогда существует число N и последовательность, не запрещённая ни одним из запрещений размера больше N.*

◀ По соображениям компактности достаточно доказать утверждение для конечных последовательностей (для некоторого N, не зависящего от длины последовательности).

Событием будет нарушение запрещения. Вероятность такого события для запрещения размера n есть 2^{-n}. Применяя лемму Ловаса, выберем в качестве соответствующего ϵ_i для события размера n значение $2^{-\beta n}$, где β — некоторая константа, бо́льшая α (как мы увидим, годится любая).

Соседями запрещения будут запрещения, пересекающиеся с ним (покрывающие общее число). Для применения леммы Ловаса надо взять запрещение размера n и проверить, что что 2^{-n} не больше $2^{-\beta n}$, умноженного на произведение множителей $(1-2^{-\beta m})$ по всем запрещениям, пересекающимся с данным.

Произведение разделим на части, соответствующие возможным точкам пересечения. Всего таких возможных точек n. (Если в пересечении несколько точек, произвольно выберем одну из них.) Кроме того, в каждой части расклассифицируем сомножители по размерам. Тогда для данной точки и для данного размера m получим не более $2^{\alpha m}$ сомножителей, каждый из которых есть $(1-2^{-\beta m})$. Дальше надо перемножить по всем m и возвести в степень n (поскольку частей n). Таким образом, нам надо проверить неравенство

$$2^{-n} \leqslant 2^{-\beta n} \prod_{m>N} (1-2^{-\beta m})^{2^{\alpha m} n},$$

или (если убрать возведение в степень n)

$$2^{\beta-1} \leqslant \prod_{m>N} (1-2^{-\beta m})^{2^{\alpha m}}.$$

По неравенству Бернулли это заведомо выполнено, если

$$2^{\beta-1} \leqslant 1 - \sum_{m>N} 2^{\alpha m} 2^{-\beta m}.$$

Поскольку левая часть меньше 1, а убывающая геометрическая прогрессия

$$\sum_m 2^{(\alpha-\beta)m}$$

сходится, то при подходящем N нужное неравенство выполнено. (Порядок выбора: сначала берём $\beta \in (\alpha, 1)$, затем выбираем N, глядя на сходящийся ряд, затем для любой длины последовательности применяем лемму Ловаса, и, наконец, ссылаемся на компактность.) ►

Другое доказательство теоремы 161, не ссылающееся на лемму Ловаса (а проводящее аналогичное ей рассуждение непосредственно), предложили Ан. А. Мучник и А. Л. Семёнов. Пусть фиксировано множество запрещений, удовлетворяющее предположениям теоремы; пусть N — минимальный размер запрещений, входящих в это множество.

Для каждого конечного множества индексов $I \subset \mathbb{N}$ через $c(I)$ обозначим количество разрешённых раскрасок I, то есть отображений I в $\{0,1\}$, которые не нарушают запрещений. (При этом рассматриваются только запрещения (F, X) с $F \subset I$, поскольку вне I отображение не определено и сравнивать его с запрещением нельзя.) Для пустого I естественно положить $c(I) = 1$.

Выберем $\beta \in (\alpha, 1)$ и будем доказывать, что при добавлении одной новой точки к множеству I величина $c(I)$ увеличивается в 2^β раз (в предположении, что N достаточно велико). Отсюда будет следовать, что число разрешённых раскрасок на множестве размера k не меньше $2^{\beta k}$ и уж заведомо положительно — нам нужно только это, но по индукции приходится доказывать более сильное утверждение.

Итак, пусть мы добавили к I новую точку i, получив $I' = I \cup \{i\}$. Каждая разрешённая раскраска I порождает две раскраски на I' (для двух возможных значений в новой точке). Получаем $2c(I)$ раскрасок, но часть из них может быть запрещёнными и их надо вычесть. Кто может их запрещать? Поскольку раскраска на I разрешена, то запрещение должно содержать i в дополнение к некоторым точкам в I; пусть K — множество этих точек, а k — их количество. Сколько раскрасок нужно вычесть из $2c(I)$ из-за одного такого запрещения? Их не больше, чем разрешённых раскрасок $I \setminus K$, а по предположению индукции таковых не больше $c(I)/2^{\beta k}$. (В самом деле, если добавление точки увеличивает число разрешённых раскрасок по крайней мере в 2^β раз, то удаление одной точки уменьшает это число в то же число раз, а удаление k точек уменьшает это число по крайней мере в $2^{\beta k}$ раз.) Теперь вычитаемое надо просуммировать по всем k от $N-1$ до I, а для каждого k по не более чем $2^{\alpha(k+1)}$ запрещениям, содержащим i и k элементов в I.

В итоге получаем оценку

$$c(I') \geqslant 2c(I) - \sum_{k=N-1}^{|I|} 2^{\alpha(k+1)} \frac{c(I)}{2^{\beta k}},$$

которую можно ослабить (заменяя 2^α на 2 и продолжая суммирование до бесконечности) до

$$c(I') \geqslant 2c(I)\left(1 - \sum_{k \geqslant N-1} \frac{2^{\alpha k}}{2^{\beta k}}\right).$$

Поскольку ряд в правой части сходится, то при достаточно большом N множитель в скобках не меньше $2^{\beta-1}$, и индуктивное рассуждение завершается. (Заметим, что мы применяли предположение индукции лишь к множествам размера меньше $|I|$, так что порочного круга не получается.)

8.5.5. Сложные подпоследовательности

Мы хотим перевести доказанную теорему на сложностной язык и доказать, что существует последовательность со сложными подпоследовательностями (а не только подсловами).

Можно ли надеяться, что сложность любой подпоследовательности достаточно большой длины m не меньше αm для некоторого $\alpha < 1$ (как это было для подслов)? Очевидно, нет: ведь можно включить в подпоследовательности те индексы, где стоят нули (или единицы). Но при этом сложность перейдёт в сложность набора индексов.

И действительно, мы уже встречали подобное утверждение в задаче 145 (для равномерной меры — пункт (д) теоремы 94, с. 172): если последовательность ω

случайна в смысле Мартин-Лёфа по равномерной мере, то

$$KP(F, \omega(F)) \geqslant |F| - c$$

для некоторого c и для всех конечных множеств F.

Но это не то, к чему мы стремимся (это скорее аналог теоремы о том, что у случайной последовательности сложные начальные отрезки, а не существования последовательности со сложными подсловами). Нас будет интересовать другое утверждение [135]:

Теорема 162. *Пусть $\alpha \in (0,1)$ — действительное число. Существует последовательность ω и константа N, для которых*

$$\max_{t \in F} KS(F, \omega(F) \mid t) \geqslant \alpha |F|$$

при всех F, содержащих не менее N элементов.

Заметим, что отсюда вытекает такое свойство последовательности ω: всякое конечное множество F размера не менее N имеет элемент t, для которого

$$KS(\omega(F) \mid F, t) \geqslant \alpha |F| - 2\,KS(F \mid t)$$

(константа 2 написана для простоты, её можно уменьшить). Опуская t в левой части, получаем, что для любого конечного F выполнено неравенство

$$KS(\omega(F) \mid F) \geqslant \alpha |F| - 2 \max_{t \in F} KS(F \mid t).$$

В частности, если индексы представить себе расположенными в плоскости, а в качестве F брать прямоугольники, то вычитаемое в правой части будет логарифмическим и потому его можно скомпенсировать изменением параметра α. Получаем утверждение задачи 251 (с. 279).

◀ Теорема 162 является сложностной переформулировкой теоремы 161. В самом деле, в ней рассматривается множество запрещений (F, Z), для которых неравенство $KS(F, Z \mid t) < \alpha |F|$ выполнено для всех $t \in F$. Значит, для любого индекса t количество таких запрещений размера k, у которых F содержит t, не превосходит $2^{\alpha k}$, что и составляет предположение теоремы 161. ▶

8.5.6. «Эффективное» доказательство леммы Ловаса

Вероятностные доказательства существования объекта с заданными свойствами не позволяют явно указать пример такого объекта (если не считать перебор явным описанием). Однако если вероятность появления нужного объекта очень близка к единице, можно пренебречь малой вероятностью ошибки и считать случайный выбор способом построения нужного объекта.

В этом смысле лемма Ловаса ещё менее «эффективна» — поскольку вероятность в ней экспоненциально мала (хотя и положительна), то непосредственное использование датчика случайных битов (с распределением, фигурирующим в лемме) ничего хорошего не даёт. Однако случайные биты можно использовать и более

хитрым способом. Пусть нам надо, скажем, построить двоичное слово, имея некоторый набор запрещений (другими словами, выполнить КНФ, см. раздел 8.5.4). Сначала выберем биты случайно. Небольшая доля запрещений при этом (скорее всего) нарушится. Возьмём какое-то из них и попробуем его исправить, выбрав значения входящих в него переменных заново (снова бросив монету для каждой из переменных). Скорее всего, это поможет (вряд ли второй раз получится тот же запрещённый набор значений). Правда, могут появиться новые нарушения. Что ж, повторим процесс: выберем нарушенное запрещение и заново выберем значения входящий в него переменных. И так далее.

Формально говоря, мы сначала случайно назначаем значения всем переменным, а потом повторяем такое действие: выбираем одно из нарушенных запрещений (скажем, первое в каком-то порядке, или по какому-то другому правилу), и присваиваем входящим в него переменным новые случайные значения. Так мы делаем до тех пор, пока нарушенных запрещений не останется. Удивительно, но этот простой способ приводит к цели, как недавно доказали Мозер и Тардош [111, 112]. (Удивительно и то, что раньше до этого простого способа не додумались, и доказывали более слабые варианты более сложными способами.)

Мы разберём упрощённое рассуждение, которое предполагает, что все запрещения одинакового размера, и что они исправляются в некотором определённом порядке (с помощью рекурсивных вызовов соседей) — тем более что его удобно изложить на языке колмогоровской сложности, следуя Фортноу.

Итак, пусть наша формула (КНФ) содержит n переменных и N дизъюнктов, каждый из которых имеет длину m, то есть включает ровно m переменных. Будем считать соседями дизъюнкты, имеющие общую переменную (и потому являющиеся зависимыми событиями). Пусть у каждого дизъюнкта не более t соседей. Мы утверждаем, что если t не слишком велико, то по лемме Ловаса формула выполнима.

Какие нужны ограничения? Поскольку все дизъюнкты одного размера, естественно выбрать в лемме Ловаса одно и то же значение ε. Оно должно удовлетворять неравенству

$$2^{-m} \leqslant \varepsilon(1-\varepsilon)^t$$

(слева стоит вероятность нарушения в данном дизъюнкте). Правая часть максимальна при $\varepsilon = 1/(t+1)$, но для упрощения формул положим $\varepsilon = 1/t$, получится $(1-1/t)^t/t$, то есть (с большой точностью) $1/et$. Другими словами, выполнимость гарантируется условием $t \leqslant 2^m/e$. В конструктивном доказательстве нам понадобится более сильное условие $t \leqslant 2^m/8$.

Теорема 163. *Существует вероятностный алгоритм, который ищет выполняющий набор для любой КНФ с n переменными и N дизъюнктами размера m, у каждого из которых менее $2^m/8$ соседей, за полиномиальное от $n+N$ время с вероятностью не менее $1/2$.*

(Как обычно, вероятность успеха можно приблизить к 1, если повторить алгоритм несколько раз.)

◀ Алгоритм использует рекурсивную процедуру $Fix(d)$ (где d — дизъюнкт):

для всех дизъюнктов d формулы:
  если d не выполнен, то $Fix(d)$

Имеется в виду, что все дизъюнкты просматриваются по очереди, и процедура *Fix* вызывается для тех из них, которые не выполнены к моменту её вызова. Корректность гарантирована, если процедура *Fix* обладает таким свойством: она делает переданный ей ложный дизъюнкт истинным, не портя те дизъюнкты, которые и так были истинны. (При этом другие ложные дизъюнкты могут стать истинными, это только лучше.)

Текст процедуры $Fix(d)$ тоже совсем прост:

выбрать новые случайные значения для всех переменных в d;
для всех соседей d' дизъюнкта d:
  если d' не выполнен, то $Fix(d')$

В принципе может оказаться, что новые случайные значения совпадут со старыми. Технически удобнее не повторять выборку, пока не выйдет что-то новое (как было бы естественно), а считать каждый дизъюнкт собственным соседом (тем самым включив повторный вызов в цикл, если к тому моменту надобность в нём не отпадёт). Корректность процедуры (при условии корректности вызовов): на первом шаге могут нарушиться лишь дизъюнкты, соседние к d, и все они потом будут исправлены, включая сам d.

Остаётся убедиться, что с большой вероятностью процесс закончится за полиномиальное время. Для этого посмотрим, как в нём используются случайные биты (представляя себе, что они заранее заготовлены). Сначала n битов используются в качестве начальных значений переменных, а затем при каждом вызове $Fix(d)$ очередные m битов используются для переменных, входящих в d. (Именно здесь удобно, что мы не повторяем выборку в том же вызове *Fix*, если новые значения совпали со старыми.)

Наблюдение: полный комплект случайных битов, использованных к данному моменту работы алгоритма, можно восстановить по текущим значениям переменных и по списку дизъюнктов, для которых вызывалась процедура *Fix*.

В самом деле, из текста алгоритма видно, что $Fix(d)$ вызывается только в случае, когда d не выполнен, что однозначно определяет значения переменных d до вызова. А значения их после вызова — те самые случайные биты. Значит, можно восстанавливать события с конца, и дойти до начала, получив начальные n случайных битов.

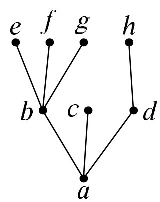
Таким образом, список дизъюнктов плюс текущие значения позволяет восстановить случайные биты. Что противоречит их несжимаемости, если список и текущие значения можно задать короче. Чтобы получить такое противоречие (при долго работающем алгоритме), надо оценить сложность списка дизъюнктов. Ключевую роль здесь играет то, что при рекурсивном вызове мы от дизъюнкта переходим к соседу, которого можно задать порядковым номером в списке соседей.[1]

[1] Удивительно, что утверждение, как доказали Мозер и Тардош, верно и при другом порядке обработки ложных дизъюнктов, но доказывать его надо иначе.

Итак, подсчитаем, сколько используется случайных битов и сколько нужно битов, чтобы задать их (описанным способом). Рассмотрим ситуацию после k вызовов процедуры *Fix*. К этому моменту использовано $n + km$ случайных битов. Чтобы восстановить их, нужно знать:

- текущие значений переменных;
- для каких дизъюнктов вызывалась процедура *Fix* из основной программы;
- какие вызовы процедуры *Fix* были сделаны рекурсивно для каждого из этих дизъюнктов.

Первое требует n битов. Второе требует не более N битов (для каждого дизъюнкта мы указываем, вызывался он или нет; порядок просмотра дизъюнктов в основной программе мы считаем фиксированным, так что его указывать не надо). Чтобы оценить третье, нарисуем деревья рекурсивных вызовов. Например, на рисунке показано дерево, начинающееся с вызова для дизъюнкта a. Этот вызов рекурсивно порождает три других, для дизъюнктов b, c, d; первый из трёх порождает вызовы для e, f, g, второй ничего не порождает, третий порождает вызов для h. Хронологический порядок вызовов такой: a, b, e, f, g, c, d, h (мы располагаем сыновей вершины слева направо в порядке их вызовов). В самом деле, мы переходим к вызову c, только вернувшись из вызова b, который рекурсивно включает в себя вызовы e, f, g.



Нам будет удобно представлять себе, что мы обходим граф, начав в точке a и держась за него правой рукой (как если бы это был вид сверху на стенку): $a - b - e - b - f - b - g - b - a - c - a - d - h - d - a$ (что соответствует передаче управления между копиями программ при рекурсивных вызовах). При этом новые случайные биты требуются, когда мы в первый раз (снизу) приходим в какую-то вершину.

Теперь видно, что для указания последовательности дизъюнктов достаточно закодировать обход дерева в указанном порядке. Он состоит из шагов вверх и вниз. Идя вверх, мы должны сообщить не только тот факт, что мы идём вверх, но и по какому ребру (к какому соседу) мы идём, что требует $\log t$ битов (плюс к одному биту, указывающему направление «вверх»). Здесь t обозначает $2^m/8$ — верхнюю оценку количества соседей. Идя вниз, мы можем не указывать соседа, поскольку он однозначно определяется предыдущей информацией. (При рекурсивном вызове мы помещаем параметр в стек, и надо указать, какой, а при выходе из процедуры мы удаляем вершину стека, и ничего указывать не надо).

В итоге на каждую вершину (кроме корня) уходит $\log t + 2$ бита (было бы $\log t + 1$, если были бы только шаги вверх, плюс по одному на шаги вниз, которых

не больше, чем шагов вверх). Таким образом, всего получается $N+n+k(\log t+2)$. Если случайные биты действительно имеют максимальную сложность, то $N+n+$ $+k(\log t+2)\geqslant n+km$, что даёт ограничение сверху на k, поскольку $\log t+2=$ $=m-1$, а именно $k\leqslant N$. То есть не более N вызовов процедуры *Fix*, что гарантирует полиномиальность алгоритма.

Правда, ещё нужно сделать некоторые уточнения и комментарии.

1. Если формально говорить о колмогоровской сложности, то появятся константы. Как обычно, проще с ними не связываться и перейти к подсчёту вероятностей: если $k=N+c$, то разница между левой и правой частью (числом случайных битов и числом битов в описании) есть c. Это означает, что количество исходов бросаний, при которых процедура *Fix* вызывается $N+c$ раз (или более), в 2^c раз меньше общего количества исходов, поэтому вероятность такого события не больше 2^{-c}.

2. Когда мы описываем с помощью битов какие-то объекты, надо проверить, что не требуется отдельно указывать разделители. В данном случае это действительно так: число переменных и дизъюнктов известно, а при описании деревьев мы сначала указываем, следует ли идти вверх или вниз, а после этого уже ясно, сколько битов дальше нужно читать.

3. Наконец, есть ещё одна проблема: вполне может оказаться, что мы прервали работу алгоритма в середине обхода одного из деревьев, так что мы должны уметь описывать незаконченный обход дерева. Но наш способ описания позволяет описать и незаконченный обход вокруг дерева с теми же оценками (число ходов вниз не превышает число ходов вверх, которое и есть число рекурсивных вызовов процедуры *Fix*). ►

8.6. Доказательство одного неравенства

Мы уже говорили (с. 22), что неравенства для колмогоровской сложности имеют разные любопытные следствия. Подробно этот вопрос разбирается в главе 10, но одно следствие подобного рода (вариант неравенства Лумиса–Уитни) мы приведём уже сейчас.

Теорема 164. *Пусть X, Y и Z — конечные множества, а $f\colon X\times Y\to\mathbb{R}$, $g\colon Y\times Z\to\mathbb{R}$ и $h\colon X\times Z\to\mathbb{R}$ — функции с неотрицательными значениями. Тогда*

$$\Bigl(\sum_{x,y,z} f(x,y)g(y,z)h(x,z)\Bigr)^2\leqslant\Bigl(\sum_{x,y} f^2(x,y)\Bigr)\cdot\Bigl(\sum_{y,z} g^2(y,z)\Bigr)\cdot\Bigl(\sum_{x,z} h^2(x,z)\Bigr).$$

◄ Как ни странно, это неравенство можно доказать с использованием неравенства

$$2\,KP(x,y,z)\leqslant KP(x,y)+KP(y,z)+KP(x,z)+O(\log n)$$

(теорема 26, с. 59). Нам удобно записать последнее неравенство для префиксной сложности, а не для обычной, как раньше, но это несущественно, так как разница есть $O(\log n)$. (На самом деле для префиксной сложности это неравенство верно и с точностью $O(1)$ (задача 114, с. 129), но для нас бóльшая точность не нужна.)

Будем для удобства считать, что множества X, Y и Z, о которых идёт речь в теореме, состоят из двоичных слов. Достаточно доказать, что если суммы в правой части неравенства равны единице, то сумма в левой части не превосходит 1. (В самом деле, от умножения функции f на число и левая, и правая части возрастают в одинаковое число раз, так что можно «нормировать» f; аналогично для g и h.)

Итак, пусть $\sum_{x,y} f^2(x,y) = 1$ и аналогично для двух других сумм. Нам надо доказать, что $\sum_{x,y,z} f(x,y)g(y,z)h(x,z) \leqslant 1$.

Идея проста: функция f^2 задаёт распределение вероятностей на парах (x,y), поэтому $KP(x,y) \leqslant -\log f^2(x,y) = -2\log f(x,y)$ (если временно считать, что это распределение вероятностей меньше априорного не с точностью до константы, а прямо так). Аналогично $KP(y,z) \leqslant -2\log g(y,z)$ и $KP(x,z) \leqslant -2\log h(x,z)$. Поэтому (вспоминаем неравенство для $KP(x,y,z)$ и временно забываем о логарифмической добавке)

$$2\,KP(x,y,z) \leqslant -2\log f(x,y) - 2\log g(y,z) - 2\log h(x,z),$$

то есть

$$f(x,y)g(y,z)h(x,z) \leqslant 2^{-KP(x,y,z)}.$$

Поскольку сумма $2^{-KP(x,y,z)}$ по всем тройкам x,y,z не превосходит единицы (теорема 57, с. 108), получаем искомое неравенство.

На самом деле, конечно, все наши оценки носят асимптотический характер, и потому для настоящего доказательства нам надо перейти от отдельных слов к наборам слов. Вот как это делается.

Для начала заметим, что нам (по непрерывности) достаточно доказать неравенство для функций f, g и h с рациональными значениями.

Пусть N — произвольное натуральное число (которое потом будет стремиться к бесконечности). Рассмотрим множества X^N, Y^N и Z^N, элементы которых являются кортежами из N слов. Рассмотрим распределение вероятностей на $X^N \times Y^N = (X \times Y)^N$, соответствующее N независимым копиям распределения f^2 на $X \times Y$. (Формально говоря, вероятность точки $\langle\langle x_1,\ldots,x_N\rangle,\langle y_1,\ldots,y_N\rangle\rangle$ равна произведению $f^2(x_1,y_1)\ldots f^2(x_N,y_N)$.) Мы получаем семейство распределений, вычислимо зависящее от N, и потому найдётся такая константа c, что

$$KP(\langle x_1,\ldots,x_N\rangle,\langle y_1,\ldots,y_N\rangle \mid N) \leqslant 2\sum_i(-\log f(x_i,y_i)) + c$$

при всех N и при всех $x_1,\ldots,x_N,y_1,\ldots,y_N$ (сравниваем построенное распределение с априорной вероятностью). Можно убрать условие N в левой части неравенства, написав справа $c\log N$ вместо c. Далее, как и раньше, сложим три таких неравенства и оценим сложность тройки через сложность пар. При этом получится

$$\begin{aligned}&KP(\langle x_1,\ldots,x_N\rangle,\langle y_1,\ldots,y_N\rangle,\langle z_1,\ldots,z_N\rangle) \leqslant\\ &\qquad\leqslant \sum_i(-\log f(x_i,y_i)) + \sum_i(-\log g(y_i,z_i)) + \sum_i(-\log h(x_i,z_i)) + c\log N\end{aligned}$$

для некоторой константы c и для всех N, $x_1,\ldots,x_N$, $y_1,\ldots,y_N$, $z_1,\ldots,z_N$ (заметим, что суммарная длина всех слов x_i,y_i,z_i при $i=1,\ldots,N$ есть $O(N)$, поэтому

все логарифмические добавки поглощаются $c\log N$). Подставляя эту оценку в неравенство $\sum_u 2^{-KP(u)} \leqslant 1$, получаем, что при любом N сумма

$$\sum\prod_i f(x_i,y_i)g(y_i,z_i)h(x_i,z_i)$$

по всем наборам $x_1,\ldots,x_N$, $y_1,\ldots,y_N$, $z_1,\ldots,z_N$ не превосходит $2^{O(\log N)}$, то есть некоторого полинома от N. Но эта сумма есть N-я степень суммы

$$\sum_{\langle x,y,z\rangle\in X\times Y\times Z} f(x,y)g(y,z)h(x,z),$$

поэтому если бы эта последняя сумма была больше 1, то получилось бы противоречие. Неравенство доказано. ►

$\boxed{\mathbf{252}}$ Покажите, что из доказанного неравенства следует оценка для объёма трёхмерного тела через площади его плоских проекций, упомянутая на с. 22. [Указание: в качестве f, g, h можно взять характеристические функции проекций тела; это даёт необходимую оценку в дискретном случае, а для непрерывного надо перейти к пределу либо разбивая тело на кубики, либо устремляя сумму к интегралу.]

Для сравнения приведём два других доказательства того же неравенства. Вот (простое) доказательство того же неравенства с помощью неравенства Коши (которое гласит, что $(u,v)^2 \leqslant \|u\|^2\cdot\|v\|^2$, то есть $\left(\sum u_iv_i\right)^2 \leqslant \left(\sum u_i^2\right)\left(\sum v_i^2\right)$). В самом деле,

$$\left(\sum_{x,y,z} f(x,y)g(y,z)h(x,z)\right)^2 \leqslant \left(\sum_{x,y} f^2(x,y)\right)\left(\sum_{x,y}\left(\sum_z g(y,z)h(x,z)\right)^2\right) \leqslant$$
$$\leqslant \left(\sum_{x,y} f^2(x,y)\right)\sum_{x,y}\left(\left(\sum_z g^2(y,z)\right)\left(\sum_z h^2(x,z)\right)\right) =$$
$$= \left(\sum_{x,y} f^2(x,y)\right)\left(\sum_{y,z} g^2(y,z)\right)\left(\sum_{x,z} h^2(x,z)\right).$$

Ещё одно доказательство может быть получено с использованием шенноновской энтропии (и в каком-то смысле представляет собой «перевод» доказательства с колмогоровской сложностью). Итак, пусть $\sum f^2 = \sum g^2 = \sum h^2 = 1$. Мы хотим доказать, что $\sum_{x,y,z} p(x,y,z) \leqslant 1$, где $p(x,y,z) = f(x,y)g(y,z)h(x,z)$. Рассуждая от противного, предположим, что эта сумма равна $c > 1$. Тогда пропорционально уменьшим её члены, получив некоторое распределение вероятностей p' на $X\times Y\times Z$:

$$p'(x,y,z) = \tfrac{1}{c} f(x,y)g(y,z)h(x,z).$$

Соответствующую случайную величину со значениями в $X\times Y\times Z$ обозначим через ξ. Её можно рассматривать как тройку (вообще говоря, зависимых) случайных величин ξ_x, ξ_y и ξ_z; можно также рассматривать двумерные проекции $\xi_{xy} = \langle\xi_x,\xi_y\rangle$ и т. п. Например, величина ξ_{xy} принимает значение $\langle x,y\rangle$ с вероятностью $\sum_z p'(x,y,z)$. Напомним, что по определению шенноновская энтропия распределения $(p_1,\ldots,p_k)$ равна сумме $\sum p_i(-\log p_i)$ и не превосходит $\sum p_i(-\log q_i)$

для любого другого распределения $q_1 + \ldots + q_k = 1$. Поэтому энтропию $H(\xi_{xy})$ можно оценить сверху, взяв в качестве другого распределения значения $f^2(x, y)$:

$$H(\xi_{xy}) \leqslant \sum_{x,y} \left(\sum_z p'(x, y, z) \right) (-2 \log f(x, y)).$$

Записав аналогичные оценки для двух других направлений проекции и применив неравенство

$$H(\xi) = H(\xi_x, \xi_y, \xi_z) \leqslant \tfrac{1}{2}(H(\xi_{xy}) + H(\xi_{yz}) + H(\xi_{xz}))$$

(задача 230, с. 253), получим, что

$$H(\xi) \leqslant \sum_{x,y,z} p'(x, y, z)(-\log f(x, y) - \log g(y, z) - \log h(x, z)) =$$
$$= \sum_{x,y,z} p'(x, y, z)(-\log p(x, y, z)).$$

Но по определению $H(\xi) = \sum_{x,y,z} p'(x, y, z)(-\log p'(x, y, z))$, и получается противоречие, так как p' меньше p в c раз (и потому $-\log p'$ больше $-\log p$ на $\log c$).

8.7. Нетранзитивность липшицевых преобразований

В этом разделе мы рассмотрим приложение колмогоровской сложности к анализу свойств бесконечных последовательностей. Начнём с такого определения, связанного с канторовским пространством Ω бесконечных двоичных последовательностей.

Отображение $f\colon \Omega \to \Omega$ называется *липшицевым*, если

$$d(f(\omega_1), f(\omega_2)) \leqslant c d(\omega_1, \omega_2)$$

для некоторой константы c и для всех $\omega_1, \omega_2 \in \Omega$. Здесь d — расстояние между последовательностями в канторовском пространстве, определяемое как 2^{-k}, где k — номер первого места, где последовательности различаются.

Неформально говоря, это свойство означает, что первые n знаков последовательности $f(\omega)$ определяются первыми $n + O(1)$ знаками последовательности ω. Поэтому, в частности, все отображения, задаваемые локальными правилами (каждый бит в $f(\omega)$ определяется некоторой окрестностью этого бита в ω), являются липшицевыми.

Нас будет интересовать такое свойство отображения f: для любых последовательностей ω_1 и ω_2 и любого $\varepsilon > 0$ существует число N и последовательности ω_1' и ω_2', для которых

$$\omega_2' = f(f(f(\ldots f(\omega_1') \ldots))) \quad (N \text{ раз})$$

и

$$d(\omega_1, \omega_1') < \varepsilon, \quad d(\omega_2, \omega_2') < \varepsilon.$$

(Другими словами, для любых двух открытых окрестностей найдётся орбита, начинающаяся в одной окрестности и попадающая в другую.) Это свойство мы будем для краткости называть «транзитивностью».

Легко понять, что отображение сдвига влево (отбрасывание первого члена последовательности) транзитивно: если мы хотим, чтобы последовательность начиналась на некоторое слово x_1, а после нескольких сдвигов начиналась на другое слово x_2, достаточно начать её с x_1x_2.

Возникает вопрос, сохранит ли сдвиг свойство транзитивности, если заменить d на так называемое *расстояние Безиковича*

$$\rho(\omega_1, \omega_2) = \limsup_{n\to\infty} d_n(\omega_1, \omega_2)/n,$$

где d_n — количество несовпадений среди первых n членов последовательностей, то есть количество тех $i < n$, при которых i-е члены ω_1 и ω_2 различны? Очевидно, что нет, поскольку расстояние Безиковича не изменяется при сдвиге последовательностей: $\rho(f(\alpha), f(\beta)) = \rho(\alpha, \beta)$. Поэтому любая окрестность последовательности из одних нулей (которая устойчива относительно сдвига) отображается сдвигом в себя. Значит с помощью сдвигов невозможно попасть из любой ее точки в достаточно малую окрестность любой далёкой от неё точки (например, последовательности из одних единиц).

Более того, имеет место следующий результат (мы воспроизводим доказательство из [10]):

Теорема 165. *Никакое липшицево отображение не является B-транзитивным.*

(Говоря о липшицевом отображении, мы имеем в виду исходное определение, с канторовским расстоянием.)

Причина этого проста: липшицево отображение почти не увеличивает сложности начальных отрезков последовательности, так как n битов выходной последовательности определяются $n + O(1)$ битами входной (и соответствующее правило можно считать вычислимым, выбрав подходящий оракул). С другой стороны, близкие в смысле Безиковича последовательности имеют близкие сложности начальных отрезков (поскольку изменение небольшой доли среди первых n битов кодируется малым количеством битов по сравнению с n).

◄ Удобно воспользоваться понятием эффективной хаусдорфовой размерности последовательности (которая равна $\liminf KS(\omega_0 \ldots \omega_{n-1})/n$ для одноэлементного множества $\{\omega\}$, см. теорему 120 на с. 198 в разделе 5.8).

Лемма 1. *Вычислимое липшицево отображение не увеличивает эффективную хаусдорфову размерность последовательности.*

(Говоря о вычислимости липшицева отображения $f\colon \Omega \to \Omega$, мы имеем в виду, что n первых знаков $f(\omega)$ алгоритмически определяются по $n + c$ знакам ω для некоторого c.)

В самом деле, если $f(\omega_1) = \omega_2$, то сложность начального отрезка длины n последовательности ω_2 не превосходит константы (зависящей от сложности алгоритма для f) плюс сложность начального отрезка длины $n + c$ последовательности ω_1, и в пределе эти константы не играют роли.

Лемма 2. *Если $\rho(\omega_1, \omega_2) < \varepsilon$, то эффективные хаусдорфовы размерности последовательностей ω_1 и ω_2 отличаются не более чем на $h(\varepsilon)$.*

(Здесь $h(\varepsilon)$ — шенноновская энтропия случайной величины с двумя значениями, имеющими вероятности ε и $1 - \varepsilon$.)

В самом деле, если первые n членов последовательностей ω_1 и ω_2 отличаются в k позициях, то сложности соответствующих начальных отрезков отличаются не более чем на сложность их разности (или суммы) по модулю 2, а последовательность длины n с k единицами имеет сложность не более $nh(k/n) + O(\log n)$ (см. раздел 7.3.1, теорема 146, с. 254); отсюда легко следует утверждение леммы.

Таким образом, если мы берём последовательность нулевой размерности (скажем, вычислимую), то любая близкая к ней последовательность (в смысле Безиковича) имеет малую размерность, а вычислимое липшицево отображение не увеличивает размерность, и поэтому могут получиться только последовательности малой размерности. С другой стороны, любая последовательность, близкая по Безиковичу к случайной последовательности (которая имеет размерность 1), имеет (по той же лемме 2) размерность, близкую к единице.

Таким образом, для вычислимых липшицевых отображений теорема доказана. Остаётся заметить, что все эти рассуждения можно релятивизовать относительно любого оракула и что любое липшицево отображение вычислимо относительно некоторого оракула. ▶

9. Частотный и игровой подходы к определению случайности

9.1. Исходный замысел фон Мизеса

Сейчас, когда все привыкли к построению теории вероятностей на основе теории меры, требуются специальные усилия, чтобы забыть об этом и попытаться оценить замысел Рихарда фон Мизеса, который в начале XX века предложил строить теорию вероятностей на основе понятия случайной последовательности, или, как он говорил, *коллектива*. Тем не менее попробуем.

В природе бывают явления, которые легко предсказывать (после того как изучены управляющие ими законы), например, движение планет. Но бывают и другие явления: как бы мы ни старались предсказать результат бросания монеты, обычно около половины предсказаний оказываются неверными. Именно такие явления и составляют предмет теории вероятностей.

Тем самым основным понятием теории вероятностей становится (более или менее неопределяемое) понятие *коллектива* — трудно предсказуемой последовательности символов (букв некоторого алфавита, который мы для простоты будем считать состоящим из нуля и единицы). А основным свойством коллектива $\omega = \omega_0\omega_1 \ldots$ является следующее свойство *устойчивости частот*:

> существует предел
>
> $$p = \lim_{n\to\infty} \frac{\omega_0 + \omega_1 + \ldots + \omega_{n-1}}{n}$$
>
> и, более того, это свойство сохраняется (с тем же p), если из всей последовательности выбрать некоторую подпоследовательность (скажем, оставить каждый второй член, или члены с составными номерами, или члены, стоящие после единиц).

Указанное p и называют *вероятностью* появления единицы (в данном коллективе).

Существование коллективов обосновывается «экспериментально», ссылкой на успешную деятельность игорных домов, которые бы разорились, если бы существовала система игры, позволяющая выбирать подпоследовательности, не обладающие свойством устойчивости частот.

Так — или примерно так — говорил фон Мизес (см., например, его книгу “Wahrscheinlichkeit, Statistik und Wahrheit” [108]; эта книга была переведена на русский язык под названием «Вероятность и статистика» — видимо, слово Wahrheit (истина) в названии показалось чрезмерно смелым посягательством на прерогативы

единственно верного и подлинно научного учения). Но говорил он это не во времена Евклида или Спинозы, а в начале XX века, когда требования математической строгости были достаточно настойчивы. Допустим, существование последовательностей с некоторыми свойствами можно объявить аксиомой или экспериментальным фактом, это ещё куда ни шло (хотя, конечно, экспериментальный факт, говорящий нечто о *пределах* последовательностей, вызывает сомнения, если не насмешки). Но уж сформулировать математически точно свойство устойчивости частот следовало бы.

Проблема здесь в том, что не указано, какие подпоследовательности разрешено выбирать. Сам Мизес ограничивался лишь некоторыми примерами допустимых правил выбора подпоследовательностей (три таких правила мы привели), а также замечал, что решение о включении или невключении какого-либо члена последовательности в выбираемую подпоследовательность не должно зависеть от значения этого члена. В самом деле, иначе мы могли бы из любой последовательности выбрать подпоследовательность из одних единиц (или нулей), нарушив свойство устойчивости частот.

Поэтому возможны разные уточнения замысла Мизеса, и несколько таких уточнений мы рассмотрим. Для каждого из них прежде всего возникает вопрос о существовании коллективов (который после уточнения определения становится математическим); при этом вопрос о том, действительно ли бросание монеты даёт коллектив, позволительно считать естественно-научным или философским и оставить в стороне.

Для простоты мы в основном ограничиваемся случаем симметричной монеты ($p = 1/2$), если обратное не оговорено явно. Будем называть бесконечную последовательность нулей и единиц *сбалансированной*, если предел частоты единиц (или нулей) в ней существует и равен $1/2$.

9.2. Правила выбора как множества слов

Пожалуй, самое естественное толкование допустимого правила выбора состоит в следующем: мы принимаем решение о включении ω_n в подпоследовательность, глядя на все предыдущие члены, то есть в зависимости от $\omega_0\omega_1\ldots\omega_{n-1}$. Таким образом, допустимое правило выбора есть функция, отображающая все двоичные слова $\omega_0\ldots\omega_{n-1}$ в двухэлементное множество $\{\text{включать}, \text{не включать}\}$. Другими словами, допустимое правило представляет собой некоторое множество R двоичных слов (соответствующих значению «включать»).

Формально говоря, для произвольного множества $R \subset \Xi$ мы определяем соответствующее ему правило выбора как отображение S_R, которое ставит в соответствие произвольной последовательности $\omega \in \Omega$ некоторую (конечную или бесконечную) подпоследовательность $S_R(\omega) \in \Sigma$. А именно, $S_R(\omega)$ состоит из тех ω_n, для которых $\omega_0\ldots\omega_{n-1} \in R$ (в том же порядке, в котором они шли в ω).

Пример: если принадлежность слова x к множеству R зависит лишь от длины слова x, то последовательность $S_R(\omega)$ получается из ω выбором членов ω_n, стоящих на заранее известных местах (при тех n, для которых слова длины n принад-

лежат R). Другой пример: правило «выбирать члены после единиц» соответствует множеству R, состоящему из всех слов, оканчивающихся на единицу.

Естественно ожидать, что если мы приходим в казино, где играют в орлянку, заранее наметив какое-то правило выбора такого рода, и делаем ставки в те моменты, когда правило выбора включает очередной член в подпоследовательность, то никакого преимущества мы не получим и предельная частота единиц в подпоследовательности будет равна $1/2$. Но это верно лишь при условии, что правило выбора фиксировано заранее: задним числом (зная последовательность) легко можно подобрать правило выбора, которое бы у этой последовательности выиграло. Другими словами, имеет место такой очевидный факт: для каждой последовательности ω существует множество R, при котором последовательность $S_R(\omega)$ состоит из одних нулей или одних единиц, и потому не сбалансирована. Поэтому нельзя определить коллектив как последовательность ω, для которой подпоследовательность $S_R(\omega)$ сбалансирована (или конечна) для всех множеств R: при таком определении коллективов вообще не будет.

Однако, как заметил Вальд [188], для любого *счётного* семейства правил выбора S_{R_i} (соответствующего счётному семейству множеств R_i) существует последовательность ω, которая обладает свойством устойчивости частот относительно всех правил этого семейства: при любом i последовательность $S_{R_i}(\omega)$ сбалансирована (или конечна).

Это легко следует из вероятностных соображений:

Теорема 166. *Пусть R — произвольное множество слов. Тогда множество тех последовательностей $\omega \in \Omega$, для которых $S_R(\omega)$ бесконечна, но не сбалансирована, имеет меру нуль* (*относительно равномерной меры на пространстве бесконечных последовательностей нулей и единиц*).

Согласно этой теореме каждое правило выбора отбраковывает лишь нулевое множество. Счётное число правил выбора даёт счётное семейство нулевых множеств, объединение которых будет нулевым. Значит, неотбракованные последовательности останутся (и даже образуют множество меры 1).

◄ Утверждение теоремы вытекает из усиленного закона больших чисел (согласно которому все последовательности, за исключением множества меры нуль, сбалансированы, см. раздел 3.2), а также следующей леммы.

Лемма. *Пусть $U \subset \Omega$ — множество меры нуль. Тогда его прообраз $S_R^{-1}(U)$ имеет меру нуль.*

Доказательство. В самом деле, каждый следующий бит последовательности $S_R(\omega)$ имеет равные шансы оказаться нулём и единицей (при известных предыдущих битах), и разница с равномерным распределением лишь в том, что он может вообще не появиться — но от этого вероятность только уменьшается.

(Наглядно это иллюстрируется старинной загадкой: изменится ли доля мужчин, если после рождения первого сына женщины больше не будут рожать детей, чтобы наследник был единственным? Ответ отрицательный по тем же причинам.)

Формально это рассуждение проводится так:

Рассмотрим множество Σ_x всех конечных и бесконечных продолжений слова x и два его подмножества Σ_{x0} и Σ_{x1}. Докажем, что прообразы этих подмножеств при

отображении S_R имеют равную меру (другими словами, что появление в подпоследовательности нуля и единицы после данного слова x одинаково вероятно).

В самом деле, рассмотрим все слова z, из которых при выборе по правилу S_R получается x и для которых $z \in R$. Они соответствуют ситуациям, когда слово x уже выбрано и вот-вот будет выбран следующий бит. Из определения следует, что два таких слова не сравнимы друг с другом. Поэтому множества Ω_{z0} для всех таких z не пересекаются, и, как легко понять, вместе составляют прообраз множества Σ_{x0}. Аналогично прообраз множества Σ_{x1} есть объединение непересекающихся множеств вида Ω_{z1}. Мы разбили прообразы на равные части и, значит, меры прообразов равны.

Из доказанного по индукции легко следует, что мера прообраза множества Σ_x не превосходит $2^{-l(x)}$. Поэтому прообраз нулевого множества $U \subset \Omega$ является нулевым. В самом деле, возьмём покрытие множества U интервалами Ω_{x_i} с малой суммой мер; рассмотрим прообразы множеств Σ_{x_i}; каждый из них есть объединение счётного числа интервалов. Объединив все эти интервалы, получим покрытие множества $S_R^{-1}(U)$ с малой суммой мер. Лемма (а с ней и теорема 166) доказана. ▶

Отметим ещё, что из доказанного стандартным для теории меры способом (для измеримого множества мера есть точная нижняя грань меры покрытий) следует, что

$$\mu(S_R^{-1}(U)) \leqslant \mu(U)$$

для любого измеримого множества $U \subset \Omega$. В случае, когда $S_R(\omega)$ бесконечно для всех (или почти всех) ω, верно и более сильное утверждение: $S_R(\omega)$ имеет равномерное распределение, то есть

$$\mu(S_R^{-1}(U)) = \mu(U)$$

для любого измеримого U. (Рассмотрим U и его дополнение.)

253 Пусть фиксировано некоторое множество R. Покажите, что если ω имеет бернуллиево распределение (испытания независимы и имеют одну и ту же вероятность успеха), то подпоследовательность $S_R(\omega)$ имеет то же распределение (предполагая, что она бесконечна с вероятностью 1).

Итак, понятие коллектива становится непустым (коллективы существуют), если из всех возможных правил выбора (соответствующих всем возможным подмножествам $R \subset \Xi$) выбрать некоторое счётное семейство. Но что это может быть за семейство?

9.3. Случайность по Мизесу – Чёрчу

С появлением теории вычислимых функций стало ясно, что естественно рассматривать правила выбора S_R, соответствующие всевозможным разрешимым множествам R. Это предложил А. Чёрч [32], поэтому такие правила выбора называют *допустимыми по Чёрчу*, а соответствующие последовательности (сохраняющие

устойчивость частот при всех допустимых по Чёрчу правилах) называют *случайными по Мизесу – Чёрчу*. В англоязычной литературе используют термин *Church stochastic*.

Мы уже знаем, что таковые существуют и образуют множество меры 1. Верно и более сильное утверждение:

Теорема 167. *Всякая случайная в смысле Мартин-Лёфа последовательность (относительно равномерной меры) случайна по Мизесу – Чёрчу.*

◀ Эффективный вариант усиленного закона больших чисел (теорема 32, с. 78; см. также раздел 8.4) говорит, что множество U несбалансированных последовательностей (у которых предел частот не существует или не равен $1/2$) является эффективно нулевым.

Покажем, что если S_R — допустимое по Чёрчу правило выбора, то прообраз эффективно нулевого множества будет эффективно нулевым. В самом деле, при разрешимом R конструкция в доказательстве теоремы 166 становится эффективной (все интервалы, составляющие прообраз данного интервала, можно эффективно перечислять). Поэтому случайная по Мартин-Лёфу последовательность не принадлежит этому прообразу, то есть её образ сбалансирован (или конечен). ▶

Что можно доказать про случайные по Мизесу – Чёрчу последовательности, кроме усиленного закона больших чисел, которому они удовлетворяют по определению? Например, можно доказать, что частоты появления не только букв, но и подслов таковы, какими они должны быть:

Теорема 168. *Пусть ω — случайная по Мизесу – Чёрчу последовательность, а U — некоторое двоичное слово. Рассмотрим те позиции k, начиная с которых U встречается в ω (то есть $U_0 U_1 \ldots = \omega_k \omega_{k+1} \ldots$). Тогда доля таких позиций среди первых N позиций стремится к $1/2^{l(U)}$ при $N \to \infty$.*

◀ По условию нули встречаются примерно в половине позиций. Рассмотрим правило «выбирать после нулей». По предположению в выбранной подпоследовательности будет примерно половина нулей, что означает, что после нулей одинаково часто бывают нули и единицы (то есть что частота групп 00 и 01 стремится к $1/4$). Теперь рассмотрим правило «выбирать после 00» (или после 01) и так далее. ▶

$\boxed{\textbf{254}}$ Докажите, что если случайную по Мизесу – Чёрчу последовательность разрезать на блоки длиной k, то каждый из 2^k возможных блоков будет встречаться с частотой, стремящейся к $1/2^k$. [Указание. Это утверждение отличается от предыдущей теоремы тем, что здесь учитываются блоки не на всех местах, а лишь на кратных k; его доказательство аналогично.]

Последовательности из теоремы 168, в которых каждая комбинация символов встречается с одинаковой частотой (в пределе), рассматривались и до Мизеса и назывались «нормальными». Действительные числа, двоичными записями которых являются нормальные последовательности, называются *нормальными по основанию* 2. Аналогично определяются действительные числа, *нормальные по основанию b* (*normal in base b*); числа, нормальные по всем основаниям, называются *абсолютно нормальными* (*absolutely normal*).

$\boxed{255}$ Покажите, что нормальность числа по основаниям b и b^k равносильны.

Нормальность по разным основаниям, вообще говоря, не равносильна, но это — нетривиальный теоретико-числовой результат [138], и мы не приводим его доказательства.

$\boxed{256}$ Будем рассматривать последовательность битов как двоичную запись числа в промежутке $[0,1]$. Тогда её хвосты образуют последовательность точек в $[0,1]$, являющуюся орбитой отображения $x \mapsto \{2x\}$, где фигурные скобки обозначают дробную часть. Покажите, что исходная последовательность битов нормальна тогда и только тогда, когда эта последовательность точек равномерно распределена на отрезке (в том смысле, что предельная доля точек, попадающих в некоторый подотрезок, пропорциональна его длине).

$\boxed{257}$ Покажите, что при умножение на целое число нормальность действительного числа сохраняется. [Указание. Используйте предыдущую задачу. В результате n-кратного применения отображения $x \mapsto \{2x\}$ к исходной точке x получается число $\{2^n x\}$. Ясно, что для любого целого k число $\{2^n(kx)\}$ получается из числа $\{2^n x\}$ преобразованием $y \mapsto \{ky\}$. Поэтому достаточно доказать, что последнее преобразование сохраняет равномерное распределение на отрезке $[0,1]$.]

Нормальность сохраняется и при делении на натуральное число (и, следовательно, при умножении на любое ненулевое рациональное число), как доказал Уолл [189] (см. также [71]); мы приводим и этот факт без доказательства.

Среди нормальных последовательностей есть и вычислимые: например, если выписать подряд числа $1, 2, 3, \ldots$ в двоичной записи и объединить цифры в одну последовательность, то получится нормальная последовательность

$$\underline{1}\,\underline{10}\,\underline{11}\,\underline{100}\,\underline{101}\,\underline{110}\,\underline{111}\,\underline{1000}\,\underline{1001}\,\underline{1010}\,\underline{1011}\ldots$$

(пример Чампернауна [26]; он рассматривал десятичную систему счисления, но большой разницы тут нет).

$\boxed{258}$ Докажите это. [Указание. Каков бы ни был размер блока k, начиная с некоторого места границы между записями натуральных чисел не сильно меняют частоты, а в среднем по всем натуральным числам заданной длины частоты правильны. (Тут нужна аккуратность, если мы останавливаемся посередине чисел некоторой длины.)]

Это построение годится только для одной системы счисления. Но существуют и числа, запись которых в любой системе счисления будет вычислимой нормальной последовательностью. (Это наблюдение было сделано ещё Тьюрингом в оставшихся неопубликованными заметках, см. [4].)

$\boxed{259}$ Докажите существование таких чисел. [Указание. Ненормальные в данной системе числа образуют нулевое в смысле Шнорра множество; объединение таких множеств по всем системам тоже будет нулевым по Шнорру множеством, и вне него существует вычислимая последовательность.]

В отличие от нормальных последовательностей, случайная по Мизесу – Чёрчу последовательность, разумеется, не может быть вычислимой, иначе было бы вычи-

слимо правило выбора, отбирающее члены на тех местах, на которых в последовательности нули (единицы). Более того, легко видеть, что верна такая

Теорема 169. *Для любого всюду определённого алгоритма, который пытается предсказывать следующий член последовательности по её предыдущим членам, доля успешных предсказаний на случайной по Мизесу – Чёрчу последовательности стремится к* $1/2$.

◀ В самом деле, каждому предсказывающему алгоритму соответствуют два правила выбора: одно отбирает те члены, где предсказан нуль, другое — где единица. Тем самым последовательность разбивается в «смесь» двух своих подпоследовательностей. Каждая из них сбалансирована (или конечна, но тогда всё очевидно), и потому на ней доля успешных предсказаний стремится к $1/2$. Значит, и общая доля успешных предсказаний стремится к $1/2$. ▶

Это утверждение можно ещё несколько обобщить. Представим себе такую игру: перед появлением следующего члена последовательности мы можем поставить некоторую сумму (неотрицательное рациональное число, не превосходящее единицы) на один из исходов (нуль или единицу). Если мы угадали, то ставка удваивается, если не угадали — пропадает. Наша стратегия в такой игре задаётся функцией S, определённой на двоичных словах и принимающей рациональные значения в отрезке $[-1,1]$. (Положительные значения соответствуют ставке на 0, отрицательные — на 1.) Суммарный выигрыш стратегии S на начальном отрезке $\omega_0 \ldots \omega_{n-1}$ можно записать как
$$\sum_{i=0}^{n-1} S(\omega_0 \ldots \omega_{i-1}) \cdot (-1)^{\omega_i}.$$
(Отрицательные значения соответствуют проигрышу — мы сейчас разрешаем игру в долг.)

Теорема 170. *Пусть S — всюду определённая вычислимая стратегия описанного вида, а ω — случайная по Мизесу – Чёрчу последовательность. Тогда выигрыш S на начальном отрезке длины n последовательности ω есть* $o(n)$.

◀ Пусть сначала стратегия принимает только значения 1 и -1. Тогда её деятельность сводится к уже рассмотренному угадыванию следующего члена; доля успехов стремится к $1/2$, что в точности и означает, что средний выигрыш (в расчёте на одну игру) стремится к нулю.

Если стратегия S принимает лишь значения, кратные $1/k$ для некоторого целого k (от -1 до 1 с шагом $1/k$), то её можно представить как среднее арифметическое $2k$ стратегий, принимающих значения -1 и 1; при этом выигрыш S будет средним арифметическим выигрышей всех этих стратегий, и если для каждой из них выигрыш есть $o(n)$, то и для среднего арифметического он будет равен $o(n)$. (Мы предполагаем, что число k фиксировано и не зависит от n.)

Осталось перейти к случаю вычислимой стратегии с произвольными рациональными значениями. Для произвольного $\varepsilon > 0$ мы должны показать, что выигрыш S на начальном отрезке длины n не превосходит (по модулю) εn для всех достаточно бóльших n.

Выберем k так, чтобы $1/k$ было меньше ε, и приблизим нашу стратегию S стратегией S' со значениями, кратными $1/k$ (заменив каждое значение на ближайшее кратное; ошибка будет не больше $\varepsilon/2$). Для стратегии S' выигрыш есть $o(n)$ и потому меньше $(\varepsilon/2)n$ для достаточно больших n, а разница между выигрышем для S и S' не превосходит $(\varepsilon/2)n$. ►

Ещё одно естественное свойство (упоминавшееся фон Мизесом как одно из основных свойств коллективов):

Теорема 171. *Применение допустимого по Чёрчу правила выбора к случайной по Мизесу – Чёрчу последовательности даёт либо конечную, либо случайную по Мизесу – Чёрчу последовательность.*

◄ В самом деле, последовательное применение двух правил выбора сводится к однократному: если мы сначала предварительно отбираем члены в подпоследовательность, а потом просматриваем члены подпоследовательности и выбираем из них только часть, то в конечном счёте решение определяется предыдущими членами. (А композиция разрешимых правил будет, очевидно, разрешимой.) ►

(В дальнейшем (раздел 9.12, с. 325) мы рассмотрим более общий класс правил (немонотонные правила выбора) и соответственно изменённое определение случайности по Мизесу. При этом окажется, что новый класс правил не замкнут относительно композиции, и, более того, соответствующий класс последовательностей не замкнут относительно правил выбора, см. теорему 203, с. 345.)

Естественный вопрос: как соотносится случайность по Мизесу – Чёрчу и Мартин-Лёфу? Как мы впоследствии увидим, случайность по Мартин-Лёфу — более сильное требование. Но сначала ещё несколько замечаний по поводу определения Мизеса.

9.4. Пример Вилля

Мы уже знаем, что для любого счётного семейства множеств R_i существует последовательность, которая обладает свойством устойчивости частот относительно всех правил выбора S_{R_i} — поскольку множество таких последовательностей имеет меру 1. Но можно дать и более прямую конструкцию такой последовательности, следуя работам А. Вальда [188], Ж. Вилля [180] и Д. Лавлэнда [88].

Для начала рассмотрим случай единственного правила S_R, соответствующего множеству R. В этом случае легко построить последовательность ω, для которой $S_R(\omega) = 01010101\ldots$ (нули и единицы чередуются) и потому последовательность $S_R(\omega)$ сбалансирована. В самом деле, будем строить ω постепенно (слева направо). Когда правило S_R предлагает выбрать очередной член подпоследовательности, мы смотрим, каким по счёту будет этот член среди выбранных (чётным или нечётным) и в зависимости от этого ставим в ω нуль или единицу. (Те члены ω, которые не выбраны правилом S_R, можно сделать любыми.)

Пусть теперь имеется конечное число множеств $R_1, \ldots, R_m$. Мы хотим построить последовательность ω, в которой выполнено свойство устойчивости частот относительно любого из правил S_{R_i}. Для каждого члена ω_n последовательности

ω посмотрим, какие из правил S_{R_i} (при $i = 1, \ldots, m$) включают его в подпоследовательность, и составим m-битовый вектор, содержащий эту информацию. Тем самым любая последовательность ω расслаивается на 2^m подпоследовательностей, соответствующих 2^m значениям этого битового вектора — мы включаем член ω_n в подпоследовательность, задаваемую m-битовым вектором v, если те (и только те) правила S_{R_i} выбирают ω_n, для которых i-й бит v равен 1. Некоторые из этих подпоследовательностей могут быть конечны.

Построим последовательность ω, для которой все эти 2^m подпоследовательностей будут иметь вид $010101\ldots$ (конечная или бесконечная последовательность, в которой нули и единицы чередуются, начиная с нуля). В самом деле, пусть $\omega_0 \ldots \omega_{n-1}$ уже известны. Как определить ω_n? К этому моменту уже ясно, какой битовый вектор соответствует ω_n (какие правила его выберут) и тем самым известно, в какую из 2^m подпоследовательностей попадёт ω_n. Остаётся посмотреть, каким по счёту (чётным или нечётным) членом этой подпоследовательности будет ω_n.

Заметим, что $S_{R_i}(\omega)$ будет смесью 2^{m-1} подпоследовательностей, а именно, тех, у которых в битовом векторе на i-м месте стоит единица. Поэтому $S_{R_i}(\omega)$ обладает свойством устойчивости частот (любой начальный отрезок последовательности $S_{R_i}(\omega)$ содержит не меньше нулей, чем единиц, а разница в количестве нулей и единиц ограничена числом 2^{m-1}, поскольку в каждой подпоследовательности разница не больше единицы).

Осталось рассмотреть общий случай счётного семейства правил R_i. Мы будем постепенно подключать эти правила, в каждый момент имея дело лишь с конечным числом правил. Вот как это делается.

Пусть уже построен некоторый начальный отрезок $\omega_0 \ldots \omega_{n-1}$ последовательности ω. Тем самым уже известно, каким из множеств R_i он принадлежит (и какие правила S_{R_i} выберут следующий, пока ещё не построенный, член ω_n последовательности). Теперь эта информация является не m-битовым вектором, а бесконечной последовательностью битов $u_1 u_2 \ldots$ (бит u_i равен единице, если правило S_{R_i} выбирает следующий член). Другими словами, у нас имеется путь $u_1 u_2 \ldots$ в бесконечном двоичном дереве.

Зафиксируем достаточно быстро растущую последовательность $k_0 < k_1 < k_2 < \ldots$ натуральных чисел. Например, пусть $k_i = 2^{2i}$. На каждом шаге конструкции (то есть для каждого члена последовательности) одна из вершин двоичного дерева будет считаться *активной.* А именно, двигаясь вдоль пути $u_1 u_2 \ldots$, найдём первую вершину, которая была активной менее k_i раз, где i — высота вершины в дереве, и объявим активной её. Другими словами, активной вершиной в момент построения ω_n будет кратчайшее слово x, для которого

- i-й бит x равен единице тогда и только тогда, когда правило S_{R_i} выбирает ω_n;
- до сих пор (в процессе построения начального отрезка $\omega_0 \ldots \omega_{n-1}$) вершина x была активна менее $k_{l(x)}$ раз.

Построенная таким способом последовательность $\omega_0\omega_1 \ldots$ разлагается в смесь счётного числа (конечных) подпоследовательностей, соответствующих счётному числу

возможных активных вершин. Подпоследовательность, соответствующая вершине x, имеет длину не более $k_{l(x)}$, но может быть короче.

Как и раньше, мы строим последовательность ω так, чтобы каждая из этих подпоследовательностей имела вид $010101\ldots$; теперь, правда, последовательность ω будет смесью *бесконечного* числа (конечных) последовательностей, и надо аккуратно проверить, что правило выбора S_{R_i} выберет из неё сбалансированную последовательность.

Выбранная правилом R_i подпоследовательность состоит из членов двух типов. Во-первых, она содержит члены, для которых активные вершины короче i (и правило S_{R_i} вообще не учитывалось при определении этих членов). Во-вторых, она содержит члены, которым соответствуют активные вершины, в которых i-й бит равен единице. Количество членов первого типа ограничено (не больше $2^0 k_0 + \ldots + 2^{i-1} k_{i-1}$) и в дальнейшем мы ими пренебрегаем.

Что касается членов второго типа, то для каждой использованной активной вершины числа нулей и единиц, ей соответствующих, отличаются не более чем на 1. Поэтому, если самая длинная активная вершина, встретившаяся при построении некоторого начального отрезка последовательности S_{R_i}, имеет длину $N \geqslant i$, то разница между числом нулей и единиц (второго типа) в этом начальном отрезке не больше числа активных вершин, использованных при его построении, то есть $O(2^N)$, а длина отрезка не меньше k_{N-1}, поскольку предшествующие активные вершины должны быть использованы полностью, прежде чем мы перейдём к более длинной. Остаётся заметить, что $2^N = o(k_{N-1})$.

Итак, мы описали явную конструкцию последовательности, обладающей свойством устойчивости частот для любого счётного семейства правил выбора. Что даёт нам эта конструкция (по сравнению с уже известным нам доказательством существования по соображениям меры)? Например, можно заметить, что поскольку каждая из смешиваемых последовательностей $010101\ldots$ начинается с нуля, то любой начальный отрезок их смеси содержит не меньше нулей, чем единиц. Отсюда получаем такое следствие (*пример Вилля* [180]):

Теорема 172. *Существует случайная по Мизесу – Чёрчу последовательность, в любом начальном отрезке которой не меньше нулей, чем единиц.*

Из этого можно было бы уже вывести, что существует последовательность, случайная по Мизесу – Чёрчу, но не по Мартин-Лёфу, если бы мы знали, что множество всех последовательностей, в которых все начальные отрезки содержат не меньше нулей, чем единиц, является эффективно нулевым. Это действительно так и следует из эффективного варианта закона повторного логарифма — но, к сожалению, не той его части, которую мы доказали в разделе 8.4 (теорема 156).

$\boxed{\textbf{260}}$ Покажите, что в данном случае достаточно знать, что указанное множество нулевое, и что из этого можно вывести, что оно эффективно нулевое. [Указание. Пусть p_n — вероятность того, что вплоть до длины n начальные отрезки содержат не меньше нулей, чем единиц. Очевидно, последовательность p_n убывает и вычислима, и её предел есть мера указанного множества, поэтому для любого рационального $\varepsilon > 0$ можно дождаться момента, когда p_n станет меньше ε. Можно также использовать понятие случайности по Курцу (с.83).]

$\boxed{\mathbf{261}}$ Докажите, что указанное множество является эффективно нулевым, не ссылаясь на закон повторного логарифма. [Указание. Для каждого n вероятность того, что в начальном отрезке длины n нулей не меньше, чем единиц, примерно равна $1/2$. Если взять быстро растущие значения n, то эти события близки к независимым (отклонение на меньшем отрезке мало по сравнению с ожидаемым отклонением на большем отрезке).]

Мы не будем проводить эти рассуждения подробно. Вместо этого мы покажем, что существуют случайные по Мизесу – Чёрчу последовательности с логарифмической сложностью начальных отрезков (что невозможно для случайных по Мартин-Лёфу последовательностей) с помощью той же конструкции.

Теорема 173. *Существует случайная по Мизесу – Чёрчу последовательность* $\omega = \omega_0\omega_1\ldots$, *для которой*

$$KS(\omega_0\ldots\omega_{n-1}) = O(\log n).$$

◄ Мы хотим применить описанное построение ко всем правилам выбора, используемым при определении случайности по Мизесу – Чёрчу, то есть взять в качестве R_i все разрешимые множества слов. При этом построение не удаётся сделать алгоритмическим, поскольку мы не можем эффективно перечислять разрешающие алгоритмы для всех разрешимых множеств. (Это не удивительно — иначе бы получилась вычислимая случайная по Мизесу – Чёрчу последовательность.)

Мы можем подряд перебирать все программы, но надо знать, какие из этих программ задают разрешимые множества (остальные программы можно пропустить, заменив каким-либо фиксированным правилом выбора). Такая информация для первых m программ занимает m битов (по одному для каждой программы) и позволяет проводить нашу конструкцию, пока мы не дойдём до активных слов длины m. А к этому моменту построенная последовательность будет иметь длину не меньше $k_{m-1} = 2^{2m-2}$. Тем самым объём дополнительной информации — логарифмический (по отношению к длине начального отрезка), что и требовалось. ►

Отметим ещё раз уже упоминавшееся следствие:

Теорема 174. *Существует случайная по Мизесу – Чёрчу последовательность, не являющаяся случайной по Мартин-Лёфу.*

Возникает естественный вопрос: не следует ли усилить требования к последовательности, предъявляемые определением случайности по Мизесу – Чёрчу? Например, можно разрешить переменные ставки, а также немонотонные правила выбора. В следующих разделах этой главы мы рассмотрим получающиеся при этом определения.

9.5. Мартингалы

Говоря о существовании коллективов, мы апеллировали к азартным играм. Но с этой точки зрения, надо признать, рассматриваемая ситуация выглядит довольно странно: клиент приходит в казино, где бросают честную монету, выбирает некоторые бросания и затем «выигрывает» (точнее, опровергает гипотезу о чест-

ности монеты), если результаты выбранных бросаний не имеют предельной частоты $1/2$.

Приближаясь к игровой терминологии, мы можем сказать, что клиент делает ставку фиксированного размера (скажем, рубль) в избранных раундах игры (при этом казино при необходимости верит ему в долг); требуется, чтобы его средний выигрыш (в расчёте на одну ставку) стремился к нулю. А если это не так, то клиент объявляет монету нечестной. (Как мы видели, ставки фиксированного размера можно заменить ставками ограниченного размера.)

Более естественным представляется другой вариант игры, который предложил Вилль. Пусть мы приходим в казино, имея рубль. Перед каждым бросанием мы делим наш капитал на две части, которые ставим на нуль и на единицу. Та часть, где мы не угадали, пропадает, а где угадали — удваивается, и мы продолжаем игру. (Например, если мы поделили капитал поровну между 0 и 1, то в любом случае останемся «при своих». Отсюда ясно, что отдельно предусматривать возможность вообще не ставить часть капитала не нужно.)

Наша стратегия в такой игре представляет собой функцию, которая (по начальному отрезку последовательности) говорит, сколько нужно ставить на нуль и сколько на единицу. Технически удобнее говорить о несколько другой функции. А именно, пусть $m(x)$ — капитал, который у нас будет после начального отрезка x при данной стратегии. Такая функция однозначно определяет стратегию: после появления слова x мы ставим $m(x0)/2$ на нуль и $m(x1)/2$ на единицу. При этом

- $m(\Lambda) = 1$ (в начале игры, при пустом слове, наш капитал равен единице);
- $m(x) = (m(x0) + m(x1))/2$ (мы ставим то, что у нас есть к данному моменту).

Функцию m, обладающую этими двумя свойствами, будем называть *мартингалом* относительно равномерной меры на пространстве нулей и единиц. (Впоследствии мы будем рассматривать мартингалы относительно других мер на Ω. В теории вероятностей рассматриваются и более общие виды мартингалов, но для наших целей бóльшая общность не потребуется.) Таким образом, вместо стратегий мы будем говорить о мартингалах, им соответствующих.

Пусть ν — произвольная мера на пространстве нулей и единиц. Легко проверить, что отношение $\nu(\Omega_x)/\mu(\Omega_x)$ (где μ — равномерная мера на Ω, а Ω_x есть множество всех продолжений слова x) есть мартингал, и что всякий мартингал соответствует некоторой мере ν.

262 Убедитесь в этом.

Имеет место следующий почти очевидный факт (называемый иногда *неравенством Колмогорова*):

Теорема 175. *Пусть фиксирован некоторый мартингал m и число k. Рассмотрим те слова, на которых значение мартингала не меньше k, и множество всех последовательностей, у которых есть начальный отрезок с таким свойством. Тогда (равномерная) мера этого множества не превосходит $1/k$.*

◀ Рассмотрим стратегию, соответствующую мартингалу m, и решим, что как только наш капитал достигнет (или превысит) k, мы заканчиваем игру и уходим из казино. Ясно, что в силу «честности» игры средний выигрыш при любой стратегии не больше 1, поэтому доля тех случаев, когда он больше или равен k, не превосходит $1/k$.

Формально это проще сказать на языке мер. Пусть $m(x)$ равно $\nu(\Omega_x)/\mu(\Omega_x)$ для некоторой меры ν (а μ — равномерная мера). Мы рассматриваем те x, у которых ν-мера конуса Ω_x превосходит μ-меру этого же конуса в k или более раз. Из них можно оставить только минимальные (не имеющие начал с тем же свойством), которым соответствуют непересекающиеся конусы. Суммарная μ-мера этих конусов по крайней мере в k раз меньше их суммарной ν-меры, которая не превосходит единицы. ▶

$\boxed{\textbf{263}}$ Покажите, что если мартингал m перечислим снизу, то точная верхняя грань его значений на начальных отрезках последовательности, рассматриваемая как функция последовательности, представляет собой ограниченный по вероятности тест в смысле раздела 3.5.

Верно и обратное утверждение: если есть множество малой меры, то можно построить стратегию (мартингал), которая много выигрывает у любой последовательности из этого множества.

Теорема 176. *Пусть имеется открытое подмножество $U \subset \Omega$ меры $\varepsilon > 0$. Тогда существует мартингал m с таким свойством: у каждой последовательности $\omega \in U$ существует начальный отрезок, на котором m не меньше $1/\varepsilon$.*

◀ Рассмотрим меру ν, для которой $\nu(X) = \mu(X \cap U)/\varepsilon$. (Вне множества U — нуль, внутри U — в $(1/\varepsilon)$ раз больше равномерной.) Отношение $m(x) = \nu(\Omega_x)/\mu(\Omega_x)$ и будет искомым мартингалом. В самом деле, если $\omega \in U$, то существует начало x последовательности ω, при котором $\Omega_x \subset U$ и $m(x) = 1/\varepsilon$.▶

Эту теорему можно объяснить следующим образом. Пусть сотрудники казино — жулики и торгуют, как теперь говорят, «инсайдерской информацией». А именно, они готовы заранее указать открытое множество U, в котором окажется будущая последовательность бросаний. Какова «рыночная цена» такой информации вместе с правом после этого начать игру с капиталом 1? Ответ: $1/\mu(U)$. Например, зная результаты первых N бросаний (что соответствует множеству меры $1/2^N$), мы можем N раз подряд выигрывать и в итоге получить 2^N. Доказанная теорема говорит, что это верно и для более сложно устроенных множеств U. Например, если известно, что некоторая последовательность невозможна («в нашем казино никогда не бывает N нулей подряд с начала игры»), то и это позволит гарантированно выиграть (правда, совсем немного, увеличив начальный капитал 1 до $2^N/(2^N - 1)$).

Легко понять, какова соответствующая стратегия игрока (из доказательства теоремы). Если гарантированное множество U на первом ходу делится в пропорции $a_0 : a_1$ (между последовательностями, начинающимися на нуль и на единицу), то мы должны ставить наш капитал в этой самой пропорции. (Например, если все элементы U начинаются на 0, то весь капитал нужно ставить на нуль.) Тогда

отношение

$$\frac{\text{текущий капитал}}{\text{доля } U \text{ среди продолжений текущей позиции}}$$

не меняется в процессе игры. В начале числитель равен 1, а знаменатель равен ε, а при попадании в открытое множество U (что обязательно случится, если только нам не продали ложную информацию) знаменатель равен 1, а потому числитель равен $1/\varepsilon$.

Аналогичные утверждения о связи мартингалов и мер можно сделать и в «предельном» случае. Будем говорить, что мартингал m *выигрывает* на последовательности ω, если значения m на начальных отрезках последовательности ω не ограничены. Следующая теорема Вилля была для него одной из главных мотивировок понятия мартингала (которое он и ввёл):

Теорема 177. (а) *Пусть m — произвольный мартингал. Тогда множество тех последовательностей $\omega \in \Omega$, на которых он выигрывает, имеет меру нуль.*

(б) *Пусть X — произвольное множество меры нуль. Тогда существует мартингал m, который выигрывает на всех последовательностях из X.*

◀ (а) Множество U_k тех последовательностей, на начальных отрезках которых мартингал достигает значений k или больше, имеет меру не более $1/k$ (и является открытым); все последовательности, на которых m выигрывает, принадлежат этому множеству.

(б) Для каждого k рассмотрим открытое множество U_k, имеющее меру не больше $1/k$, содержащее X, и соответствующий мартингал m_k, который гарантирует выигрыш не меньше k на всех элементах U_k (и тем самым на всех элементах X). Теперь из всех этих мартингалов нужно собрать один. Заметим, что взвешенная сумма мартингалов есть мартингал (мы можем разделить наш капитал на конечное или счётное число частей, и с каждой частью играть отдельно, применяя свой мартингал). Будем использовать мартингалы m_{4^n} с начальным капиталом 2^{-n} параллельно при всех n (разложив начальный капитал 1 в сумму ряда $\sum 2^{-n}$); тогда для последовательностей из U_{4^n} (и потому для всех последовательностей из X) гарантирован выигрыш $4^n \cdot 2^{-n} = 2^n$. Значит, общий выигрыш на любой последовательности из X не ограничен. ▶

Доказательство этой теоремы сильно напоминает доказательство критерия случайности (теорема 90, с. 167), который можно рассматривать как эффективный вариант только что доказанной теоремы.

По существу мы доказали более сильное утверждение. Будем говорить, что мартингал m *сильно выигрывает* на последовательности ω, если его значения на начальных отрезках последовательности ω стремятся к бесконечности. Построенный в теореме 177 мартингал, как легко видеть, *сильно* выигрывает на всех элементах X. (В самом деле, мартингал из доказательства теоремы 176 равен $1/\varepsilon$ на *всех* достаточно длинных начальных отрезках.)

Заметим, что теорему 93, с. 170 можно рассматривать как конструктивный аналог утверждения о существовании мартингалов, сильно выигрывающих у последовательностей из множества меры нуль. (О различных конструктивизациях понятий и результатов, связанных с мартингалами, мы ещё будем много говорить.)

Отметим ещё, что переход от выигрыша к сильному выигрышу можно провести и непосредственно:

Теорема 178. *Для всякого мартингала m существует мартингал m', который сильно выигрывает у всех последовательностей, у которых выигрывает мартингал m.*

◀ Мартингал m' должен действовать как запасливый игрок: достигнув выигрыша 2 (по m-стратегии), он откладывает половину «на чёрный день» (это означает, что эту часть выигрыша игрок ставит поровну на нуль и единицу), а с другой половиной поступает как m (но только все суммы вдвое меньше). Когда будет достигнут выигрыш 4 (что соответствует выигрышу 8 для m), откладываем половину (то есть 2) на чёрный день, а остаток снова пускаем в игру в соответствии с m, и так далее.

Можно привести и другое рассуждение (которое имеет то преимущество, что годится для перечислимых снизу мартингалов). Для каждого мартингала m и числа $c > 0$ рассмотрим мартингал m_c, который играет как m, пока и если капитал не достигнет c, после чего игра прекращается. Ясно, что m_c в пределе достигает c на всех последовательностях, где m хотя бы раз достиг c. Остаётся сложить m_c с весами, скажем, взяв сумму m_{4^k} с весами 2^{-k}. ▶

Более аккуратный выбор весов позволяет доказать общее утверждение [148, 35]: пусть $f\colon [1, +\infty) \to [0, +\infty)$ — неубывающая непрерывная функция, причём интеграл $\int_1^\infty f(t)/t^2\, dt$ не превосходит 1. Тогда для всякого мартингала m существует мартингал m' с таким свойством: если на какой-то последовательности мартингал m хотя бы раз был больше c, то с этого момента мартингал m' на той же последовательности будет больше $f(c)$. (Это утверждение точное: оценку на f нельзя ослабить.)

До сих пор мы предполагали, что вероятности выпадения нуля и единицы (декларируемые казино) равны, и потому ставки и на нуль, и на единицу удваиваются. Но это не обязательно. Пусть, например, вероятность появления нуля считается равной $1/3$, а вероятность единицы, соответственно, $2/3$. Тогда ставка на нуль должна утраиваться, а на единицу — увеличиваться всего лишь в полтора раза. Соответственно и определение мартингала изменится: $m(x)$ (капитал после x) должен быть равен сумме его части, поставленной на нуль (то есть $m(x0)/3$) и его части, поставленной на единицу (то есть $2m(x1)/3$):

$$m(x) = \frac{1}{3} m(x0) + \frac{2}{3} m(x1).$$

Это равенство можно прочесть и так: капитал до очередной игры равен математическому ожиданию капитала после этой игры.

Дадим формальное определение. Пусть π — произвольное распределение вероятностей на Ω (декларируемое казино распределение вероятностей для монеты). Соответствующую функцию на словах будем также обозначать буквой π, положив $\pi(x) = \pi(\Omega_x)$.

Функция $m(x)$, определённая на двоичных словах и принимающая неотрицательные действительные значения, называется *мартингалом относительно* π (с

единичным начальным капиталом — в дальнейшем это особо не оговаривается), если $m(\Lambda) = 1$ и

$$m(x)\pi(x) = m(x0)\pi(x0) + m(x1)\pi(x1)$$

при всех x. (Это определение соответствует предыдущему: поделив на $\pi(x)$, мы получим в правой части условные вероятности $\pi(x0)/\pi(x)$ и $\pi(x1)/\pi(x)$ появления нуля и единицы после x.)

Другими словами, функция $m(x)\pi(x)$ задаёт меру, так что мартингал относительно π (π-мартингал) — это отношение некоторой другой меры к π.

Далее всё аналогично случаю равномерной меры. А именно:

(1) вероятность того, что π-мартингал достигнет k на начальном отрезке последовательности ω, распределённой по мере π, не превосходит $1/k$;

(2) для всякого открытого множества U существует мартингал, который достигает значения $1/\pi(U)$ на всех последовательностях из U;

(3) множество X имеет π-меру нуль тогда и только тогда, когда существует π-мартингал, который выигрывает (вариант: сильно выигрывает) у любой последовательности из X.

Неравенство Колмогорова показывает, что любой мартингал ограничен на почти всех последовательностях. Верно и более сильное утверждение (*теорема Дуба*):

Теорема 179. *Для любого мартингала m множество тех последовательностей, вдоль которых его значения не имеют конечного предела, имеет меру нуль.*

◀ Согласно неравенству Колмогорова мартингал ограничен с вероятностью 1. Остаётся доказать, что для любых рациональных p, q с $0 < p < q$ множество тех последовательностей, на которых мартингал бесконечное число раз колеблется, переходя промежуток $[p, q]$ в обе стороны, имеет меру нуль. Это можно сделать, указав другой мартингал, который неограничен на всех таких последовательностях. А именно, будем следовать стратегии “buy low, sell high”, применяя стратегию исходного мартингала с того момента, когда он имел значение меньше p, до того момента, когда он приобрёл значение больше q (и сохраняя капитал в промежутках). На каждом таком участке мы наращиваем капитал в (q/p) раз, а участков бесконечно много по предположению. ▶

Эта теорема может быть использована для определения условной вероятности. Пусть дана некоторая мера μ на произведении $\Omega \times \Omega$. Тогда можно рассмотреть проекцию μ_1 меры μ на первую координату. Кроме того, возникает желание определить условную вероятность по второй координате при условии «первая координата равна α», где α — некоторая последовательность. Обычное определение условной вероятности тут не годится, поскольку само по себе событие «первая координата равна α» имеет меру нуль. Однако можно применить теорему Дуба, чтобы определить условную вероятность для почти всех α. Пусть A — произвольное событие по второй координате. Тогда можно определить условную вероятность A при условии «первая координата начинается на $a_0 \ldots a_{k-1}$». Заметим, что при фиксированном A и переменном начале мы получаем μ_1-мартингал (после нормировки), и по теореме Дуба предел существует почти всегда. (Возможен случай, когда некоторое

начало имеет вероятность нуль и условная вероятность не определена. Но тогда последовательности, имеющие это начало, и так образуют нулевое множество.)

В заключение отметим следующее почти очевидное обстоятельство:

Теорема 180. *Для всякого мартингала существует последовательность, у которой он не выигрывает* (*и даже не превосходит единицы на всех её начальных отрезках*).

◀ В самом деле, из определения следует, что одно из чисел $m(x0)$ и $m(x1)$ не больше $m(x)$, поэтому любое слово можно продолжить на один бит, не увеличивая мартингал. ▶

(Если крупье может манипулировать монетой, то он может всегда добиться, чтобы клиент ничего не выиграл.)

9.6. Отступление: мартингалы в теории вероятностей

Теорему 177 можно интерпретировать так:

(а) чтобы доказать, что какое-то множество нулевое, достаточно построить мартингал, который выигрывает на всех его элементах;

(б) этот метод можно применить к любому нулевому множеству (найдя соответствующий мартингал).

Эта интерпретация имеет глубокий смысл. Во-первых, с чисто технической точки зрения мы получаем удобный способ доказательства различных теорем теории вероятностей. Чтобы доказать, что некоторое множество имеет меру нуль, достаточно предъявить мартингал, который выигрывает на всех последовательностях интересующего нас множества.

В качестве иллюстрации изложим в таком стиле доказательство усиленного закона больших чисел. Рассмотрим бернуллиево распределение вероятностей B_p, в котором испытания независимы и вероятность появления единицы равна p.

Для данного $q > p$ докажем, что вероятность события «частота единиц в последовательности бесконечно много раз превосходит q» равна нулю. (Для $q < p$ и для события «частота меньше q» рассуждение симметрично.) Для этого рассмотрим мартингал (относительно меры B_p), равный отношению мер B_q/B_p. Другими словами, рассмотрим мартингал, который на последовательности z длины n с частотой единиц r (из nr единиц и $n(1-r)$ нулей) принимает значение

$$\frac{q^{nr}(1-q)^{n(1-r)}}{p^{nr}(1-p)^{n(1-r)}},$$

а его логарифм равен

$$n[(r\log q + (1-r)\log(1-q)) - (r\log p + (1-r)\log(1-p))].$$

Поскольку $q > p$, то это выражение является возрастающей (линейной) функцией от r: коэффициент при r равен $\log[q/p] + \log[(1-p)/(1-q)]$, и оба слагаемых положительны. Значит, при $r > q$ (интересующий нас случай) замена r на q лишь

уменьшит это выражение и логарифм мартингала не меньше

$$n[(q\log q + (1-q)\log(1-q)) - (q\log p + (1-q)\log(1-p)].$$

Согласно неравенству Гиббса (с. 241) коэффициент в квадратных скобках, то есть расстояние Кульбака – Лейблера между распределениями $(q, 1-q)$ и $(p, 1-p)$, положителен. Таким образом, построенный мартингал неограничен на интересующих нас последовательностях (где частота бесконечно много раз превосходит q).

Это доказательство усиленного закона больших чисел не совсем соответствует описанной схеме, так как для разных q получаются свои мартингалы. Каждый из них используется для оценки меры своего множества, а лишь затем мы замечаем, что объединение счётного числа нулевых множеств нулевое.

Можно действовать и в другом порядке: взять счётное семейство q_i, соответствующие мартингалы, а затем сложить эти мартингалы с положительными весами. Если хотя бы один из мартингалов неограничен, то и сумма будет неограниченной.

Примерно то же самое мы уже делали в разделе 3.2 (задача 67, с. 69), только говорили о конечных последовательностях и не употребляли слова «мартингал», ограничиваясь упоминаниями двух мер. (По существу те же рассуждения встретятся и в дальнейшем, в разделе 9.13.)

Во-вторых, можно смотреть на мартингалы с более философской точки зрения. Что значит доказать некоторую теорему теории вероятностей с помощью мартингала? Это значит предъявить некоторое свойство последовательностей L (скажем, закон больших чисел) и некоторый мартингал m и доказать, что *для любой* последовательности ω верно (хотя бы) одно из двух:

- последовательность ω обладает свойством L;
- мартингал m выигрывает на ω.

Двигаясь в этом направлении, можно предложить следующий «рыночный» (или «игровой») подход к понятию случайности и сказать, что

> *случайность последовательности нулей и единиц — это не свойство последовательности как таковой, а тип гарантии, с которой она продаётся.*

Что это значит? Представим себе продавца, который за рубль продаёт случайную последовательность на карточке. Покупатель может стирать краску, открывая бит за битом слева направо. (Последовательность считаем бесконечной.) Продавец даёт гарантию, что мартингал покупателя (копию которого покупатель отдаёт продавцу в запечатанном конверте в момент покупки) много не выиграет. Точнее, если этот мартингал в какой-то момент примет значение k на уже открытых битах, то покупатель может вернуть неиспользованные биты, истребовав с продавца k рублей неустойки. (Здесь существенно, что дальнейшие биты не известны покупателю, иначе он мог бы сжульничать: подглядеть вперёд и отказаться от использования невыгодных его мартингалу битов.)

Такой договор (как, наверно, сказали бы финансисты) позволяет «хеджировать» риски покупателя: если покупатель заправит купленную последовательность в свою

машину, и машина потерпит ущерб из-за того, что последовательность не удовлетворяет закону больших чисел, то продавец вынужден будет возместить ущерб согласно мартингалу, предусмотрительно заявленному покупателем при покупке. Для разных целей покупатель будет приобретать последовательности с разными мартингалами. (Конечно, на «практике» речь скорее будет идти о конечных последовательностях, но сама схема остаётся без изменений.) Скажем, если последовательность должна использоваться в вероятностном алгоритме получения большого простого числа (который может дать и составное число, но с очень малой вероятностью), покупатель может оговорить мартингал, который велик на последовательностях, приводящих к составным числам. Тогда, если покупатель пострадает из-за того, что число составное, то по крайней мере он сможет получить страховую выплату от поставщика случайности. Заметим, что при подписании гарантийного договора стороны оговаривают меру (поскольку от неё зависит класс мартингалов). Таким образом, при таком подходе мера из чего-то существующего в природе превращается в тип договора, заключаемого покупателем и продавцом. (Но, конечно, продавец должен учитывать даваемую гарантию при производстве последовательностей.)

Теорема 177, если следовать этой метафоре, показывает, что любой закон теории вероятности можно (более или менее естественным образом) уложить в эту схему.

Подробно такой подход к теории вероятностей обсуждается в книге Вовка и Шейфера [149].

9.7. Перечислимые мартингалы

Доказанные нами результаты о мартингалах имеют естественные эффективные аналоги. Нулевые множества, как мы видели, соответствуют мартингалам. Поэтому можно ожидать, что эффективно нулевые множества соответствуют некоторому классу «эффективных» мартингалов, и так оно и есть.

Мы будем рассматривать вычислимые меры и перечислимые снизу мартингалы относительно этих мер. Фиксируем некоторую вычислимую меру и будем рассматривать мартингалы (а также нулевые множества и случайные по Мартин-Лёфу последовательности) относительно этой меры, если не оговорено иное. При этом мы будем разрешать начальному капиталу (значению мартингала на пустом слове) быть произвольным числом, не превосходящим единицы. (Определение перечислимости снизу было дано в разделе 4.1; функция m перечислима снизу, если множество пар $\langle r, x\rangle$, для которых рациональное число r меньше $m(x)$, перечислимо.)

$\boxed{\mathbf{264}}$ Покажите, что перечислимый снизу мартингал m, для которого $m(\Lambda) = 1$, вычислим.

Вычислимых мартингалов нам недостаточно, поэтому мы и отказались от требования $m(\Lambda) = 1$. Теперь можно сформулировать эффективный вариант теоремы Вилля (теорема 177), предложенный Шнорром [143]:

Теорема 181. (а) *Пусть m — произвольный перечислимый снизу мартингал. Тогда множество тех последовательностей $\omega \in \Omega$, на которых он выигрывает, является эффективно нулевым.*

(б) *Пусть X — произвольное эффективно нулевое множество. Тогда существует перечислимый снизу мартингал m, который выигрывает на всех последовательностях из X.*

◄ Если мартингал перечислим снизу, то множество тех последовательностей, на которых он больше (целого) k, является эффективно открытым и имеет меру не больше $1/k$. (При этом мы имеем в виду именно меру множества, а не сумму мер возникающих в процессе перечисления интервалов, поскольку интервалы могут пересекаться: сначала появляется меньший интервал, а потом содержащий его больший. Но это не страшно, так как можно представить разность большего и меньшего как объединение конечного числа интервалов, и взять их.)

Наоборот, если множество эффективно открыто и имеет меру меньше $1/k$, то строимый по нему в доказательстве теоремы 176 мартингал будет перечислим снизу (появление нового интервала в множестве увеличивает мартингал). Надо только делить не на меру множества (она не обязательно вычислима, и помещение в знаменатель перечислимого снизу числа может сделать частное неперечислимым снизу), а на её верхнюю оценку $1/k$ (то есть умножать на k). При этом значение мартингала в корне может оказаться меньше единицы, но теперь это разрешено.

Остаётся (как в доказательстве теоремы 177) сложить построенные для разных чисел k мартингалы с надлежащими (вычислимыми) коэффициентами; это не нарушает перечислимость снизу. ►

Утверждение этой теоремы можно усилить, причём сразу в двух отношениях. Во-первых, как мы уже говорили, построенный мартингал будет *сильно* выигрывать на всех последовательностях из множества X. (Впрочем, это не особо важно, так как теорема 178, точнее, второе из её доказательств, позволяет преобразовать таким образом любой мартингал, сохраняя перечислимость снизу.)

Во вторых, можно рассматривать не только перечислимые снизу мартингалы, но и перечислимые снизу *супермартингалы*, или, как иногда говорят, *полумартингалы*. Они соответствуют играм, в которых игрок может отказываться от части выигранных денег. То есть, мы разрешаем неравенство

$$m(x) \geqslant (m(x0) + m(x1))/2$$

(для равномерной меры) или

$$m(x)\pi(x) \geqslant m(x0)\pi(x0) + m(x1)\pi(x1)$$

(для произвольной меры) вместо соответствующего равенства. Поскольку отказ лишь уменьшает капитал игрока, верхняя оценка вероятности выигрыша остаётся в силе, и в теореме можно заменить (перечислимые снизу) мартингалы на супермартингалы.

Из определения ясно, что супермартингалы — это отношения *полумер* к мере π, которая предполагается вычислимой. Поэтому, как мы знаем, среди перечислимых снизу супермартингалов существует максимальный (строго говоря, следовало бы сказать — наибольший)

$$m(x) = a(x)/\pi(x),$$

где a — максимальная перечислимая полумера на дереве (см. раздел 5.2), а $\pi(x)$ — вероятность множества Ω_x по мере π. Максимальность понимается с точностью до мультипликативной константы.

Отсюда получается новое доказательство теоремы Левина – Шнорра для априорной вероятности (теорема 91, с. 169): последовательность ω случайна по вычислимой мере π тогда и только тогда, когда на её начальных отрезках отношение $a(x)/\pi(x)$ ограничено.

9.8. Вычислимые мартингалы

С точки зрения стратегий игры в казино рассматривать перечислимые мартингалы странно: пропорция, в которой делается ставка, в этом случае представляет собой отношение перечислимых снизу чисел, и этому сложно придать естественный игровой смысл.

Поэтому интересно посмотреть, что получится, если ограничиться лишь вычислимыми мартингалами. Пусть фиксирована некоторая вычислимая мера π. Для простоты будем предполагать, что все значения $\pi(x) = \pi(\Omega_x)$ положительны (это нужно, поскольку эти значения стоят в знаменателях). В этом случае вычислимость π-мартингала равносильна вычислимости соответствующей стратегии в игре.

Будем называть последовательность ω *случайной по мере π относительно вычислимых мартингалов*, если никакой вычислимый π-мартингал на ней не выигрывает — другими словами, если любой вычислимый π-мартингал ограничен на её начальных отрезках. В англоязычной литературе для таких последовательностей употребляется термин *computably random*.

$\boxed{\textbf{265}}$ (**а**) Покажите, что это определение не изменится, если ограничиться мартингалами, отделёнными от нуля (скажем, потребовав, чтобы все значения были не меньше $1/2$); (**б**) если $\pi(\Omega_x)$ — рациональные числа, вычислимые по x (как конечные объекты), то можно ограничиться мартингалами, принимающими лишь рациональные значения (и вычислимыми в том же смысле). [Указание. (а) Можно взять среднее арифметическое любого мартингала и постоянного мартингала 1. (б) Если мартингал отделён от нуля, то можно приближать рациональными числами пропорцию, в которой делится капитал, с быстро возрастающей точностью.]

Как связан этот вариант понятия случайности с другими?

Первые два утверждения следующей теоремы относятся к общему случаю случайности по вычислимой мере π; два других — к случайности по равномерной мере (и легко обобщаются на бернуллиеву меру B_p с вычислимой вероятностью успеха p).

Теорема 182. (а) *Всякая случайная в смысле Мартин-Лёфа последовательность случайна относительно вычислимых мартингалов.*

(б) *Существует случайная относительно вычислимых мартингалов последовательность, начальные отрезки которой имеют логарифмическую сложность.* (*Отсюда следует, что предыдущее утверждение теоремы нельзя обратить.*)

(в) *Всякая случайная относительно вычислимых мартингалов последовательность случайна по Мизесу – Чёрчу.*

(г) *Не всякая случайная по Мизесу – Чёрчу последовательность случайна относительно вычислимых мартингалов.*

◀ (а) Случайность по Мартин-Лёфу гарантирует, что даже и перечислимые снизу мартингалы не выигрывают (теорема 181).

(б) Мы уже отмечали, что для любого мартингала можно найти последовательность, на которой он ограничен (достаточно из двух исходов игры выбирать тот, где капитал игрока не увеличивается).

Если мартингал вычислим, то можно найти вычислимую последовательность, на которой он ограничен. Это немного сложнее: мы не можем сравнить значения мартингала на двух продолжениях $x0$ и $x1$ слова x, так как знаем эти значения лишь с некоторой точностью. Но это и не нужно, достаточно на n-м шаге выбрать такое продолжение, где мартингал растёт менее чем на $1/2^n$ (а это можно сделать, находя всё более точные приближения и ожидая приближения, которое это гарантирует).

(Кстати, отсюда немедленно следует, что среди вычислимых мартингалов не существует максимального. Именно поэтому переход к перечислимым снизу мартингалам или супермартингалам существен.)

Следующий шаг: пусть даны два вычислимых мартингала; как найти вычислимую последовательность, на которой они оба ограничены? Для этого достаточно взять взвешенную сумму этих мартингалов, в которую они оба входят с положительным коэффициентом (например, полусумму). Получится снова вычислимый мартингал. Если хотя бы один из двух мартингалов не ограничен на некоторой последовательности ω, то и их полусумма не ограничена на ω (напомним, что мартингалы всегда неотрицательны), поэтому можно применить предыдущее утверждение.

То же рассуждение годится, если есть вычислимая последовательность вычислимых мартингалов (точнее, программ, эти мартингалы задающих). Тогда их можно сложить, взяв i-й мартингал с весом 2^{-i}.

Все вычислимые мартингалы нельзя расположить в вычислимую последовательность (это, кстати, следует из только что нами доказанного: ведь тогда можно было бы найти вычислимую последовательность, на которой все вычислимые мартингалы ограничены). Поэтому построение последовательности ω, на которой все вычислимые мартингалы ограничены, не может быть алгоритмическим. Дополнительная информация, которая нам нужна — это сведения о том, какие из программ задают мартингалы, а какие нет, то есть по биту на программу. Чтобы получить последовательность с логарифмической сложностью, мы должны эти биты использовать очень понемногу, включая в рассмотрение i-ю программу, когда последовательность будет уже достаточно длинной (экспоненциальной от i длины или ещё длиннее).

Опишем этот процесс подробнее. В каждый момент мы имеем некоторое слово x, а также некоторую линейную комбинацию с положительными коэффициентами, $m_1(x) + \varepsilon_2 m_2(x) + \ldots + \varepsilon_k m_k(x)$, где m_i — мартингал, задаваемый i-й программой (или его заменитель — скажем, тождественный нуль, если дополнительная информация об i-й программе говорит, что она не задаёт мартингал). При этом

мы следим, чтобы это выражение (для текущего слова x) было строго меньше 2. (Вначале, когда у нас только m_1 и слово x пусто, это выражение равно 1.)

Как мы видели, слово x всегда можно продолжить на один бит так, чтобы это выражение оставалось меньше двух (и такое продолжение можно алгоритмически найти, если мы знаем программы для мартингалов). Также время от времени мы будем добавлять новый член $\varepsilon_k m_k(x)$ к сумме, подбирая ε_k настолько малым, чтобы сумма осталась меньше 2 (чем ближе текущая сумма к 2 и чем больше значение $m_k(x)$ для текущего слова x, тем меньше берётся ε_k).

На полученной последовательности все m_i будут ограничены, так как каждый из них входит пусть с малым, но с постоянным ненулевым коэффициентом.

Сложность разрешения начального отрезка такой последовательности не превышает числа использованных битов дополнительной информации, и может быть очень мала, если мы добавляем новые мартингалы редко. А обычная колмогоровская сложность отрезков длины n будет $O(\log n)$, что и требовалось.

(в) Вспомним, что усиленный закон больших чисел говорит, что некоторое множество нулевое, и потому существует соответствующий мартингал (см. раздел 9.6, где мы строили мартингалы для каждой границы, а потом их смешивали). Ясно, что получающийся при этом мартингал можно сделать вычислимым (напомним, что p вычислимо).

Более того, если R — некоторое множество слов, а S_R — соответствующее правило выбора, то легко построить мартингал, который выигрывает на любой последовательности ω, для которой подпоследовательность $S_R(\omega)$ не сбалансирована. В самом деле, нужно играть лишь с теми членами последовательности, которые отбираются правилом S_R (а в остальных случаях сохранять значение мартингала неизменным, пропуская ход).

При этом, если множество R разрешимо, то мы получаем вычислимый мартингал, который выигрывает на любой последовательности ω, для которой подпоследовательность $S_R(\omega)$ бесконечна, но не сбалансирована.

Поэтому для всякой неслучайной по Мизесу – Чёрчу последовательности существует вычислимый мартингал, который на ней не ограничен.

(г) Рассмотрим (для случая равномерной меры) случайную по Мизесу – Чёрчу последовательность, у которой любой начальный отрезок содержит не меньше нулей, чем единиц (теорема 172). Пусть p_n — вероятность того, что все начальные отрезки длины не более n содержат не меньше нулей, чем единиц. Как мы уже обсуждали (задачи 260, 261), последовательность p_n вычислима, убывает и стремится к нулю. Для каждого n можно эффективно указать мартингал M_n, который выигрывает $1/p_n$ у любой последовательности, начальные отрезки (вплоть до длины n) которой содержат не меньше нулей, чем единиц. Взяв теперь взвешенную сумму некоторых M_n (выбранных так, чтобы выигрыши росли быстрее, чем убывали коэффициенты), получим вычислимый мартингал, который не ограничен на любой последовательности, у которой во всех начальных отрезках не меньше нулей, чем единиц.

Это же рассуждение можно после некоторых модификаций использовать для произвольной бернуллиевой меры (одновременно модифицировав конструкцию те-

оремы 172). Можно также применить конструкцию с двумя мерами (см. ниже раздел 9.13). ►

266 Укажите мартингал, описанный в доказательстве п. (г), явно. [Указание. Пусть мы пришли в казино и нам говорят, что в любом начальном отрезке игры не меньше орлов, чем решек. Тогда мы можем ставить на орла фиксированную сумму, зная, что не разоримся. Если разница между орлами и решками стремится к бесконечности, это и будет выигрышной стратегий. Если это не так, найдётся такой момент t и такое число n, что с момента t разница между орлами и решками не меньше n и бесконечно много раз равна n (нижний предел). Тогда после t, когда разница равна n, мы можем смело ставить на орла, не рискуя проиграть. Получаем один мартингал для первого случая и семейство мартингалов для второго с параметрами n, t. Остаётся соединить их в один мартингал.]

Заметим, что из утверждений (б) и (в) вытекает, что существуют случайные по Мизесу – Чёрчу последовательности с логарифмической сложностью начальных отрезков. Тем самым мы получили другое доказательство теоремы 173 (следуя работе [102]).

267 Докажите следующее усиление п. (б): если f — произвольная неубывающая неограниченная вычислимая функция, то существует случайная относительно вычислимых мартингалов последовательность ω, для которой $KS(\omega_0 \ldots \omega_{n-1} \,|\, n) \leqslant f(n) + O(1)$ при всех n. [Указание: новый мартингал, обходящийся в один бит дополнительной информации, мы добавляем, когда увеличившееся значение функции f это позволяет.]

Более того, можно построить одну последовательность ω, которая случайна относительно вычислимых мартингалов и для которой указанное неравенство выполнено для любой неубывающей неограниченной вычислимой функции f [102].

268 Покажите, что для вычислимой меры P на пространстве $\Omega \times \Omega$ и для любой последовательности ω, случайной относительно вычислимых мартингалов по первой проекции меры P, можно корректно определить условную вероятность, используя вычислимый вариант теоремы Дуба (с. 307).

В заключение раздела отметим ещё, что всё сказанное о мартингалах можно перевести на язык стратегий — алгоритмов, которые смотрят на уже происшедшие испытания и определяют, в какой пропорции (рациональной) нужно делать ставки на следующем шаге. (Мы предполагаем, что мера всех интервалов положительна, и мартингалы отделены от нуля — этого можно добиться, беря среднее с единичным мартингалом.)

9.9. Мартингалы и случайность по Шнорру

Понятие случайности относительно вычислимых мартингалов тесно связано с понятием случайности по Шнорру (см. раздел 3.4). Оба этих понятия были введены и изучены в книге Шнорра [143]. Там же была доказана и следующая теорема.

Теорема 183. *Пусть дана некоторая вычислимая мера π, относительно которой все интервалы Ω_x имеют положительную меру. Последовательность ω*

не случайна по Шнорру относительно π тогда и только тогда, когда существует вычислимый π-мартингал m и вычислимая неубывающая неограниченная функция $g\colon \mathbb{N}\to\mathbb{N}$ с таким свойством:

$$m(\omega_0\omega_1\ldots\omega_{n-1})\geqslant g(n)$$

для бесконечно многих n.

Эта теорема говорит, что неслучайность по Шнорру влечёт не только существование вычислимого мартингала, выигрывающего на последовательности, но и некоторую нижнюю оценку скорости выигрыша (справедливую для бесконечно многих начальных отрезков).

◄ Пусть последовательность ω не случайна по Шнорру. Тогда, как мы видели в разделе 3.4 (задача 90, с. 83), существует последовательность слов $x_0, x_1, x_2\ldots$, для которых ряд $\sum\pi(x_i)$ вычислимо сходится, причём бесконечно многие среди x_i являются началами ω.

Разобьём члены ряда

$$\pi(x_0)+\pi(x_1)+\pi(x_2)+\ldots+\pi(x_i)+\ldots$$

на группы из конечного числа подряд идущих членов. Сделаем это так, чтобы сумма членов k-й группы была не больше 4^{-k} (отбросим начальный отрезок ряда, если это необходимо). Поскольку ряд вычислимо сходится, это можно сделать алгоритмически. Можно считать дополнительно, что слова разных групп разделены по длинам: есть вычислимая последовательность чисел $n_0<n_1<n_2<\ldots$, и слова k-й группы имеют длину в интервале $[n_k, n_{k+1})$. В самом деле, любое слово в покрытии Соловея можно заменить на набор всех его продолжений большей длины, и мы можем по очереди обрабатывать слова разных групп, используя длины больше длин предыдущей группы. Суммы по группам от этого не изменятся.

Рассмотрим отдельно слова k-й группы. Они покрывают множество меры меньше 4^{-k} и можно построить мартингал m_k, который достигает на них значения 4^k. Теперь смешаем эти мартингалы и получим мартингал $m=\sum 2^{-k}m_k$. Этот мартингал достигает значения 2^k на словах k-й группы. Остаётся положить $g(n)=2^k$ для всех n в интервале $[n_k, n_{k+1})$ и заметить, что в бесконечно многих группах встречаются начала последовательности ω.

Напротив, пусть есть вычислимый мартингал m и неубывающая неограниченная вычислимая функция g. Нам надо покрыть последовательность ω, про которую известно лишь, что для бесконечно многих n значение $m(\omega_0\ldots\omega_{n-1})$ не меньше $g(n)$. При этом мера покрытия должна быть не больше некоторого $\varepsilon>0$. Как это сделать?

Прежде всего, увеличим функцию g на некотором начальном отрезке и будем считать, что она везде не меньше $1/\varepsilon+1$. Далее будем рассматривать все слова в порядке возрастания длин и сравнивать значения m и g на этих словах. Будем отбирать те слова, где m больше g. (Точнее говоря, поскольку m нам известно лишь с некоторой точностью, будем отбирать с некоторым запасом: при этом для всех отобранных слов $m>g-1$, и все слова с $m>g$ отобраны.)

По построению интервалы, соответствующие отобранным словам, заведомо покрывают последовательность ω. Кроме того, поскольку в них мартингал больше $1/\varepsilon$, их мера не превосходит ε. Наконец, сумма их мер вычислима: чтобы вычислить её с любой точностью $\delta > 0$, надо подождать момента, когда функция g станет больше $1/\delta + 1$: все следующие (более длинные) интервалы изменят сумму мер не более чем на δ. ▶

Этот же метод можно использовать и для доказательства такого критерия случайности по Шнорру [6]:

$\boxed{\textbf{269}}$ Докажите, что последовательность ω случайна по Шнорру относительно вычислимой меры μ тогда и только тогда, когда для любой вычислимой верхней оценки K для префиксной сложности KP и для любой монотонной вычислимой неограниченной функции h выполнено неравенство

$$K((\omega)_n) \geqslant -\log_2 \mu(\Omega_{(\omega)_n}) - h(n) - O(1)$$

при всех n (константа в $O(1)$ не зависит от n).

[Указание. Для вычислимых K и h покрытие из доказательства теоремы Левина – Шнорра имеет вычислимую меру. Рассуждение в другую сторону аналогично доказательству предыдущей теоремы: покрытие разобьём на конечные группы слов одинаковой длины μ-меры меньше 4^{-n}, увеличим суммарную меру слов из n-й группы пропорционально до 2^{-n} и по лемме о шторах получим вычислимую оценку для префиксной сложности; используя то, что в n-й группе все слова одинаковой длины, можно получить и функцию h.]

Этот результат можно использовать, чтобы доказать существование случайных по Шнорру (относительно равномерной меры) последовательностей, не случайных по Мизесу – Чёрчу [6]. В самом деле, он показывает, что если $KP((\omega)_n) \geqslant n - h(n) - O(1)$ для некоторой неограниченной монотонной функции h, которая стремится к бесконечности медленнее любой вычислимой неограниченной функции, то последовательность ω случайна по Шнорру. Такие функции h существуют (диагональный аргумент), и остаётся выбрать любую из них и построить последовательность с нужной оценкой, но не случайную по Мизесу – Чёрчу.

Как это делается? Можно взять случайную последовательность α и разбавить её редкими нулями — потом эти нули и образуют нулевую подпоследовательность, которая будет выбрана по Мизесу – Чёрчу. Нужно, чтобы (1) места этих нулей были вычислимы или хотя бы алгоритмически обнаруживаемы при чтении последовательности слева направо; (2) они были бы очень редкими, чтобы не сильно уменьшить сложность. Требование (2) означает, что функция $F\colon n \mapsto$ (позиция n-го нуля) должна расти быстрее всех вычислимых функций. Ключевая идея: по теореме Кучеры – Гача можно взять любую функцию F и затем α, относительно которой F вычислима. Правда, нам нужно вычислять $F(n)$, имея не всю последовательность α, а только начальный отрезок — но это не страшно, если $F(n)$ вычислится позже, то мы просто используем вместо $F(n)$ большее значение, а именно, то место, где она уже определилась. При оценке сложности мы используем тот факт, что нули на предсказуемых позициях можно добавлять или вычёркивать, не меняя префиксной сложности.

$\boxed{\mathbf{270}}$ Проведите это рассуждение подробно и докажите, что существуют случайные по Шнорру, но не случайные по Мизесу – Чёрчу последовательности.

(Они не будут случайными и относительно вычислимых мартингалов, так что случайность по Шнорру строго слабее случайности относительно вычислимых мартингалов.)

$\boxed{\mathbf{271}}$ Докажите, что последовательность не случайна по Курцу (см. раздел 3.4, с. 83) тогда и только тогда, когда существует вычислимый мартингал, который на этой последовательности ограничен снизу некоторой неубывающей неограниченной вычислимой функцией.

[Указание. Множества малой меры, покрывающие ω и состоящие из конечного числа интервалов, можно преобразовать в мартингалы, при этом известно, до какого момента сработает каждый из них, и можно преобразовать это в вычислимую оценку для суммарного мартингала. Напротив, зная мартингал и оценку для него на начальных отрезках последовательности ω, можно взять конечный набор интервалов, где мартингал больше заданного числа, и гарантированно покрыть ω.]

Отсюда видно, что случайные по Шнорру последовательности случайны и по Курцу (и включение это строгое, поскольку случайные по Курцу последовательности могут даже не удовлетворять усиленному закону больших чисел, см. задачу 92 на с. 83).

9.10. Мартингалы и эффективная размерность

В предыдущих разделах мы видели, как можно перевести понятия нулевого и эффективно нулевого (а также нулевого по Шнорру) множества на язык мартингалов. Аналогичный перевод возможен и для понятия размерности по Хаусдорфу. В одной фразе этот перевод можно описать так: чем меньше размерность множества, тем быстрее может расти мартингал на его элементах. (Говоря о мартингалах, в этом разделе мы имеем в виду мартингалы относительно равномерной меры.)

Начнём с классического варианта (без алгоритмов):

Теорема 184. *Множество $A \subset \Omega$ является α-нулевым тогда и только тогда, когда существует мартингал m с таким свойством: для любого $\omega \in A$ отношение*

$$\frac{m(x)}{2^{(1-\alpha)l(x)}}$$

не ограничено на начальных отрезках ω.

(При $\alpha = 1$ получается теорема 177.)

Немного более наглядная формулировка теоремы 184 такова: пусть игрок облагается налогом, в результате которого его капитал на каждом шаге (между двумя ставками) уменьшается в $2^{1-\alpha}$ раз, то есть умножается на $2^{\alpha-1}$. Тогда капитал игрока после x будет не мартингалом, а функцией m, для которой

$$2^{\alpha-1} m(x) = \frac{m(x0)}{2} + \frac{m(x1)}{2},$$

то есть
$$2^\alpha m(x) = m(x0) + m(x1).$$
Такие функции (следуя работе [92]) называют “α-gales”, что мы по-русски передаём как *α-мартингалы*. Они тоже соответствуют мерам: α-мартингал есть функция вида
$$p(x)2^{\alpha l(x)},$$
где $p(x)$ — мера интервала Ω_x относительно некоторого распределения. Аналогично можно определять и *α-супермартингалы*, где допускаются дополнительные потери игрока (помимо налога):
$$2^\alpha m(x) \geqslant m(x0) + m(x1).$$

Теперь можно пересказать утверждение теоремы 184 так: множество тех последовательностей, на начальных отрезках которых данный α-мартингал неограничен, является α-нулевым, и всякое α-нулевое множество содержится в множестве такого вида.

◄ Доказательство по существу повторяет рассуждение из теоремы 177. Пусть m — произвольный α-мартингал. Докажем, что множество тех последовательностей, для которых он неограничен (на начальных отрезках), имеет α-меру нуль. Для этого достаточно показать, что слова, где m достигает значения k (или больше) впервые, имеют сумму α-степеней мер не больше $1/k$. Если записать α-мартингал $m(x)$ как $p(x)2^{\alpha l(x)}$, то для выбранных слов x имеем $p(x) \geqslant k2^{-\alpha l(x)}$. Все выбранные слова несравнимы (одно не является началом другого), поэтому сумма p-мер не больше единицы, поэтому сумма величин $2^{-\alpha l(x)}$ по всем выбранным словам x ($=$ сумма α-мер соответствующих интервалов) не превосходит $1/k$.

Обратно, пусть дано некоторое множество α-меры нуль. Нам надо построить α-мартингал, который неограничен на всех его элементах. Возьмём покрытие интервалами с суммой α-мер меньше $1/k$, и построим α-мартингал m_k, который достигает значения k на этих элементах. (Затем можно сложить m_{4^k} с коэффициентами 2^{-k}, сумма α-мартингалов есть также α-мартингал.) Как построить m_k? Для данного слова x можно построить α-мартингал, который равен 1 на x и равен 0 на всех словах той же длины $l(x)$, продолжив его на более короткие слова (однозначно) и на более длинные (любым способом). На пустом слове значение этого α-мартингала будет $2^{-\alpha l(x)}$, то есть как раз α-мера соответствующего интервала. Значит, сумма этих α-мартингалов для всех x будет не больше $1/k$, и умножив её на k, получим искомый α-мартингал. ►

Теперь попытаемся сформулировать эффективный вариант теоремы. Пусть α — вычислимое действительное число из $(0, 1]$. Естественно определяются понятие *перечислимого снизу α-мартингала* и *α-супермартингала*. Как и в случае $\alpha = 1$, значение перечислимого снизу мартингала на пустом слове не обязательно равно единице — это произвольное (перечислимое снизу) число, не превосходящее единицы.

Можно предположить, что α-мартингалам (и супермартингалам) соответствуют введённые в разделе 5.8 эффективно α-нулевые множества и что доказательство

этого факта повторяет только что приведённое доказательство теоремы 184. В одну сторону это действительно так:

Теорема 185. *Пусть $\alpha \in (0,1]$ — вычислимое действительное число и множество $A \subset \Omega$ является эффективно α-нулевым. Тогда существует перечислимый снизу α-мартингал, который неограничен на начальных отрезках любой последовательности $\omega \in A$.*

◀ В самом деле, приведённая выше конструкция даёт перечислимые снизу α-мартингалы m_k, и смешивание сохраняет перечислимость снизу. ▶

В другую сторону мы встречаемся с трудностью. Пусть m — перечислимый снизу α-мартингал. Рассмотрим множество тех слов, на которых $m(x) > k$. Это множество перечислимо. Кроме того, сумма α-мер его минимальных элементов не превосходит $1/k$. Но множество минимальных элементов может быть неперечислимым, а если оставить все элементы (не только минимальные), не получается оценка для суммы мер. Поэтому доказать, что множество тех последовательностей, на которых перечислимый снизу α-мартингал неограничен, является *эффективно* α-нулевым, этим способом не удаётся.

На самом деле перечислимые α-мартингалы соответствуют несколько более слабому понятию α-нулевого множества, в котором ограничивается не суммарная α-мера покрытия, а лишь суммы α-мер *непересекающихся* интервалов этого покрытия. Эквивалентное определение получится, если рассматривать лишь максимальные интервалы покрытия (не содержащиеся в других). Но для характеризации эффективной хаусдорфовой размерности в терминах мартингалов эта разница не имеет значения, и нам вполне достаточно такого утверждения:

Теорема 186. *Пусть m — перечислимый снизу α-мартингал и $\beta > \alpha$. Тогда множество тех последовательностей, на (начальных отрезках) которых m неограничен, является эффективно β-нулевым.*

(Напомним, что в данных нами определениях предполагается, что α и β вычислимы.)

◀ Пусть k — положительное натуральное число. Рассмотрим те слова x, для которых $m(x) > k$, и соответствующие им интервалы. Получится покрытие интересующего нас множества. Найдём сумму β-мер этих интервалов. Как мы видели, если выбрать несравнимое множество из этих интервалов, то сумма α-мер будет не больше $1/k$. В частности, для каждой длины N сумма α-мер интервалов этой длины (из покрытия) не превосходит $1/k$. А сумма β-длин будет меньше в $2^{N(\beta-\alpha)}$ раз. Поэтому суммирование по длинам приведёт лишь к умножению на сумму бесконечно убывающей геометрической прогрессии, которая для каждого β конечна (хотя и велика, если β близко к α), так что множество оказывается эффективно β-нулевым. ▶

Из двух последних теорем вытекает такое следствие:

Теорема 187. *Эффективная хаусдорфова размерность произвольного множества $A \subset \Omega$ равна точной нижней грани тех α, при которых существует*

перечислимый снизу α-мартингал, неограниченный на всех элементах множества A.

В этом утверждении можно заменить мартингалы на супермартингалы.

Заметим, что отсюда легко получить другое доказательство теоремы 120. В самом деле, α-супермартингалы получаются из полумер умножением на $2^{\alpha l(x)}$. Поэтому среди перечислимых снизу α-супермартингалов есть максимальный (получающийся из максимальной перечислимой снизу полумеры a), и в последней теореме можно рассматривать только его. Значит, эффективная размерность множества $\{\omega\}$ равна точной нижней грани тех α, при которых произведение $a(\omega_0\omega_1\ldots\omega_{n-1})2^{\alpha n}$ не ограничено. Логарифм его равен $\alpha n - KA(\omega_0\omega_1\ldots\omega_{n-1})$, и потому точная нижняя грань таких α есть

$$\liminf \frac{KA(\omega_0\ldots\omega_{n-1})}{n}.$$

(В теореме 120 использовалась простая колмогоровская сложность, но с интересующей нас точностью это не играет роли.)

9.11. Частичные правила выбора

Вернёмся теперь от мартингалов к правилам выбора и подпоследовательностям. Определяя случайность по Мизесу – Чёрчу, мы рассматривали правила выбора, соответствующие *разрешимым* множествам R. Такое правило никогда не «зависает», решая вопрос о том, нужно или нет сделать ставку на очередной член последовательности.

Можно рассмотреть и более широкий класс правил выбора. А именно, пусть r — частичная вычислимая функция на двоичных словах, принимающая значения 0 и 1. Имея начальный отрезок $\omega_0\ldots\omega_{n-1}$ и решая вопрос о том, включать ли следующий член ω_n в подпоследовательность, мы вычисляем значение $r(\omega_0\ldots\omega_{n-1})$. Если оно равно единице, то ω_n включается, если нулю — то нет, а если не определено — процесс выбора вообще на этом обрывается и получается конечная подпоследовательность. Формально говоря, это эквивалентно рассмотрению множества R, определённого следующим образом: слово x принадлежит R, если функция r определена на x и на всех его началах, причём $r(x) = 1$. (Это множество может быть неразрешимым для вычислимых частичных r.) Таким образом, каждой вычислимой функции r описанного вида соответствует правило выбора, которое мы будем обозначать S_r.

Получаем более широкий класс правил выбора, чем допустимые по Чёрчу. Этот класс рассматривал Дэли [34] и мы будем называть такие правила *допустимыми по Чёрчу – Дэли*. Последовательность называется *случайной по Мизесу – Чёрчу – Дэли*, если любое допустимое по Чёрчу – Дэли правило выбора даёт сбалансированную или конечную последовательность.

272 Докажите, что применение допустимого по Чёрчу – Дэли правила к случайной по Мизесу – Чёрчу – Дэли последовательности даёт случайную по Мизесу – Чёрчу – Дэли последовательность.

При таком расширении класса допустимых правил класс случайных последовательностей уменьшается. Это следует из теоремы 173 (с. 302) и такого утверждения (доказанного Меркле [102]):

Теорема 188. *Не существует случайной по Мизесу – Чёрчу – Дэли последовательности ω, для которой*

$$KS(\omega_0 \ldots \omega_{n-1}) = O(\log n).$$

◄ Пусть

$$KS(\omega_0 \ldots \omega_{n-1}) < c \log n$$

для некоторого c и для всех достаточно больших n. Мы хотим показать, что такая последовательность не может быть случайной по Мизесу – Чёрчу – Дэли, то есть построить правило выбора, нарушающее устойчивость частот.

Для начала рассмотрим случай, когда $c < 1$. Множество всех слов, имеющих сложность не больше $c \log n$, представляет собой перечислимое множество из не более чем n^c элементов, и при бо́льших n число элементов в этом множестве (обозначим его C_n) не больше $n/10$. Зафиксируем одно такое n. Читая слева направо начальный отрезок длины n, будем пытаться предсказывать следующий бит по предыдущим. Оказывается, что из первых n битов можно гарантированно отгадать 90%. Это делается так. Начнём перечислять C_n и дождёмся появления первого элемента. Этот элемент мы объявляем «текущим кандидатом» и предсказываем те биты, которые в нём стоят, пока не ошибёмся. Такая ошибка означает, что начальный отрезок последовательности ω, хотя и принадлежит C_n, но отличается от найденного нами элемента. Поэтому продолжим перечисление C_n, пока в нём не обнаружится другой элемент, не противоречащий уже известным битам, и будем использовать этот элемент для предсказаний — снова до первой ошибки. Ясно, что число ошибок не больше числа различных элементов в C_n, и поэтому как минимум $0{,}9n$ членов мы предскажем правильно.

Будем применять такой метод предсказания для быстро возрастающей последовательности значений $n_0 < n_1 < n_2 \ldots$ и достаточно большого n_0 (начиная с которого сложность меньше $c \log n$). Используя C_{n_i} на участке $[n_{i-1}, n_i]$, мы допускаем не более $0{,}1n_i$ ошибок. Если n_{i-1}/n_i мало (скажем, меньше $0{,}1$), то в целом среди первых n_i предсказаний будет не более $0{,}2n_i$ ошибок. (Мы реально используем C_{n_i} на отрезке $[n_{i-1}, n_i]$ вместо $[0, n_i]$, а до этого может быть в самом худшем случае лишь $n_{i-1} < 0{,}1n_i$ ошибок.)

Осталось заметить, что (как и в теореме 169) алгоритм предсказания соответствует двум правилам выбора: одно выбирает те члены, где мы предсказывали единицу, а второе — где предсказывали нуль, и одна из подпоследовательностей будет сильно несбалансированной.

Что же делать при $c > 1$? Пусть, скажем, $c = 1{,}5$. Тогда оценка для числа элементов в C_n будет $n^{1,5}$, что много больше n, так что наше рассуждение не проходит. Но можно сделать так. Разделим слово $\omega_0 \ldots \omega_{n-1}$ на две половины u и v (по $n/2$ битов каждая). Пара $\langle u, v\rangle$ имеет сложность не более $1{,}5 \log n$. Но сложность пары равна сумме $KS(u) + KS(v \mid u)$ с точностью до членов порядка

$O(\log KS(u,v)) = O(\log\log n)$, поэтому либо $KS(u) < 0{,}8\log n$, либо $KS(v\,|\,u) < 0{,}8\log n$. В каждом случае мы можем применить уже разобранный метод угадывания, поскольку оценка сверху для числа кандидатов, $n^{0{,}8}$, менее одной десятой от числа угадываемых битов ($n/2$). (Заметим, что при предсказании битов второй половины мы уже знаем первую половину, так что можем перечислять слова, у которых условная сложность меньше $0{,}8\log n$)

Таким образом, один из двух алгоритмов предсказания будет успешным на своей половине. А если алгоритм предсказания на данной половине успешен, то одно из двух правил выбора, пропускающих все члены из другой половины и выбирающее предсказанные нули/единицы в этой, будет успешным.

Тут, однако, возникает существенная трудность. Всё сказанное относилось к одному значению n; далее нам нужно соединить наши алгоритмы предсказания для разных n_i в один. Проблема в том, что если мы пытались предсказывать биты (скажем, в левой половине), считая, что $KS(u) < 0{,}8n$, в то время как на самом деле это было не так, то наш алгоритм не только не будет успешным (это было бы не страшно), но вообще «зависнет» и тем самым никаких предсказаний не даст не только для этого n, но и для всех следующих n из последовательности n_i.

Чтобы преодолеть эту трудность, надо вспомнить, как доказывалась формула для сложности пары (теорема 21, с. 47) и встроить её доказательство в нашу конструкцию. Вот что при этом получится.

Мы по-прежнему будем отдельно предсказывать левую и правую половины (u и v) начального отрезка длины n, но делать это несколько иначе. Предсказывая v слева направо (при известном u), мы перечисляем множество C_n возможных значений начального отрезка (состоящее из всех слов длины n и сложности меньше $1{,}5\log n$), ожидая появления кандидата, согласованного с u и уже известными битами в v. Когда такой кандидат найден, мы используем его для предсказания, пока не ошибёмся (после чего ищем следующего кандидата). Успешность таких предсказаний (при данном u) зависит от того, сколько существует различных v, при которых $uv \in C_n$. Если их много, то мы можем ошибаться каждый раз (и каждый раз заменять кандидата). А если таких v мало, скажем, не больше $n^{0{,}8}$, то наши предсказания будут успешными. Важно, что в любом случае этот способ предсказания никогда не «зависнет», если только uv действительно содержится в C_n. (А это можно гарантировать, взяв n_0 достаточно большим.)

Теперь о предсказании левой половины. Здесь мы в качестве кандидатов будем рассматривать такие слова u, для которых существует не менее $n^{0{,}8}$ значений v, при которых $uv \in C_n$. Эти предсказания будут удачными, если левая половина попадёт в число кандидатов (что обязательно случится, если предсказания в правой половине будут неудачными). Но если левая половина не окажется среди кандидатов, то этот способ предсказания «зависнет» (в некоторый момент мы будем ждать появления кандидата, согласованного с уже известными членами, и никогда не дождёмся).

Что произойдёт при объединении этих алгоритмов для разных n_i? Сначала рассмотрим алгоритмы предсказания правых половин. Эти алгоритмы никогда не зависают (считаем, что n_0 достаточно велико, так что все начальные отрезки длин n_i имеют сложность меньше $1{,}5\log n_i$). Если бесконечное число среди этих алгоритмов успешно (соответствующие множества малы), то соединённый алгоритм будет

успешен (доля успешных предсказаний не стремится к $1/2$). Значит, достаточно рассмотреть случай, когда лишь конечное число среди этих алгоритмов успешно. Тогда при всех достаточно больших i число кандидатов v (при данном u) будет больше $n_i^{0,8}$, и потому, отбросив некоторый начальный отрезок, мы можем считать левый алгоритм предсказания всегда успешным. Таким образом, и в этом случае последовательность не будет случайной по Мизесу – Чёрчу – Дэли.

Что же делать, если наша константа c ещё больше? Надо действовать аналогично, только разбивать последовательность на большее число частей. Алгоритм предсказания для самой правой части никогда не зависает. Если он успешен в бесконечном числе случаев, всё доказано. Если с некоторого момента он неуспешен, то с этого момента следующий алгоритм (для предпоследней части) никогда не зависает. Если он успешен в бесконечном числе случаев, всё доказано. Если нет, то рассмотрим предыдущую часть и так далее.

(Аккуратное изложение со всеми деталями см. в статье [102].) ▶

Таким образом, сложность начальных отрезков случайных по Мизесу – Чёрчу – Дэли последовательностей не может быть логарифмической. Но она может расти лишь чуть быстрее (скажем, быть $O(\log n \log\log n)$, как показано в [102]:

Теорема 189. *Пусть $f\colon \mathbb{N} \to \mathbb{N}$ — всюду определённая неубывающая вычислимая неограниченная функция. Тогда существует случайная по Мизесу – Чёрчу – Дэли последовательность, у которой сложность начального отрезка длины n не больше $f(n)\log n + O(1)$ (при всех n).*

◀ Вспомним, как строились случайные по Мизесу – Чёрчу последовательности малой сложности (теорема 173). Тогда дополнительная информация состояла в том, какие из правил действительно являются правилами (какие программы всюду определены) — по одному биту на правило. Теперь этого уже мало. Для каждой программы надо знать, в какой момент она перестала быть определённой (то есть первый момент, когда она не определена на текущем начальном отрезке), чтобы с этого момента её заменить на что-нибудь безобидное. Таким образом, если при построении начального отрезка длины n в игру вводятся $f(n)$ программ, то для его описания при известном n достаточно $f(n)\log n$ битов информации (для каждой программы нужно $\log n$ битов, чтобы указать момент её неопределённости).

Заметим, что наличие n в качестве условия не существенно, так как меняет сложность на $O(\log n)$, что соответствует изменению функции f на $O(1)$. ▶

Аналогичное утверждение можно сформулировать и для мартингалов. Напомним, что мы рассматривали последовательности, случайные относительно вычислимых мартингалов, и при этом в качестве мартингалов (для равномерной меры) можно было рассматривать вычислимые всюду определённые функции с рациональными значениями. Теперь разрешим и частичные функции с рациональными значениями, при этом неравенство из определения мартингала должно выполняться в тех случаях, когда все три входящих в него величины определены.[1] Будем говорить,

[1] С точки зрения игры естественно предполагать также, что $m(x0)$ и $m(x1)$ определены одновременно, так как к следующему бросанию монеты можно приступить, только когда сделаны обе ставки. Это можно предполагать без ограничения общности: если $m(x)$ и $m(x0)$ определились, то по ним можно вычислить $m(x1)$.

что такой частичный мартингал *выигрывает* на некоторой последовательности, если он определён на всех её начальных отрезках и не ограничен на них. Будем говорить, что последовательность случайна (по равномерной мере) *относительно частичных вычислимых мартингалов*, если не существует вычислимого частичного мартингала, выигрывающего на ней. Теперь можно доказать, следуя [102], такое обобщение предыдущей теоремы 189:

Теорема 190. (а) *Всякая последовательность, случайная относительно частичных вычислимых мартингалов, случайна по Мизесу – Чёрчу – Дэли.*

(б) *Пусть* $f\colon \mathbb{N} \to \mathbb{N}$ *— неубывающая неограниченная вычислимая функция. Тогда существует случайная относительно частичных вычислимых мартингалов последовательность, у которой сложность начального отрезка длины* n *не больше* $f(n)\log n + O(1)$ *(при всех* n*).*

Отметим, что из этих утверждений вытекает предыдущая теорема.

◀ (а) Применим ту же самую конструкцию, преобразующую правило выбора в мартингал, что и в теореме 182, пункт (в). Если правило R не всюду определено, то и мартингал будет частичным. Но если правило в применении к некоторой последовательности ω даёт бесконечную подпоследовательность, то соответствующий ему мартингал будет определён на всех начальных отрезках последовательности ω.

(б) И здесь практически без изменений проходит рассуждение из теоремы 182, пункт (б). Для каждого вводимого в игру мартингала нам надо знать, в какой момент он становится неопределённым (чтобы с этого момента заменить его чем-нибудь безобидным, например, его последним значением). Эта информация занимает не более $\log n$ битов для каждого мартингала, использованного при построении начального отрезка длины n, и если к этому моменту ввести в дело не больше $f(n)$ мартингалов, получится как раз требуемая оценка. ▶

9.12. Немонотонные правила выбора

До сих пор наши правила выбора были монотонны (сохраняли порядок членов в исходной последовательности). Немонотонные правила были предложены Колмогоровым [62] и независимо Лавлэндом [88, 89].

Наглядно идею Колмогорова и Лавлэнда можно объяснить следующим образом. Пусть в казино бросают монету не при нас, а заранее, и результаты бросаний (нули и единицы) записывают на карточках. Эти карточки кладут «лицом вниз», так что мы не видим, что на них написано.

Глядя на последовательность карточек, мы имеем право попросить перевернуть любую из них (не делая ставки), а также сделать ставку на любую ещё не перевёрнутую карточку.

Наша стратегия в такой игре задаётся двумя функциями. Первая функция F отображает двоичные слова в натуральные числа и говорит, какую карточку надо переворачивать следующей (в зависимости от того, что мы увидели на уже перевёрнутых карточках). Мы предполагаем, что значения функции F на любых двух словах, из которых одно является началом другого, различны (никакая карточка не переворачивается дважды в ходе игры).

Вторая функция G определена также на двоичных словах и принимает значения 0 и 1; договоримся, что значение 0 означает, что очередная карточка переворачивается «для информации», а 1 означает, что на неё «делается ставка» (соответствующий член включается в подпоследовательность).

Формально говоря, для любой частичной функции F (с указанным свойством) и для любой частичной функции G мы определяем отображение (правило выбора) $S_{F,G}\colon \Omega \to \Sigma$ следующим образом: сначала строим конечную или бесконечную последовательность натуральных чисел $n_0, n_1, \ldots$ по формулам

$$n_0 = F(\Lambda),\ \ n_1 = F(\omega_{n_0}),\ \ n_2 = F(\omega_{n_0}\omega_{n_1}),\ \ \ldots$$

(построение прерывается, как только очередное значение F не определено). Наше условие на функцию F гарантирует, что все номера n_i различны.

Затем отбираются те члены ω_{n_i}, при которых значение G на слове $\omega_{n_0}\omega_{n_1}\ldots\omega_{n_{i-1}}$ определено и равно 1, а значения G на всех началах этого слова определены (и могут быть любыми). Соответствующие ω_{n_i} (в порядке возрастания i) и образуют последовательность $S_{F,G}(\omega)$.

Правила выбора $S_{F,G}$, соответствующие частичным вычислимым функциям F и G, называются *допустимыми по Колмогорову – Лавлэнду*. Последовательность $\omega \in \Omega$ называется *случайной по Мизесу – Колмогорову – Лавлэнду* относительно равномерной меры, если любое допустимое по Колмогорову – Лавлэнду правило выбора даёт сбалансированную (или конечную) последовательность. Английский термин: *Kolmogorov – Loveland stochastic* (или *KL-stochatic*).

Аналогично определяется случайность по Мизесу – Колмогорову – Лавлэнду и относительно бернуллиевой меры (независимые испытания с вероятностью успеха p) при любом p.

Следующее простое, но неожиданное наблюдение сделал Меркле [101]:

Теорема 191. *В определении случайности по Мизесу – Колмогорову – Лавлэнду можно ограничиться лишь всюду определёнными вычислимыми функциями F и G; получится тот же самый класс последовательностей.*

◀ Пусть имеются частичные функции F и G, выбирающие из некоторой последовательности ω бесконечную несбалансированную подпоследовательность. Разобьём эту последовательность на две, в зависимости от чётности порядкового номера в исходной последовательности ω. Одна из них будет бесконечной, но не сбалансированной. Поэтому без ограничения общности можно считать, что наше правило включает в подпоследовательность лишь (скажем) чётные члены последовательности. Тогда нечётные члены можно читать заранее, это не повредит, так как в последовательность включены они всё равно не будут. Поэтому, если частичное правило долго работает, можно параллельно про запас читать нечётные члены последовательности. Такое модифицированное правило задаётся всюду определёнными функциями F и G. (Если исходный алгоритм «завис», то новый будет читать подряд нечётные члены последовательности, но ничего так и не выберет.) ▶

(В этом доказательстве мы свели одно правило с частичными функциями к двум правилам со всюду определёнными функциями.)

Как соотносится новое определение с уже известными нам? Частичный ответ на этот вопрос даёт следующая теорема.

Теорема 192. (а) *Всякая случайная по Мизесу – Колмогорову – Лавлэнду последовательность случайна по Мизесу – Чёрчу – Дэли (и, следовательно, по Мизесу – Чёрчу).*

(б) *Всякая случайная по Мартин-Лёфу последовательность случайна по Мизесу – Колмогорову – Лавлэнду.*

Точнее, пункт (а) имеет место для любого действительного $p \in (0,1)$; пункт (б) дополнительно предполагает, что p вычислимо и случайность по Мартин-Лёфу понимается относительно бернуллиевой меры с параметром p.

◄ (а) Допустимые по Мизесу – Чёрчу – Дэли (и тем более по Мизесу – Чёрчу) правила выбора являются частными случаями допустимых по Колмогорову – Лавлэнду правил.

(б) По существу это рассуждение повторяет аналогичное рассуждение для случайности по Мизесу – Чёрчу. Мы предполагаем, что фиксировано некоторое вычислимое p, для которого изучается случайность по Мизесу – Колмогорову – Лавлэнду (предел частот в подпоследовательностях должен быть равен p) и по Мартин-Лёфу (рассматривается вычислимая мера μ_p на Ω, соответствующая независимым испытаниям, в каждом из которых вероятность успеха равна p).

Пусть фиксированы вычислимые функции F и G, задающее допустимое по Мизесу – Чёрчу правило выбора.

Для каждого целого n и рационального q рассмотрим множество $D_{n,q}$ всех двоичных слов длины n, у которых доля единиц больше q. Мы знаем, что при фиксированном $q > p$ и при возрастающем n мера множества $D_{n,q}$ (точнее, μ_p-мера множества $\boldsymbol{D}_{n,q}$ всех последовательностей, имеющих начало в $D_{n,q}$) экспоненциально убывает.

Рассмотрим теперь прообраз этого множества относительно $S_{F,G}$, то есть множество всех последовательностей ω, для которых описанный процесс выбора даёт подпоследовательность длины не меньше n и частота единиц среди первых n выбранных членов больше q. Легко заметить, что *μ_p-мера этого прообраза не больше $\mu_p(\boldsymbol{D}_{n,q})$*.

Неформально говоря, мы замечаем, что результат выбора имеет то же самое бернуллиево распределение, что и исходная последовательность, если отвлечься от того, что он может быть конечным. Подробнее: пусть t — некоторое двоичное слово длины $k-1$. Рассмотрим условную вероятность того, что k-я по счёту выбранная (включённая в подпоследовательность) карточка содержит единицу, при условии что предыдущие члены подпоследовательности образуют слово t и до выбора k-й карточки вообще дело дойдёт. Эта условная вероятность равна p, поскольку к моменту переворачивания k-й карточки её содержимое ещё никак не использовалось и мы можем «отложить» бросание монеты до этого момента. Более формально, событие, которое фигурирует в качестве условия, является счётным объединением непересекающихся вариантов, каждый из которых соответствует конкретному ходу процесса выбора вплоть до момента, когда выбрана (но ещё не перевёрнута) k-я карточка подпоследовательности (какие карточки перевёрнуты и что в них

оказалось). В каждом из вариантов уже определено, какая карточка будет следующей выбрана, и этот вариант разбивается на две части в пропорции $p : (1-p)$ в зависимости от содержания этой карточки. Отсюда по индукции легко следует выделенное курсивом утверждение.

Остаётся заметить, что упомянутый прообраз не только имеет малую меру (быстро убывающую с ростом n), но ещё и перечислим (любой из вариантов зависит от конечного числа значений функций F и G и потому рано или поздно обнаружится), поэтому, как и в усиленном законе больших чисел, возникает эффективно нулевое множество (для каждого q — своё). Все эти эффективно нулевые множества (а также аналогичные множества для последовательностей с частотами меньше p) содержатся в максимальном эффективно нулевом множестве. ►

В следующем разделе мы докажем такое усиление утверждения (б): если вычислимая последовательность p_n действительных чисел в интервале $(0,1)$ вычислимо сходится к пределу $p \in (0,1)$, то всякая последовательность, случайная по Мартин-Лёфу относительно произведения мер (испытания независимы, вероятность удачи в i-м испытании равна p_i), случайна по Мизесу – Колмогорову – Лавлэнду с параметром p. Это позволит нам строить примеры случайных по Мизесу – Колмогорову – Лавлэнду последовательностей с плохими свойствами (неслучайные по Мартин-Лёфу, содержащие больше единиц, чем нулей, и другие).

Пока же мы хотим действовать в другом направлении и показать — для случая равномерной меры, — что случайная по Мизесу – Колмогорову – Лавлэнду последовательность (в отличие от случайных по Мизесу – Чёрчу и Мизесу – Чёрчу – Дэли) имеет достаточно большую сложность начальных отрезков:

Теорема 193. *Пусть $\omega \in \Omega$ такова, что $KS(\omega_0 \ldots \omega_{n-1}) < \alpha n$ для некоторого $\alpha < 1$ и всех достаточно больших n. Тогда последовательность ω не случайна по Мизесу – Колмогорову – Лавлэнду.*

Этот результат (доказанный Ан. А. Мучником в конце 1980-х годов) впоследствии был усилен: оказалось, что если неравенство верно *для бесконечно многих n*, то последовательность не случайна по Мизесу – Колмогорову – Лавлэнду. Мы ограничимся доказательством первоначального результата.

Для него нам понадобится некоторое вспомогательное утверждение о цене «инсайдерской информации» в играх с ограниченными ставками. Сейчас мы его сформулируем и докажем, а потом перейдём к доказательству теоремы 193.

Представим себе, что крупье бросает монету раз за разом, и перед каждым бросанием мы можем сделать ограниченную ставку $u \in [-1,1]$ (положительные u соответствуют ставкам на единицу, отрицательные — на нуль). После этого мы получаем u рублей, если выпадает единица, и $(-u)$ рублей, если выпадает нуль.

В отличие от игр с мартингалами максимальная ставка всегда есть 1, независимо от того, сколько мы выиграли или проиграли в предыдущих партиях (и тем самым наш проигрыш может быть сколь угодно большим — а раньше он был ограничен начальным капиталом).

Во избежание путаницы подчеркнём ещё, что мы делаем ставки в том же порядке, в котором монету бросают (никаких немонотонных правил выбора пока нет).

Лемма. *Пусть заранее известно множество A слов длины n, содержащее не более 2^s элементов для некоторого $s < n$. Тогда существует стратегия, гарантирующая выигрыш не менее $n - s$ на любом элементе множества A (для любой последовательности бросаний, соответствующей слову из множества A).*

Например, если A содержит единственный элемент (то есть мы заранее знаем результаты всех бросаний), то можно выиграть n рублей (что не удивительно: в каждой партии мы выигрываем рубль). (В рассмотренном выше «мартингальном» варианте игры выигрыш был гораздо больше: в 2^n раз.) Если мы знаем результаты каких-то фиксированных k бросаний, то по лемме можем выиграть k рублей (что тоже не удивительно). Чуть более сложный пример: пусть мы знаем, что число орлов будет чётно (что соответствует $s = n - 1$). Лемма говорит, что на этом можно выиграть рубль. Легко понять, как: нужно делать нулевые ставки до последнего момента, а результат последнего бросания нам фактически известен, и на него надо сделать максимальную ставку.

После этих примеров общее доказательство выглядит довольно естественно. В каждый момент нам известен некоторый начальный отрезок w последовательности (пусть его длина j). Имеется 2^{n-j} продолжений слова w, имеющих длину n, однако не все они принадлежат A. Рассмотрим условную вероятность попадания в A после w (долю продолжений, попадающих в A); минус логарифм этой вероятности назовём *информационным капиталом игрока.*

Вначале этот капитал равен $n - s$. Мы докажем, что можно так выбрать стратегию (так определять размер ставок в каждый момент), чтобы сумма информационного и реального капитала не убывала. Тогда в конце игры, когда информационный капитал равен нулю (мы предполагаем, что последовательность исходов на самом деле оказалась внутри множества A), реальный капитал будет не меньше $n - s$, что и требовалось.

Почему это возможно? Пусть в данный момент информационный капитал равен $(-\log p)$ (где p — текущая доля элементов из A среди продолжений). Зная A, мы можем найти этот капитал, а также узнать, как он изменится после появления нуля и единицы: после появления нуля он будет равен $(-\log p_0)$, а после единицы будет равен $(-\log p_1)$, где p_0 и p_1 — соответствующие доли. Ясно, что $p = (p_0 + p_1)/2$. Нам нужно найти такой размер очередной ставки d, чтобы в обоих случаях сумма информационного и реального капитала не уменьшилась:

$$-\log p_0 - d \geqslant -\log p,$$
$$-\log p_1 + d \geqslant -\log p,$$

или, разрешая эти неравенства относительно d,

$$-\log p_0 + \log p \geqslant d \geqslant -\log p + \log p_1,$$
$$-\log(p_0/p) \geqslant d \geqslant \log(p_1/p).$$

Ясно, что такое d можно выбрать, если и только если $p/p_0 \geqslant p_1/p$, то есть $p^2 \geqslant p_0 p_1$ (а это так согласно неравенству о среднем арифметическом и геометрическом). Заметим также, что p_0 и p_1 не больше $2p$, поэтому границы для d (и тем самым само d) не выходят за пределы промежутка $[-1, 1]$. Лемма доказана.

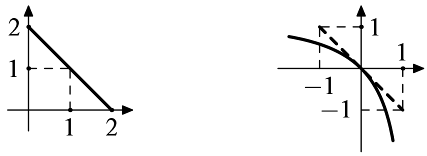


Рис. 9.1. Пространство выбора для игрока: слева в обычных координатах, справа — в логарифмических. Пунктиром показано пространство выбора для игры с суммированием ограниченных ставок.

Связь этой леммы с мартингальными играми можно пояснить так. Мартингал делает некоторый выбор (распределение ставок), в результате которого наш капитал умножится на той или иной коэффициент после очередного раунда. Возможные варианты выбора лежат на отрезке. Коэффициенты для разных раундов перемножаются, а их логарифмы складываются. Можно изобразить варианты выбора (образ этого отрезка) в логарифмических координатах, получится некоторая кривая (рис. 9.1). Несложно проверить, что эта кривая целиком лежит ниже касательной, поэтому игра станет лишь выгоднее, если заменить её на касательную. А тогда мы как раз получим игру со сложением ограниченных ставок.

Теперь легко доказать такое утверждение:

Теорема 194. *Пусть K — произвольная вычислимая всюду определённая верхняя оценка для функции KS, а последовательность $\omega \in \Omega$ такова, что*

$$K(\omega_0 \ldots \omega_{n-1}) \leqslant \alpha n$$

для некоторого $\alpha < 1$ и для всех достаточно больших n. Тогда последовательность ω не случайна по Мизесу – Чёрчу.

◀ Для каждого n можно алгоритмически построить список A_n всех слов длины n, для которых значение функции K не превосходит αn. Этот список будет содержать не более $2^{\alpha n + O(1)}$ слов. Мы знаем, что при достаточно больших n начальный отрезок последовательности ω попадает в A_n. Для этих n построенная по лемме (для множества A_n) стратегия будет выигрывать не меньше $(1-\alpha)n - O(1)$ рублей при игре с начальным отрезком последовательности ω длины n.

Рассмотрим последовательность быстро растущих n_i (насколько быстро, что $n_{i-1}/n_i \to 0$) и соединим стратегии для A_{n_i} в одну. Фактически стратегия для A_{n_i} будет применяться не на всём начальном отрезке длины n_i, а лишь после окончания предыдущего отрезка n_{i-1}, что, по нашему предположению, составляет бесконечно малую долю от n_i. Поэтому соединённая стратегия будет успешна в том смысле, что выигрыш её на начальном отрезке длины n будет превосходить εn для фиксированного $\varepsilon > 0$ и для бесконечно многих n. (Достаточно взять $\varepsilon < 1 - \alpha$ и $n = n_i$ при бо́льших i.)

А это противоречит случайности по Мизесу – Чёрчу (теорема 170, с. 298). ▶

Теперь мы уже готовы к доказательству теоремы 193.

◀ По аналогии с только что проведённым доказательством, мы можем рассмотреть множество A_n всех слов длины n, имеющих сложность не более αn. Оно

содержит примерно $2^{\alpha n}$ слов, среди которых заведомо имеется начальный отрезок последовательности ω. Но теперь мы не можем алгоритмически построить список всех слов множества A_n, а можем лишь перечислять эти слова (не зная ни в какой момент, все ли слова уже появились или ещё нет). Преодолеть эту трудность можно, вспомнив, что нам разрешено использовать немонотонные правила выбора. Вот как это делается.

Снова рассмотрим быстро растущую вычислимую последовательность n_i, например, $n_i = i!$ Разобьём последовательность на отрезки, проведя границы в точках n_i; длина i-го отрезка будет $n_i - n_{i-1}$. Немного увеличив α, мы можем считать, что сложность i-го отрезка последовательности не больше α, умноженного на его длину. Другими словами, мы будем считать, что удельная (в расчёте на букву) сложность i-го отрезка последовательности не превосходит α. Пусть A_i — множество всех слов длины $n_i - n_{i-1}$, для которых удельная сложность не больше α. Мы знаем, что i-й отрезок последовательности ω, который мы обозначим через ω^i, лежит в A_i (по крайней мере при достаточно бо́льших i), и можем перечислять множество A_i, зная i.

Пусть t_i — число шагов этого перечисления, которое нужно сделать, прежде чем ω^i появится в A_i. Будем отдельно рассматривать чётные и нечётные номера, и сравним соседние значения t_{2m} и t_{2m+1}. Тривиальное замечание: либо $t_{2m} \leqslant t_{2m+1}$, либо $t_{2m+1} \leqslant t_{2m}$ (а может, и то, и другое). Что это нам даёт? А вот что: одна стратегия может посмотреть (не делая ставок, в порядке информации) участок ω^{2m}, после чего перечислять A_{2m} до появления ω^{2m}, и тем самым найти число t_{2m}. Затем эта стратегия делает столько же (то есть t_{2m}) шагов перечисления множества A_{2m+1}, надеется на то, что неизвестное ей ω^{2m+1} принадлежит перечисленной части множества A_{2m+1} и делает ставки как в предыдущей теореме. Это приведёт к успеху, если $t_{2m} \geqslant t_{2m+1}$; в противном случае мы можем всё проиграть. Но тогда парная стратегия, которая просматривает ω^{2m+1}, находит t_{2m+1}, а затем делает столько же шагов перечисления A_{2m}, будет успешной.

Таким образом, при каждом m у нас есть пара стратегий (со ставками в промежутке $[-1, 1]$), одна из которых является успешной (выигрывает не менее $1-\alpha$ в расчёте на каждую сделанную ставку). (Формально говоря, такие стратегии имеются лишь при достаточно бо́льших m, но мы можем все меньшие пропустить.) Одна из этих стратегий монотонна и (как в теореме 170) приближается средним арифметическим конечного числа правил выбора, допустимых по Чёрчу. Число правил зависит от допустимой погрешности; нам нужно, чтобы погрешность была мала по сравнению с $1-\alpha$, и этого можно добиться для фиксированного (не зависящего от m) числа правил, которое мы обозначим через N. Другая стратегия немонотонна, и потому получаются N правил выбора, допустимых по Колмогорову – Лавлэнду. Всего получается $2N$ правил выбора, имеющих дело с двумя соседними участками последовательности.

Соединяя правила для разных отрезков, получим $2N$ допустимых по Колмогорову – Лавлэнду правил выбора. Вспоминая, что n_{i-1}/n_i мало и результат выбора на предыдущем отрезке почти не меняет частоту на следующем, получаем, что для каждого m хотя бы одно из построенных $2N$ правил даёт значительное отклонение

частоты. Значит, одно из них даёт такое отклонение для бесконечно многих m, и последовательность ω не случайна по Мизесу – Колмогорову – Лавлэнду.

Теорема 193 доказана. ►

Из неё (и теоремы 189) вытекает такое следствие:

Теорема 195. *Существуют случайные по Мизесу – Чёрчу – Дэли последовательности, не являющиеся случайными по Мизесу – Колмогорову – Лавлэнду.*

В связи с теоремой 193 возникает естественный вопрос: пусть известно, что сложность (конечного) слова x мала. Существует ли немонотонное правило выбора небольшой колмогоровской сложности, выбирающее из x сильно несбалансированную подпоследовательность. В аккуратной постановке этой задачи имеется несколько параметров: длина n слова x, его дефект случайности d, сложность правила выбора (при известном n) и несбалансированность (разница между количеством нулей и единиц) отобранной подпоследовательности. В работе [41] получен следующий (далеко не окончательный) результат в этом направлении: существует правило выбора сложности $O(\log(n/d))$ при известном n, отбирающее подпоследовательность, несбалансированность которой равна $\Omega(n/\log(n/d))$. В частности, если дефект случайности, как в теореме 193, пропорционален n, то правило отбора имеет ограниченную сложность, а несбалансированность отобранной подпоследовательности пропорциональна длине исходной последовательности n.

9.13. Случайность по изменённой мере

В этом разделе мы изложим метод (предложенный Ламбальгеном [73]), с помощью которого можно строить случайные по Мизесу – Колмогорову – Лавлэнду последовательности с разными «патологическими» свойствами (не случайные по Мартин-Лёфу, с превышением нулей над единицами в начальных отрезках и другие).

9.13.1. Случайность по двум мерам

Для начала обсудим вопрос, имеющий и более общее значение: насколько зависит случайность (относительно данной меры) от выбора меры?

Вот два примера противоположного характера.

Пример 1. Рассмотрим равномерную меру μ, а также меру μ', в которой испытания независимы, причём во всех, кроме первого, вероятность успеха равна $1/2$, а в первом она равна, скажем, $2/3$. Интуитивно ясно, что случайными должны быть одни и те же последовательности: в самом деле, в первом испытании возможны оба исхода (и по той, и по другой мерам), а дальше меры одинаковы. Это действительно так для всех разумных определений случайности. Можно заметить, что эффективно нулевые множества относительно мер μ и μ' одни и те же (поскольку меры любого множества отличаются не более чем в два раза). Отсюда ясно, что случайными в смысле Мартин-Лёфа по этим мерам будут одни и те же последовательности.

$\boxed{\textbf{273}}$ Как доказать это, исходя из критерия случайности в терминах сложности?

274 Докажите, что случайными относительно вычислимых мартингалов (по мерам μ и μ') будут одни и те же последовательности.

Пример 2. Рассмотрим равномерную меру μ, а также меру μ', в которой испытания независимы и вероятность успеха в каждом из них равна $2/3$. Может ли одна и та же последовательность быть случайной (скажем, по Мартин-Лёфу) относительно μ и μ'? В полном согласии с нашей интуицией, нет — в самом деле, для случайной последовательности предел частот равен вероятности успеха, а $1/2 \neq 2/3$.

Возникает следующий естественный вопрос. Пусть имеются две последовательности чисел $p_i, p_i' \in (0,1)$. Рассмотрим меры μ и μ' на Ω, соответствующие независимым испытаниям, вероятность успеха в i-м равна p_i для меры μ и p_i' для меры μ'. Что можно сказать о классах случайных последовательностей относительно этих двух мер? Естественно ожидать, что при близких p_i и p_i' случайными должны быть одни и те же последовательности, а при сильно отличающихся p_i и p_i' классы случайных последовательностей не должны пересекаться.

Мы докажем, что это так и есть для случая отделённых от 0 и 1 вероятностей, когда все p_i и p_i' лежат в интервале $(\varepsilon, 1-\varepsilon)$ для некоторого положительного ε. Кроме того, мы предполагаем, что последовательности p_i и p_i' являются вычислимыми последовательностями вычислимых действительных чисел (иначе не имеет смысла говорить о случайности по Мартин-Лёфу).

Имеет место такой критерий, доказанный Вовком в [181] (конструктивный аналог классической теоремы Какутани [55]):

Теорема 196. (а) *Если сумма ряда $\sum(p_i - p_i')^2$ конечна, то классы случайных по Мартин-Лёфу последовательностей относительно мер μ и μ' совпадают.*

(б) *Если сумма ряда $\sum(p_i - p_i')^2$ бесконечна, то эти классы не пересекаются.*

Классический вариант этой теоремы [55] утверждает, что в первом случае любое нулевое множество относительно одной меры является нулевым относительно другой, а во втором случае меры ортогональны (существует множество меры 0 по одной и меры 1 по другой).

◄ Кажется естественным доказывать пункт (а) следующим образом. Пусть последовательность $\omega \in \Omega$ случайна по мере μ, соответствующей вероятностям p_i. Тогда (монотонная) сложность её начального отрезка длины n близка к логарифму его вероятности, которая равна произведению

$$\prod_{i=0}^{n-1} r_i,$$

где $r_i = p_i$ при $\omega_i = 1$ и $r_i = 1 - p_i$ при $\omega_i = 0$. Если p_i близко к p_i', то r_i близко к r_i' (где r_i' определяется аналогичным образом для другой меры). Поэтому произведение всех r_i близко к произведению всех r_i', так что случайность по одной мере влечёт за собой случайность и по другой.

Это рассуждение несложно формализовать, но при этом нам придётся предположить, что разность

$$\sum_{i=0}^{n-1}(-\log r_i) - \sum_{i=0}^{n-1}(-\log r_i')$$

ограничена; легко проверить, что это так, когда $\sum |p_i - p_i'| < \infty$ (напомним, что мы предполагаем p_i и p_i' отделёнными от нуля и единицы). А это больше, чем нам дано по условию; нам известно лишь, что сумма *квадратов* ограничена.

Как же быть? Заметим, что на самом деле нам достаточно, чтобы упомянутая разность была ограничена для любой *случайной* по мере μ последовательности ω. Мы знаем, что тогда

$$KM(\omega_0 \ldots \omega_{n-1}) = \sum_{i=0}^{n-1}(-\log r_i) + O(1)$$

(сложность отличается от логарифма вероятности не более чем на константу), поэтому верхняя оценка для сложности (которая имеет место для любой вычислимой меры, в частности, для μ') даёт нам

$$\sum_{i=0}^{n-1}(-\log r_i) \leqslant \sum_{i=0}^{n-1}(-\log r_i') + O(1).$$

Обозначим разницу между $r_i' - r_i$ через δ_i. Используя это обозначение и переходя к произведениям, получим (верное с точностью до мультипликативной константы) неравенство

$$\prod_{i=0}^{n-1} r_i \geqslant \prod_{i=0}^{n-1}(r_i + \delta_i).$$

Мы знаем, что $\sum_i \delta_i^2 < \infty$ и, в частности, $\delta_i \to 0$ при $i \to \infty$. Поэтому почти все δ_i меньше числа ε, отделяющего наши вероятности от нуля и единицы. Изменив конечное число членов p_i' (что не влияет на случайность), мы можем считать, что это верно для всех i. Тогда можно рассмотреть меру μ'', «симметричную» мере μ' относительно μ (это означает, что p_i есть середина отрезка $[p_i', p_i'']$). Для этой меры можно записать аналогичное неравенство, которое будет отличаться лишь знаком при δ_i:

$$\prod_{i=0}^{n-1} r_i \geqslant \prod_{i=0}^{n-1}(r_i - \delta_i)$$

(оно также верно с точностью до константы). Перемножим эти два неравенства:

$$\prod_{i=0}^{n-1} r_i^2 \geqslant \prod_{i=0}^{n-1}(r_i^2 - \delta_i^2).$$

А это неравенство (само по себе очевидное, и даже безо всякой константы), в силу наших предположений о δ_i является равенством (с точностью до ограниченного и отделённого от нуля множителя): как известно из курса математического анализа,

бесконечное произведение $\prod(1-h_i)$ больше нуля (при $0<h_i<1$) тогда и только тогда, когда $\sum h_i<\infty$.

Следовательно, и оба перемножаемых неравенства являются равенствами с точностью до константы. После этого критерий случайности (теорема 90) говорит, что последовательность ω случайна и по мере μ' (а также μ'', что для нас неважно). Утверждение (а) доказано.

Утверждение (б) является частным случаем более общей теоремы, в которой рассматриваются зависимые испытания, и мы докажем его сразу в общем случае (см. ниже теорему 197). ▶

Классический вариант теоремы Какутани можно получить как следствие алгоритмического.

275 Пусть $p_0,p_1,p_2,\ldots$ и $p'_0,p'_1,p'_2,\ldots$ находятся в интервале $(\varepsilon,1-\varepsilon)$ для некоторого $\varepsilon>0$, а меры μ и μ' соответствуют независимым испытаниям с вероятностями p_i и p'_i. Докажите, что $\sum(p_i-p'_i)^2<\infty$, то нулевыми по мерам μ и μ' являются одни и те же множества. Докажите, что если $\sum(p_i-p'_i)^2=\infty$, то меры μ и μ' ортогональны (существует множество меры 0 по одной из них и меры 1 по другой). [Указание. Воспользуйтесь теоремой 196 в релятивизованном варианте и заметьте, что для всякого нулевого множества можно найти оракул, относительно которого меры вычислимы, а множество эффективно нулевое. Во втором случае рассмотрите множества случайных по этим мерам последовательностей.]

Пусть μ — произвольная мера на пространстве Ω, а $p(x)$ — соответствующая функция на двоичных словах: $p(x)=\mu(\Omega_x)$. Для каждого слова x, которое имеет положительную вероятность появления в качестве начального отрезка последовательности (что означает $p(x)>0$), рассмотрим условную вероятность появления единицы после x, то есть отношение $p(1\,|\,x)=p(x1)/p(x)$. (Если бросания независимы, то эта величина зависит лишь от длины слова x и обозначалась нами раньше через p_i.)

Если теперь, помимо меры, задана ещё и последовательность $\omega\in\Omega$, то можно рассмотреть условные вероятности появления единицы на каждом шаге для этой последовательности, то есть последовательность чисел

$$p_i=p(1\,|\,\omega_0\ldots\omega_{i-1})=p(\omega_0\ldots\omega_{i-1}1)/p(\omega_0\ldots\omega_{i-1})$$

(которые заведомо определены, если p не обращается в нуль на начальных отрезках последовательности). Следующий результат [181] показывает, что это предсказание не сильно зависит от выбора меры, по которой данная последовательность случайна:

Теорема 197. *Если последовательность ω случайна в смысле Мартин-Лёфа по двум мерам μ и μ', и числа p_i и p'_i, построенные описанным образом для этих мер, все находятся в интервале $(\varepsilon,1-\varepsilon)$ для некоторого $\varepsilon>0$, то*

$$\sum_i(p_i-p'_i)^2<\infty.$$

Заметим, что отсюда прямо следует утверждение теоремы 196. Отметим также, что для случайной последовательности числа p_i определены, так как знаменатель дроби $p(x1)/p(x)$ не может обращаться в нуль (слово, имеющее нулевую вероятность, не может быть началом случайной последовательности).

◄ Рассмотрим третью меру $\tilde{\mu}$, которая усредняет вероятности появления единицы для мер μ и μ'. А именно, вероятность появления единицы после любого слова x с точки зрения меры $\tilde{\mu}$ равна среднему арифметическому вероятностей появления единицы после того же x с точки зрения мер μ и μ'. (Заметим, что это *не* соответствует усреднению мер: если положить $\tilde{\mu}(X) = (\mu(X) + \mu'(X))/2$, тоже получится мера, но совсем другая!)

Если последовательность случайна, то для сложности её начальных отрезков можно записать равенство

$$KA\,(\omega_0 \ldots \omega_{n-1}) = \sum_{i=0}^{n-1} (-\log r_i) + O(1),$$

где, как и раньше, $r_i = p_i$ при $\omega_i = 1$ и $r_i = (1 - p_i)$ при $\omega_i = 0$. Аналогичное равенство можно записать и для r_i' (построенных для меры μ'), а для $\tilde{r}_i$ (построенных по мере $\tilde{\mu}$) можно записать лишь неравенство (сложность не больше минус логарифма меры).

Исключая из этих равенств и неравенств колмогоровскую сложность, получаем неравенства для условных вероятностей:

$$\sum_{i=0}^{n-1} (-\log r_i) \leqslant \sum_{i=0}^{n-1} (-\log \tilde{r}_i) + O(1),$$
$$\sum_{i=0}^{n-1} (-\log r_i') \leqslant \sum_{i=0}^{n-1} (-\log \tilde{r}_i) + O(1).$$

После сложения и деления пополам получается неравенство

$$\sum_{i=0}^{n-1} \frac{(-\log r_i) + (-\log r_i')}{2} \leqslant \sum_{i=0}^{n-1} (-\log \tilde{r}_i) + O(1).$$

Вспоминая, что $\tilde{r}_i = (r_i + r_i')/2$ и перенося всё в одну часть, получаем

$$\sum_{i=0}^{n-1} \left(\frac{(-\log r_i) + (-\log r_i')}{2} - \left(-\log \frac{r_i + r_i'}{2} \right) \right) \leqslant O(1).$$

Каждое слагаемое суммы в левой части неотрицательно (выпуклость логарифма) и по порядку величины равно $(r_i - r_i')^2$, поэтому $\sum (r_i - r_i')^2 \sim \sum (p_i - p_i')^2 < \infty$, что и требовалось доказать.

(В этом рассуждении есть один деликатный момент, который требует дополнительного обсуждения. Мы усредняли условные вероятности, что вызывает трудности, если одна из мер равна нулю на некотором цилиндре. В этом случае мы так и не сможем вычислить эти условные вероятности, и построение меры застопорится.

Однако мы можем построить полумеру, и этого нам достаточно, а последовательность ω через эти особые точки не пройдёт, так как случайна по обеим мерам.) ▶

Отметим, что в этой теореме говорится лишь о близости мер вдоль последовательности ω, случайной по обеим мерам; на других последовательностях они могут и сильно различаться.

276 Приведите соответствующий пример. [Указание. Рассмотрим две меры, которые на левой половине отрезка равномерны, а на правой сильно различаются.]

277 Покажите, что фактически мы использовали не случайность по Мартин-Лёфу, а более слабое свойство случайности относительно вычислимых мартингалов (в предположении, что обе меры ненулевые для всех слов и условные вероятности определены).

Приведённое доказательство имеет естественную игровую интерпретацию. Представим себе, что имеются два тотализатора, принимающих ставки на одну и ту же последовательность событий (каждое событие имеет два исхода: нуль и единицу). Однако их организаторы по-разному оценивают шансы и поэтому принимают ставки на разных условиях. В одном случае вероятности появления единицы и нуля считаются равными p и q (и потому поставленная на 1 [0] сумма возвращается с коэффициентом $1/p$ [$1/q$]), а в другом случае вероятности считаются равными p' и q' (а коэффициенты равны $1/p'$ и $1/q'$). (Вероятности p, p', q, q' могут меняться от события к событию).

В этом случае можно одновременно участвовать в двух играх (в каждой — со своим капиталом; передавать деньги из одного игрового зала в другой нельзя), причём делать это так, чтобы по крайней мере в одной выиграть. Более того, в случае, когда все p, q, p', q' отделены от нуля и сумма $\sum(p-p')^2$ (по всей последовательности событий) бесконечна, то хотя бы в одной из игр наш выигрыш будет неограничен.

Как этого добиться? Пусть в первой игре у нас имеется капитал u, оценки вероятностей для следующего события равна p, p', q, q'. Поделим u между двумя ставками так, чтобы после игры наш капитал был равен $(p+p')u/2p$ (если выпадет единица) и $(q+q')u/2q$ (если выпадет нуль). (Легко проверить, что это допускается правилами игры, то есть что математическое ожидание капитала после партии равно u.) Во второй игре капитал v превращается после игры в $(p+p')v/2p'$ (если выпадет единица) и $(q+q')v/2q'$ (если выпадет нуль). Будем следить за произведением капиталов в обеих играх. Оно умножается либо на $(p+p')^2/4pp'$, либо на $(q+q')^2/4qq'$. Неравенство о среднем арифметическом и среднем геометрическом показывает, что в обоих случаях произведение uv возрастает. Переходя к логарифмам и оценивая этот рост, легко заметить, что (для отделённых от нуля p, p', q, q') при $\sum(p-p')^2 = \infty$ произведение капиталов стремится к бесконечности, и, следовательно, хотя бы в одной игре капитал не ограничен.

Другими словами, мы построили для каждой из мер свой мартингал с таким свойством: на любой последовательности, для которой условные вероятности по обеим мерам отделены от 0 и 1 и различия в условных вероятностях имеют бесконечную сумму квадратов, хотя бы один из мартингалов неограничен.

9.13.2. Закон больших чисел для переменных вероятностей

Усиленный закон больших чисел говорит, что для любого $p \in (0,1)$ и соответствующей ему бернуллиевой меры μ_p (испытания независимы, вероятность успеха в каждом равна p) множество последовательностей, у которых предел частот равен p, имеет меру 1 (а его дополнение, то есть множество последовательностей, не имеющих предела частот или имеющих другой предел частот, является нулевым).

Пусть теперь вероятности не постоянны, и в i-м испытании вероятность успеха равна p_i. Естественно ожидать, что если $p_i \to p$, то предел частот для почти всех последовательностей будет равен p. Не вполне строго это можно объяснить так: возьмём какое-то $\varepsilon > 0$. Все частоты p_i, кроме конечного числа (а им можно пренебречь), меньше $p + \varepsilon/2$. Мы знаем, что если заменить вероятности p_i на $p + \varepsilon/2$, то почти наверное начиная с некоторого места частота будет меньше $p + \varepsilon$. Значит, это тем более так для $p_i < p + \varepsilon/2$ по монотонности.

Это рассуждение можно (с некоторым трудом, впрочем) сделать строгим, но мы предпочитаем доказать более общую оценку, которая полезна во многих случаях. Она относится к произвольной мере μ на пространстве Ω; пусть p — соответствующая ей функция на словах. Для любого конечного слова $x = x_0 \ldots x_{n-1}$ длины n рассмотрим число единиц m в этом слове, а также сумму $p_0 + \ldots + p_{n-1}$ условных вероятностей появления единицы в каждой его позиции:

$$p_i = p(x_0 \ldots x_{i-1}1)/p(x_0 \ldots x_{i-1}).$$

Обе эти величины зависят от слова x; наша оценка утверждает, что с большой вероятностью (по мере μ) для данной длины n эти величины будут близки. Вот точная формулировка (смысл m и p_i объяснён только что, вероятность берётся по мере μ):

Теорема 198.

$$\Pr[|m - (p_0 + \ldots + p_{n-1})| > d] \leqslant 2e^{-d^2/4n}.$$

Заметим, что бесконечные последовательности в этой формулировке по существу не нужны; фактически речь идёт о неравенстве, верном для любого распределения вероятностей на последовательностях нулей и единиц длины n. Это неравенство является ослабленной формой *неравенства Азумы–Хёфдинга*, известного в теории вероятностей; мы приведём простое рассуждение по аналогии с уже встречавшимися (сравнение мер), хотя и не дающее наиболее сильного утверждения (можно было бы заменить константу 4 на 2).

◄ Отдельно оцениваем вероятность избытка и недостачи; оба варианта симметричны, поэтому достаточно доказать, что

$$\Pr[m - (p_0 + \ldots + p_{n-1}) > d] \leqslant e^{-d^2/4n}.$$

Мы используем обычный приём: строим другую меру μ', для которой отношение μ'/μ велико на всех последовательностях, попадающих в событие из левой части.

(Отношение μ'/μ будет мартингалом, который велик, а именно, не меньше $e^{d^2/4n}$ на всех таких последовательностях.)

Поскольку мы хотим увеличения меры для последовательностей, в которых много единиц, то естественно, что в новой мере нужно условную вероятность единицы увеличить (по сравнению с μ). А именно, если раньше после некоторого слова x условные вероятности единицы и нуля были p и q, то теперь мы положим их равными

$$p' = p + \varepsilon pq, \quad q' = q - \varepsilon pq,$$

где ε — некоторое положительное число, меньшее $1/2$ (его мы выберем позже). Легко проверить, что p' и q' не выйдут за пределы отрезка $[0,1]$.

Получаем новое распределение вероятностей на последовательностях нулей и единиц длины n. Во сколько раз оно отличается от исходного на некотором слове x длины n? Каждый новый символ слова x даёт множитель p'/p (если это была единица) или q'/q (если это был нуль), где p, q, p', q' — условные вероятности единицы и нуля в исходной и изменённой мерах. Переходя к логарифмам, видим, что логарифм интересующего нас отношения равен сумме величин

$$\ln(p'/p) = \ln(1 + \varepsilon q) \geqslant \varepsilon q - (\varepsilon q)^2 \geqslant \varepsilon(1 - p) - \varepsilon^2$$

или

$$\ln(q'/q) = \ln(1 - \varepsilon p) \geqslant -\varepsilon p - (\varepsilon p)^2 \geqslant -\varepsilon p - \varepsilon^2$$

для всех букв слова x (с соответствующими вероятностями p и q); первый вариант для единицы и второй вариант для нуля. (Мы использовали неравенство $\ln(1 + h) \geqslant h - h^2$, выполненное при всех $|h| \leqslant 1/2$.)

Что получится, если сложить все величины? Помимо общего множителя ε, в правой части будет число единиц m (от каждого слагаемого $(1 - p)$ по единице) минус сумма всех p_i, где p_i — условная вероятность единицы на очередном месте в слове x, и ещё минус $n\varepsilon^2$:

$$\ln \frac{p'(x)}{p(x)} \geqslant \varepsilon(m - \sum p_i) - n\varepsilon^2.$$

Для тех x, которые нас интересуют (у которых превышение больше d) имеем

$$\ln \frac{p'(x)}{p(x)} > \varepsilon d - n\varepsilon^2 = \varepsilon(d - n\varepsilon).$$

Это верно при всех $\varepsilon \in (0, 1/2)$, поэтому выберем то из них, при котором правая часть наибольшая, а именно $\varepsilon = d/2n$ (ясно, что случай $d > n$ можно не рассматривать, так как число единиц никогда не превосходит n, поэтому $\varepsilon = d/2n < 1/2$).

Получаем, что

$$\ln \frac{p'(x)}{p(x)} > d^2/4n \quad \text{и} \quad \frac{p'(x)}{p(x)} > e^{d^2/4n},$$

что нам и требовалось. ▶

(Поучительно посмотреть, какая оценка получается для простейшего случая, когда нули и единицы равновероятны. Вероятность того, что число единиц превосходит ожидаемое $n/2$ более чем на $2\sqrt{n}$, меньше $1/e$, а превышение более чем на $2k\sqrt{n}$ имеет вероятность не больше $1/e^{k^2}$. Видно, что наша оценка не оптимальная — теорема Муавра – Лапласа даёт немного лучше — но близка к ней, отличие лишь в коэффициенте в показателе экспоненты.)

Теперь можно повторить доказательство усиленного закона больших чисел с этой новой оценкой и получить такую теорему, верную для произвольной меры μ на пространстве Ω:

Теорема 199. *Для последовательности $\omega = \omega_0\omega_1 \ldots$, распределённой по мере μ, почти наверное выполнено такое свойство*:

$$\lim_{n\to\infty}\left(\frac{m}{n} - \frac{p_0 + \ldots + p_{n-1}}{n}\right) = 0,$$

где m — число единиц среди $\omega_0, \omega_1, \ldots, \omega_{n-1}$, а p_i — условная вероятность появления единицы после $\omega_0\omega_1 \ldots \omega_{i-1}$.

В частности, из этой теоремы следует, что если испытания независимы, а вероятности успеха p_i стремятся к некоторому пределу p, то с вероятностью единица предел частот равен p (поскольку предел средних $(p_0 + \ldots + p_{n-1})/n$ равен p).

Мы, однако, хотим большего — чтобы с вероятностью единица последовательность ω была случайной по Мизесу (в том или ином варианте), то есть чтобы не только для самой последовательности, но и для её «законно выбранных» подпоследовательностей предел частот равнялся p.

Для начала рассмотрим случай монотонных правил выбора (задаваемых множествами слов, после которых делается ставка).

9.13.3. Закон больших чисел для подпоследовательностей

Докажем обобщение теоремы 198, включив туда правила выбора. В этой теореме мы рассматривали меру μ на пространстве Ω и соответствующие вероятности $p(x) = \mu(\Omega_x)$ и $p(1\,|\,x) = p(x1)/p(x)$. Для данной последовательности ω и данного числа n можно рассмотреть две величины:

- число единиц среди первых n битов ω;
- сумму условных вероятностей единицы на первых n позициях, то есть

$$p(1\,|\,\Lambda) + p(1\,|\,\omega_0) + p(1\,|\,\omega_0\omega_1) + \ldots + p(1\,|\,(\omega_0 \ldots \omega_{n-1})).$$

Теорема 198 гарантирует, что при любом n две эти величины могут сильно отличаться лишь для небольшой (по мере μ) части последовательностей ω, а именно, μ-мера множества тех последовательностей ω, где разница превышает d, не больше $2e^{-d^2/4n}$.

Теперь мы добавляем в эту схему правило выбора S_R, задаваемое некоторым множеством R (тех слов, после которых надо выбирать следующий член). В любой последовательности ω правило выбора S_R указывает позиции $i_0, i_1, \ldots$, в которых стоят члены выбираемой подпоследовательности ($S_R(\omega) = \omega_{i_0}\omega\, i_1 \ldots$). Теперь мы сравниваем

- число единиц среди первых n *выбранных* членов;
- сумму условных вероятностей появления единицы на первых n *выбранных* позициях, то есть

$$p(1\,|\,\omega_0\omega_1\ldots\omega_{i_0-1})+p(1\,|\,\omega_0\omega_1\ldots\omega_{i_1-1})+\ldots+p(1\,|\,\omega_0\omega_1\ldots\omega_{i_n-1}).$$

Если R содержит все слова, то получаются величины из предыдущей теоремы, но теперь мы хотим доказать, что та же самая оценка верна и в общем случае:

Теорема 200. *При любых R и n вероятность (по мере μ) того, что эти две величины отличаются на d, не превосходит $2e^{-d^2/4n}$.*

Вообще говоря, для некоторых последовательностей ω выбранная подпоследовательность $S_R(\omega)$ содержит менее n членов; такие ω не входят в событие, вероятность которого оценивается в этой теореме.

Пример. Рассмотрим произвольную меру и правило выбора, которое отбирает те члены последовательности, для которых условная вероятность появления единицы меньше 0,5. Тогда утверждение теоремы гарантирует, что доля единиц среди первых n выбранных членов превышает 51% лишь с малой (экспоненциально убывающей с ростом n) вероятностью.

◀ Эта теорема также является простым следствием общего неравенства Азумы – Хёфдинга, для мартингалов в смысле теории вероятностей, но можно обойтись и без ссылки на него, действуя так же, как при доказательстве теоремы 198.

Теперь в формулировке теоремы нельзя ограничиться фиксированным числом битов в последовательности ω (поскольку выбранные n членов могут быть сколь угодно далеко). Но можно доказать конечный вариант этой теоремы для последовательностей длины N, а затем по непрерывности меры устремить N к бесконечности.

Как и раньше, мы сравниваем ω с другой мерой, но теперь мы меняем условные вероятности только в тех позициях, где очередной член включается в подпоследовательность (а в остальных случаях условные вероятности остаются неизменными). Тогда в оценку для отношения вероятностей будет входить длина подпоследовательности (вместо длины всей последовательности), число выбранных единиц (вместо числа всех единиц) и сумма вероятностей в моменты выбора (вместо суммы всех вероятностей), то есть получится ровно то, что требуется. ▶

Доказанная теорема позволяет, следуя Ламбальгену [73], строить случайные по Мизесу – Чёрчу – Дэли последовательности. Вот как это делается.

Пусть $p_0, p_1, \ldots$ — вычислимая последовательность вычислимых действительных чисел, вычислимо сходящаяся к числу $p \in (0,1)$. (Вычислимая сходимость означает, что по данному $\varepsilon > 0$ можно алгоритмически указать то N, начиная с которого члены последовательности отстоят от предела меньше чем на ε; ясно, что тогда p вычислимо.) Рассмотрим вычислимую меру μ, соответствующую независимым испытаниям с вероятностями p_i.

Теорема 201. *Всякая случайная в смысле Мартин-Лёфа по мере μ последовательность является случайной по Мизесу – Чёрчу – Дэли с пределом частот p.*

◀ Пусть фиксированы некоторое (рациональное) $\varepsilon > 0$ и некоторое допустимое в смысле Чёрча – Дэли (вычислимое, возможно не всюду определённое) правило выбора R. Мы должны показать, что множество U тех последовательностей, для которых применение этого правила даёт бесконечную последовательность и частота единиц в ней бесконечно много раз превышает $p + \varepsilon$, является эффективно нулевым. (Аналогично для частот, меньших $p - \varepsilon$.)

Фиксируем некоторое n и рассмотрим множество U_n тех ω, для которых выбранная подпоследовательность имеет длину не менее n и доля единиц среди первых n членов больше $p + \varepsilon$. Это множество эффективно открыто (применяя правило выбора вдоль всех ветвей, мы можем перечислять все последовательности, продолжения которых попадают в U_n).

Теорема 200 даёт верхнюю оценку для множества U_n; для достаточно больших n среднее арифметическое вероятностей не больше $p + \varepsilon/2$ (и границу можно найти, пользуясь вычислимой сходимостью), и даваемая этой теоремой оценка экспоненциально убывает с ростом n. Поэтому можно покрыть U перечислимым семейством интервалов сколь угодно малой меры (взяв все интервалы, образующие множества $U_N, U_{N+1}, U_{N+2}, \ldots$ для достаточно большого N; напомним, что всякий элемент U по определению входит в бесконечно много разных U_n). Следовательно, U является эффективно нулевым и не может содержать случайной в смысле Мартин-Лёфа последовательности. ▶

На самом деле в тех же предположениях верно и более сильное утверждение:

Теорема 202. *Всякая случайная в смысле Мартин-Лёфа по мере μ последовательность является случайной по Мизесу – Колмогорову – Лавлэнду с тем же пределом частот p.*

◀ Применение допустимого по Колмогорову – Лавлэнду правила $S_{F,G}$ к последовательности можно разбить на два этапа. Сначала с помощью функции F отбирается подпоследовательность просмотренных членов (как выбранных, так и не выбранных). Затем с помощью функции G производится выбор, и эта вторая часть соответствует применению допустимого по Чёрчу-Дэли правила.

Посмотрим на результат первого этапа. Как распределена получающаяся последовательность ω_F, если исходная последовательность ω получается в результате независимых испытаний с вероятностью успеха p_i в i-м испытании? Будем для начала считать, что функции F и G всюду определены.

Первый член ω_F имеет в ω заранее известный номер $n_0 = F(\Lambda)$ (наше правило начинает с просмотра этого члена). Вероятность появления единицы в начале ω_F есть, таким образом, p_{n_0}. Какой член будет вторым в ω_F? Это уже зависит от результата первого просмотра. Соответственно для вероятности появления единицы на втором месте есть две возможности, соответствующие двум разным позициям в ω. Общее правило: вероятность появления 1 после некоторого слова x в ω_F равна $p_{F(x)}$, поскольку в этом случае следующим будет просмотрен член с номером $F(x)$. Заметим, что в силу ограничений на F (никакой член не просматривается дважды) условные вероятности вдоль любой ветви образуют подпоследовательность последовательности p_i без повторений, и потому среди них вне интервала $(p - \varepsilon/2, p + \varepsilon/2)$ окажется не больше, чем в исходной последовательности.

Это позволяет применить оценку теоремы 200 к последовательности ω_F и правилу выбора, задаваемому функцией G, как и в доказательстве теоремы 201.

Та же самая оценка верна и для частичных F и G, поскольку от их произвольного продолжения (вычислимого может не быть, но это и не нужно) оцениваемое множество может только увеличиться.

Поскольку функции F и G вычислимы, множество тех последовательностей ω, у которых правило выбора $S_{F,G}$ даёт подпоследовательность, начинающееся на данное слово x, может быть эффективно перечислено по x, так что после получения оценки на вероятность больших уклонений частоты от p мы легко строим нужное нам эффективно нулевое множество. ▶

Замечание. Можно было бы дать и более прямое доказательство теоремы 202. Вот одно из возможных рассуждений (заимствованное из [153]).

Пусть фиксированы некоторое (рациональное) $\varepsilon > 0$ и функции F, G, задающие допустимое в смысле Колмогорова – Лавлэнда правило выбора R. Мы должны показать, что множество U тех последовательностей, для которых применение этого правила даёт бесконечную последовательность и частота единиц в ней бесконечно много раз превышает $p + \varepsilon$, является эффективно нулевым относительно меры μ. (Аналогично для частоты, меньшей $p - \varepsilon$.)

Обозначим через U_n множество тех последовательностей ω, для которых $S_{F,G}(\omega)$ содержит не менее n членов и частота единиц среди первых n членов больше $p + \varepsilon$. Достаточно доказать, что ряд $\sum \mu(U_n)$ вычислимо сходится (где μ — рассматриваемая нами мера, соответствующая независимым испытаниям с вероятностями успеха p_i).

Обозначим через $\alpha_{n,k}(r_1, \ldots, r_n)$ вероятность того, что в n независимых испытаниях с вероятностями успеха $r_1, \ldots, r_n$ случилось не менее k успехов. Функция $\alpha_{n,k}$ неубывает по всем своим аргументам (и, кстати, является полилинейной, то есть многочленом от r_i степени не более 1 по каждому аргументу). Кроме того, $\alpha_{n,k} \leqslant \alpha_{n,l}$ при $k \geqslant l$.

Мы утверждаем, что $\mu(U_n) \leqslant \alpha_{n,k}(r_1, \ldots, r_n)$, где $k = n(p + \varepsilon)$, а r_i — i-й по величине член последовательности $p_0, p_1, \ldots$ (r_1 — максимальный, r_2 — максимальный из оставшихся и т. д.; точнее, r_1 — точная верхняя грань всех p_i; r_2 — точная верхняя грань минимумов пар $\min(p_i, p_j)$ при $i \neq j$ и так далее).

Покажем, как из этой оценки можно получить сходимость ряда $\mu(U_n)$. Очевидно, что $r_1 \geqslant r_2 \geqslant \ldots$ и $\lim r_i = p$. Заменим те r_i, которые больше $p + \varepsilon/2$, на единицу (пусть их количество равно s), а остальные — на $p + \varepsilon/2$. Получим, что

$$\mu(U_n) \leqslant \alpha_{n-s,k-s}(p + \varepsilon/2, \ldots, p + \varepsilon/2),$$

так что мы свели дело к оценке уклонений для случая постоянных вероятностей, который мы уже многократно разбирали. (Важно отметить, что при больших n отношение $(k - s)/(n - s)$ примерно равно $p + \varepsilon$ и заметно больше $p + \varepsilon/2$.)

Осталось доказать указанную оценку для $\mu(U_n)$. Можно представлять себе (как мы уже обсуждали), что члены последовательности написаны на карточках, лежащих лицом вниз, и что правило выбора говорит, какие карточки нужно перевернуть для просмотра и какие — для включения в выбранную подпоследовательность.

При этом сведения о том, какие карточки переворачивались, и что на них оказалось, записываются в протокол применения правила выбора к последовательности. Пусть π — начальный отрезок такого протокола. Через $n(\pi)$ обозначим количество членов, включённых в подпоследовательность на отрезке π, а через $k(\pi)$ — количество единиц среди этих членов. Через $r_i(\pi)$ обозначим i-й по величине член последовательности, получаемой из $p_0, p_1, \ldots$ выкидыванием членов, соответствующим перевёрнутым в ходе π карточкам (независимо от того, были ли они включены в подпоследовательность или только просмотрены).

Через $\mu(U_n \mid \pi)$ обозначим условную вероятность события U_n при условии того, что протокол применения правила выбора начинается на π. Мы докажем следующее обобщение интересующего нас неравенства: при $n(\pi) \leqslant n$

$$\mu(U_n \mid \pi) \leqslant \alpha_{n-n(\pi),k-k(\pi)}(r_1(\pi), r_2(\pi), \ldots) \tag{$*$}$$

(при пустом π получаем интересующую нас оценку). Если $n(\pi) = n$, то это неравенство превращается в равенство (левая и правая части одновременно равны либо нулю, либо единице). Пусть $n(\pi) < n$ и m — номер карточки, переворачиваемой сразу после π (если такой нет, то $\mu(U_n \mid \pi) = 0$). Тогда

$$\mu(U_n \mid \pi) = p_m \mu(U_n \mid \pi_1) + (1 - p_m)\mu(U_n \mid \pi_0),$$

где π_0 и π_1 — протоколы, получаемые добавлением к π информации о нуле (единице) на m-й карточке. Покажем, что если доказываемое нами неравенство ($*$) верно для π_0 и π_1, то оно верно и для π. В самом деле, тогда $\mu(U_n \mid \pi)$ не превосходит

$$p_m \alpha_{n-n(\pi_1),k-k(\pi_1)}(r_1(\pi_1), \ldots) + (1 - p_m)\alpha_{n-n(\pi_0),k-k(\pi_0)}(r_1(\pi_0), \ldots). \tag{$**$}$$

Если m-я карточка была выбрана только для просмотра, то $n(\pi_0) = n(\pi_1) = n(\pi)$ и $k(\pi_0) = k(\pi_1) = k(\pi)$, и остаётся воспользоваться монотонностью $\alpha_{n,k}$ и тем, что $r_i(\pi_0) = r_i(\pi_1) \leqslant r_i(\pi)$. Если же m-я карточка включена в подпоследовательность, то $n(\pi_0) = n(\pi_1) = n(\pi) + 1$, $k(\pi_0) = k(\pi)$ и $k(\pi_1) = k(\pi) + 1$, а выражение ($**$) равно

$$\alpha_{n-n(\pi),k-k(\pi)}(p_m, r_1(\pi_1), r_2(\pi_1), \ldots)$$

(заметим, что $r_i(\pi_1) = r_i(\pi_0)$) и тем самым не превосходит

$$\alpha_{n-n(\pi),k-k(\pi)}(r_1(\pi), r_2(\pi), \ldots).$$

Это завершает доказательство неравенства ($*$) в том случае, когда все начальные отрезки протоколов π с $n(\pi) = n$ имеют ограниченную длину. Если же это не так, то приведённые рассуждения позволяют получить оценку для $\mu(U_{n,t} \mid \pi)$, где $U_{n,t}$ — множество тех последовательностей, для которых после не более чем t переворачиваний карточек будет выбрана подпоследовательность длины не меньше n с числом единиц (в начальном отрезке длины n) не менее k. Остаётся лишь перейти к пределу при $t \to \infty$.

9.13.4. Примеры

Теперь уже легко построить примеры случайных по Мизесу – Колмогорову – Лавлэнду последовательностей с различными патологическими свойствами.

Теорема 203. (а) *Существует случайная по Мизесу – Колмогорову – Лавлэнду последовательность с частотой* $1/2$, *не случайная по Мартин-Лёфу относительно равномерной меры.*

(б) *Существует случайная по Мизесу – Колмогорову – Лавлэнду последовательность с частотой* $1/2$, *любой начальный отрезок которой содержит не меньше нулей, чем единиц.*

(в) *Существует случайная по Мизесу – Колмогорову – Лавлэнду последовательность с частотой* $1/2$, *из которой с помощью допустимого по Мизесу – Колмогорову – Лавлэнду* (*и даже по Чёрчу*) *правила выбора получается не случайная по Мизесу – Колмогорову – Лавлэнду последовательность.*

(г) *Существует случайная по Мизесу – Чёрчу–Дэли последовательность, становящаяся неслучайной по Мизесу – Чёрчу после вычислимой перестановки.*

◄ (а) Рассмотрим вычислимую последовательность рациональных чисел из интервала $(0,1)$, которая сходится к $1/2$, но медленно, например,

$$p_i = \tfrac{1}{2} - \tfrac{1}{\sqrt{i+5}}$$

(число 5 добавлено, чтобы p_i были положительны). Рассмотрим (вычислимую) меру μ, соответствующую независимым испытаниям с вероятностями успеха p_i.

Поскольку ряд $\sum(p_i - 1/2)^2$ расходится, никакая случайная по Мартин-Лёфу (относительно меры μ) последовательность ω не будет случайной по Мартин-Лёфу относительно равномерной меры (теорема 197). С другой стороны, любая такая последовательность будет случайной по Мизесу – Колмогорову – Лавлэнду с предельной частотой $1/2$ (теорема 202).

(б) Это утверждение можно доказать аналогично предыдущему, только нужно взять последовательность p_i, ещё медленнее стремящуюся к пределу $1/2$. Пусть, например,

$$p_i = \tfrac{1}{2} - \tfrac{1}{\log(i+5)}.$$

Какова вероятность (по мере μ) события «начальный отрезок длины n содержит меньше нулей, чем единиц»? Другими словам, насколько вероятно, что частота единиц в этом начальном отрезке больше $1/2$? По теореме 198 эта вероятность (обозначим её δ_n) не больше $e^{-n/O(\log^2 n)}$ (граница d в этой теореме есть примерно $n/O(\log n)$, и $d^2/4n = n/O(\log^2 n)$). Ряд $\sum_n \delta_n$ сходится, и поэтому для некоторого N остаточный член ряда меньше единицы, и тем самым вероятность события «все начальные отрезки, имеющие длину N и более, содержат не меньше нулей, чем единиц» положительна. Множество положительной μ-меры обязано содержать хотя бы одну случайную в смысле Мартин-Лёфа (относительно меры μ) последовательность, поэтому существует случайная в смысле Мартин-Лёфа относительно меры μ последовательность, у которой все начальные отрезки, начиная с длины N, содержат не меньше нулей, чем единиц. Легко понять, что заменив первые N членов

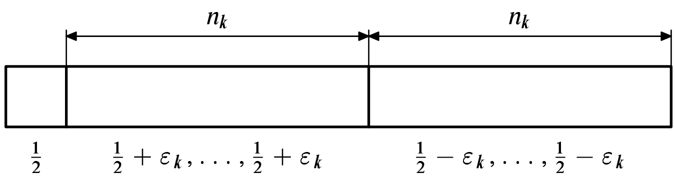


Рис. 9.2. Участок номер k: вероятности.

на нули, мы получим случайную по Мартин-Лёфу относительно меры μ последовательность, у которой *все* начальные отрезки содержат не меньше нулей, чем единиц. Остаётся сослаться на теорему 202.

(в) И здесь нужный пример, как выяснил Меркле [101], доставляет последовательность, случайная относительно меры μ, соответствующей независимым испытаниям с вероятностью успеха p_i. Однако p_i будут устроены некоторым более сложным образом. Разобьём последовательность на участки; k-й по счёту участок состоит из одного члена, для которого вероятность равна $1/2$, и двух кусков одинаковой (и достаточно большой) длины n_k. В первом из них вероятности одинаковы и чуть больше $1/2$, а во втором — одинаковы и чуть меньше $1/2$ (рис. 9.2).

При этом ε_k положительно и стремится к нулю с ростом k, но достаточно медленно. Точнее говоря, нам нужно такое соотношение между n_k и ε_k: вероятность того, что среди n_k независимых испытаний, в каждом из которых вероятность единицы равна $\frac{1}{2}+\varepsilon_k$, будет больше единиц, чем нулей, не меньше $1-2^{-(k+3)}$. (Этого несложно добиться для любых ε_k, если взять n_k достаточно большими.)

Если так, то с положительной вероятностью мы получим последовательность, в которой при всех k в каждом из кусков (для каждого k имеется два куска длиной n_k) имеется дисбаланс между нулями и единицами в ожидаемую сторону (в левом куске длиной n_k больше единиц, чем нулей, а в правом — наоборот).

Следовательно, существует случайная по Мартин-Лёфу (относительно построенной нами меры μ) последовательность ω с таким дисбалансом. По теореме 202 она будет случайной по Мизесу – Колмогорову – Лавлэнду.

Покажем теперь, как использовать дисбаланс, чтобы построить (допустимое по Чёрчу) правило выбора, выбирающее из ω неслучайную подпоследовательность ω'. Правило это совсем простое: первый член каждого участка («ключ») выбирается всегда, а в зависимости от его значения (0 или 1) выбирается целиком левый кусок (без правого) или целиком правый (без левого).

Почему ω' не будет случайна по Мизесу – Колмогорову – Лавлэнду? Это совсем просто: условие дисбаланса позволяет восстановить значение ключа по числу единиц в следующем за нём куске длиной n_k. (Мы сначала читаем «в порядке информации» n_k членов, идущих за ключом, а потом отгадываем значение ключа.)

Это доказывает утверждение (в). Попутно можно заметить, что последовательность ω' случайна по Мизесу – Чёрчу – Дэли (поскольку двукратное применение допустимого по Чёрчу – Дэли правила сводится к однократному). С другой стороны, если мы подвергнем ω' вычислимой перестановке членов, переместив ключ

в позицию после управляемого им участка, то получится последовательность, не случайная по Мизесу – Чёрчу. Утверждение (г) доказано. ►

Заметим, что эта теорема в значительной степени дискредитирует наши уточнения случайности в стиле Мизеса. Дело в том (такое свойство коллективов отмечал ещё сам Мизес), что применение допустимого правила выбора к коллективу должно давать коллектив. Утверждение (в) показывает, что определение Мизеса – Колмогорова – Лавлэнда этим свойством, увы, не обладает. Определения с монотонными правилами (Мизеса – Чёрча и Мизеса – Чёрча – Дэли) этим свойством обладают (двукратное применение монотонных правил сводится к однократному), но зато неустойчивы относительно вычислимых перестановок, что тоже нехорошо.

278 Докажите, что применение допустимого по Мизесу – Колмогорову – Лавлэнду правила к случайной по Мизесу – Колмогорову – Лавлэнду последовательности даёт случайную по Мизесу – Чёрчу – Дэли последовательность.

279 Покажите, что существует случайная по Мизесу – Колмогорову – Лавлэнду последовательность (с частотой 1/2), не случайная даже по Курцу. [Указание. Возьмём случайную по сдвинутой мере последовательность: для неё один из мартингалов, построенных в доказательстве теоремы 197, ограничен, поэтому второй имеет вычислимую стремящуюся к бесконечности нижнюю оценку.]

Приведённые в этом разделе примеры показывают, что определение случайности по Мизесу – Колмогорову – Лавлэнду, видимо, накладывает слишком слабые требования на последовательность. То же самое можно сказать и про определение случайности относительно вычислимых мартингалов. Возникает естественный вопрос: что будет, если скомбинировать эти определения, то есть разрешить играть с последовательностью битов, написанных на карточках, переворачивая карточки в произвольном порядке (по выбору игрока) и ставя любую часть капитала на тот или иной исход (алгоритмически)? Легко видеть, что получится класс последовательностей (их иногда называют «непредсказуемыми»; английское название — *Kolmogorov – Loveland random sequences*), содержащий все случайные по Мартин-Лёфу последовательности, но до сих пор так и не выяснено, совпадают ли эти два класса. Можно заметить ещё, что для немонотонных мартингалов не имеет значения, являются ли они всюду определёнными (применимо то же соображение Меркле, что и в теореме 191, с. 326).

В заключение попытаемся резюмировать соотношения между рассмотренными определениями случайности (рис. 9.3). У нас есть две линии последовательно ослабляющихся определений: одна линия (игровая, восходящая к Виллю) говорит о мартингалах, другая (частотная, восходящая к Мизесу) — о правилах выбора подпоследовательностей. Обе они начинаются с определения Мартин-Лёфа (которое равносильно определению с перечислимыми мартингалами) и определения с немонотонными вычислимыми мартингалами, которое то ли более слабое, то ли эквивалентно (как мы только что говорили). Правая колонка соответствует уменьшающимся классам допустимых правил выбора. Левая колонка соответствует усиливающимся требованиям к выигрышу на неслучайных последовательностей (см. раздел 9.9, теорема 183 и задача 271). Отсюда сразу видно, что соответствующие классы последовательностей возрастают (пока что нестрого) сверху вниз.

Чтобы понять, почему других стрелок (кроме нарисованных на картинке и их следствий, а также, возможно, стрелки с вопросом) нет, посмотрим на связи между колонками. Как мы видели, из случайности по Мизесу – Колмогорову – Лавлэнду не следует даже случайность по Курцу (см. задачу 279), поэтому ни из какого утверждения правой колонки не следует никакое утверждение левой.

Слева направо: из случайности относительно частичных вычислимых мартингалов следует случайность по Мизесу – Чёрчу – Дэли (теорема 190, с. 325), из случайности относительно вычислимых мартингалов следует случайность по Мизесу – Чёрчу (теорема 182, с. 312), а из случайности по Шнорру вытекает усиленный закон больших чисел (задача 92, с. 83). Но более сильные утверждения уже вывести нельзя: из случайности по Курцу не вытекает закон больших чисел (задача 92, с. 83), из случайности по Шнорру не следует случайность по Мизесу – Чёрчу (задача 270, с. 318), из случайности относительно вычислимых мартингалов не следует случайность по Мизесу – Чёрчу – Дэли, поскольку в первом случае возможен $O(\log n)$-рост сложности для начал длины n (теорема 182, с. 312), а во втором — нет (теорема 188, с. 322); наконец, из случайности относительно частичных вычислимых мартингалов не следует случайность по Мизесу – Колмогорову – Лавлэнду, поскольку в первом случае возможна сублинейная сложность начальных отрезков (теорема 190, с. 325), а во втором нет (теорема 193, с. 328).

Отсюда видно, что стрелки в каждой из колонок необратимы (поскольку утверждения в левой колонке имеют различные следствия, а утверждения в правой колонке следуют из разных утверждений), так что действительно в нашей таблице нет пропущенных стрелок.

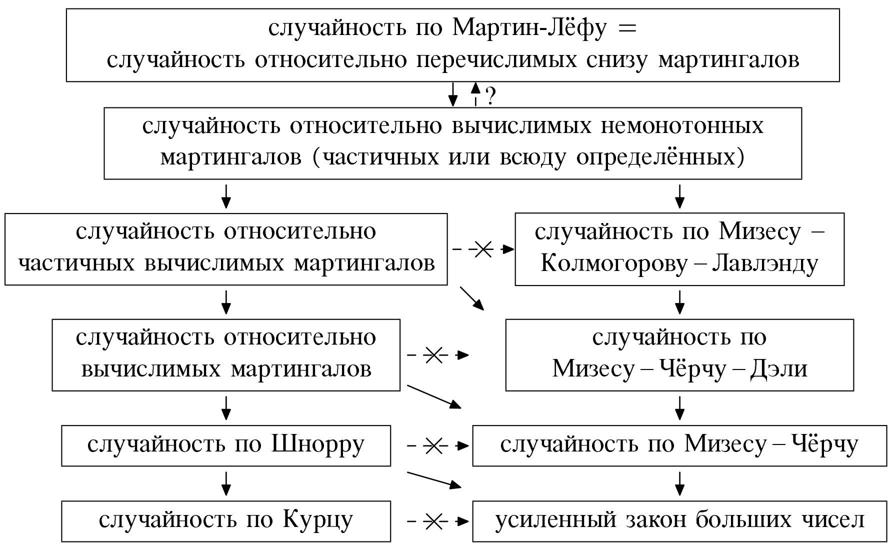


Рис. 9.3. Соотношения между разными определениями случайности.

10. Неравенства для энтропии, сложности и размера

10.1. Постановка задачи и результаты

Первой публикацией Колмогорова, где давалось определение сложности конечного объекта, была статья «Три подхода к определению понятия „количество информации"» [63]. Эти три подхода назывались там комбинаторным, вероятностным и алгоритмическим.

При *алгоритмическом* подходе количество информации в сообщении измеряется его колмогоровской сложностью (как мы говорим теперь; естественно, в оригинальной статье такого названия не было). При *вероятностном* подходе сообщение рассматривается как одно из значений случайной величины, и количество информации в нём определяется как шенноновская энтропия этой величины. Но самым первым упоминался *комбинаторный* подход, основанный на таком тривиальном наблюдении: если имеется N различных сообщений, то нужно предусмотреть $\log N$ битов, чтобы их закодировать. (Вариант: если требуется отгадать один из N объектов, задавая да-нет-вопросы, то нужно $\log N$ вопросов.)

Мы уже приводили некоторые результаты о связи этих трёх подходов. Например, теорема 8 (с. 29) связывает комбинаторный и алгоритмический подходы, будучи уточнением такого (абсурдного при буквальном понимании) утверждения: «слово x имеет сложность меньше n, если и только если оно принадлежит множеству из менее чем 2^n элементов». С другой стороны, результаты раздела 7.3 связывают колмогоровскую сложность и шенноновскую энтропию.

В этой главе мы, следуя [51, 134], хотим установить более формальные связи между тремя подходами к определению количества информации, ограничившись достаточно узким классом утверждений: линейными неравенствами для энтропии и сложности (и соответствующими им комбинаторными утверждениями).

Пусть $x_1, \ldots, x_n$ — двоичные слова. Для каждого непустого множества индексов $I \subset \{1, 2, \ldots, n\}$ рассмотрим набор слов с индексами из I, который мы будем обозначать x_I, и его колмогоровскую сложность. Например, при $n = 3$ имеется 7 таких наборов и, соответственно, 7 сложностей:

$$KS(x_1), KS(x_2), KS(x_3), KS(x_1, x_2), KS(x_1, x_3), KS(x_2, x_3), KS(x_1, x_2, x_3).$$

Несколько примеров линейных неравенств, их связывающих:

- $KS(x_1, x_2) \leqslant KS(x_1) + KS(x_2) + O(\log N)$;
- $KS(x_1, x_2, x_3) \leqslant KS(x_1) + KS(x_2, x_3) + O(\log N)$;

- $KS(x_1, x_2, x_3) + KS(x_1) \leqslant KS(x_1, x_2) + KS(x_1, x_3) + O(\log N)$;
- $2\,KS(x_1, x_2, x_3) \leqslant KS(x_1, x_3) + KS(x_2, x_3) + KS(x_1, x_2) + O(\log N)$

(мы предполагаем, что x_1, x_2, x_3 — слова длины не больше N).

Общий вид линейного неравенства для сложностей:

$$\sum_I \lambda_I\, KS(x_I) \leqslant O(\log N)$$

(суммирование по всем непустым подмножествам множества $\{1, \ldots, n\}$, коэффициенты λ_I могут быть любого знака; предполагается, что все слова имеют длину не больше N).

Нас интересует, какие неравенства такого вида верны, то есть, говоря более формально, при каких наборах действительных чисел λ_I найдётся такое число c, что

$$\sum_I \lambda_I\, KS(x_I) \leqslant c \log N$$

для любого N и для любых слов $x_1, \ldots, x_n$ длины не больше N. Ответ на этот вопрос неизвестен, есть лишь некоторые частичные результаты.

Первый из них (доказанный А. Ромащенко) говорит, что неравенство для сложностей верно тогда и только тогда, когда верно неравенство для шенноновских энтропий с теми же коэффициентами. Оно получается, если вместо слов $x_1, \ldots, x_n$ рассматривать случайные величины $\xi_1, \ldots, \xi_n$ с произвольным совместным распределением:

$$\sum_I \lambda_I H(\xi_I) \leqslant 0.$$

Здесь ξ_I — случайная величина, составленная из величин ξ_i при $i \in I$, или, другими словами, проекция случайного вектора $\langle \xi_1, \ldots, \xi_n \rangle$ на I-координаты.

В одну сторону это сразу же вытекает из результатов раздела 7.3: теорема 147 (с. 256) говорит, что энтропия есть математическое ожидание сложности, и если линейное неравенство верно для сложностей, то оно верно и для их математических ожиданий (с нулём в правой части, поскольку при $N \to \infty$ отношение $O(\log N)/N$ стремится к нулю).

Более подробно. Пусть величина ξ_i принимает значения в множестве X_i. Тогда значение $\xi = \langle \xi_1, \ldots, \xi_n \rangle$ можно изобразить в виде столбца высоты n. Готовясь применить теорему 147, для произвольного N рассмотрим N независимых величин, имеющих то же распределение, что и ξ. Вместе они образуют величину ξ^N, значениями которой являются прямоугольные таблицы шириной N и высотой n. По теореме 147 математическое ожидание сложности такой таблицы есть $NH(\xi) + O(\log N)$ (в этой теореме речь шла о префиксной сложности при условии N, но с точностью до $O(\log N)$ это не играет роли).

Такую таблицу можно также читать по строкам, рассматривая её как набор из n строк, каждая из которых есть слово длины N в соответствующем алфавите. При этом теорему 147 не обязательно применять ко всем строкам: можно оставить лишь строки с номерами из некоторого множества $I \subset \{1, 2, \ldots, n\}$. Математическое ожидание сложности этой части таблицы будет $NH(\xi_I) + O(\log N)$.

Если неравенство
$$\sum_I \lambda_I\, KS(x_I) \leqslant O(\log N)$$
справедливо для любых слов $x_1, \ldots, x_n$, то оно справедливо и для строк нашей таблицы. Поэтому, переходя к математическим ожиданиям, имеем
$$\sum_I \lambda_I NH(\xi_i) \leqslant O(\log N).$$
Левая часть равна
$$\left(\sum_I \lambda_I H(\xi_i)\right) \cdot N,$$
поэтому это возможно лишь при
$$\sum_I \lambda_I H(\xi_i) \leqslant 0,$$
что и требовалось.

Более сложен обратный переход (если неравенство верно для энтропий, то оно верно и для сложностей с логарифмической точностью). Здесь надо, начав с некоторого набора слов $x_1, \ldots, x_n$, построить набор случайных величин $\xi_1, \ldots, \xi_n$ так, чтобы энтропии этих величин и их комбинаций были близки к соответствующим сложностям. Это делается с помощью предложенного А. Ромащенко метода «типизации»: мы рассматриваем множество всех наборов $x'_1, \ldots, x'_n$, которые имеют не бóльшие сложности и условные сложности, чем $x_1, \ldots, x_n$, и затем рассматриваем случайный элемент этого множества. Подробности см. ниже, в разделе 10.6 (теорема 212).

Дальнейшие результаты так или иначе связаны с комбинаторной интерпретацией неравенств. Для начала рассмотрим простейшее неравенство
$$KS(x_1, x_2) \leqslant KS(x_1) + KS(x_2) + O(\log N)$$
и попытаемся понять, что ему соответствует при комбинаторном подходе. Пусть X_1 и X_2 — конечные множества, из которых берутся сообщения x_1 и x_2, а $A \subset \subset X_1 \times X_2$ — множество возможных пар сообщений. Тогда для пары $\langle x_1, x_2 \rangle$ возможно $|A|$ вариантов (здесь $|A|$ — число элементов в множестве A). Для первой компоненты x_1 число возможностей равно числу элементов в первой проекции множества A (в множестве тех x_1, для которых $\langle x_1, x_2 \rangle \in A$ при некотором $x_2 \in X_2$). Обозначая это число $m(1)$, а размер проекции на вторую ось $m(2)$, можно записать комбинаторный аналог рассматриваемого неравенства:
$$\log |A| \leqslant \log m(1) + \log m(2)$$
или, в мультипликативной форме,
$$|A| \leqslant m(1)\, m(2)$$

(размер множества не больше произведения размеров его проекций, что очевидно).

Менее очевидное неравенство получается из другого неравенства для сложностей (теорема 26; аналогичное неравенство для энтропий составляет содержание задачи 230):

$$2\,KS(x_1,x_2,x_3) \leqslant KS(x_1,x_2) + KS(x_1,x_3) + KS(x_2,x_3) + O(\log N).$$

Действуя по аналогии, можно предположить, что для любого подмножества A декартова произведения $X_1 \times X_2 \times X_3$ справедливо неравенство:

$$2\log|A| \leqslant \log m(1,2) + \log m(1,3) + \log m(2,3)$$

(здесь $m(i,j)$ — число элементов в проекции множества A на оси i и j, рис. 10.1). В мультипликативной записи:

$$|A|^2 \leqslant m(1,2)m(1,3)m(2,3).$$

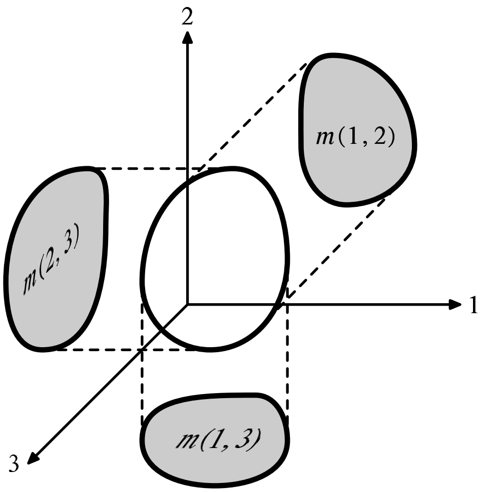


Рис. 10.1. Три проекции.

И действительно, это не только верно, но может быть выведено из неравенства для сложностей с помощью следующего простого рассуждения. Рассмотрим произвольное натуральное N и множество A^N. Изображая элемент $\langle x_1,x_2,x_3\rangle \in A$ в виде столбца высоты 3, мы представляем элемент A^N как таблицу ширины N и высоты 3. Таких таблиц имеется $|A|^N$, и потому среди них есть таблицы сложности не меньше $\log(|A|^N) = N\log|A|$. Но каждую такую таблицу можно считать тройкой строк $\bar{x}_1,\bar{x}_2,\bar{x}_3$ (каждая из трёх строк имеет длину N) и применить неравенство для сложностей:

$$2\,KS(\bar{x}_1,\bar{x}_2,\bar{x}_3) \leqslant KS(\bar{x}_1,\bar{x}_2) + KS(\bar{x}_1,\bar{x}_3) + KS(\bar{x}_2,\bar{x}_3) + O(\log N).$$

Далее можно оценить каждое слагаемое: например, пара $\langle \bar{x}_1, \bar{x}_2 \rangle$, которую можно представить таблицей ширины N и высоты 2, представляет собой набор из N столбцов, каждый из которых принадлежит проекции множества A на первую и вторую координату. Для каждого столбца имеется $m(1,2)$ возможностей, а для всей таблицы $m(1,2)^N$ возможностей, и потому её сложность (при известных N и A) не превосходит $N \log m(1,2) + O(1)$. Множество A не зависит от N, а сложность N есть $O(\log N)$, поэтому в итоге мы получаем

$$2N \log |A| \leqslant N \log m(1,2) + N \log m(1,3) + N \log m(2,3) + O(\log N),$$

что при $N \to \infty$ даёт нам искомое неравенство

$$2 \log |A| \leqslant \log m(1,2) + \log m(1,3) + \log m(2,3).$$

280 Докажите то же неравенство, исходя из неравенства

$$2H(\xi_1, \xi_2, \xi_3) \leqslant H(\xi_1, \xi_2) + H(\xi_1, \xi_3) + H(\xi_2, \xi_3).$$

[Указание: рассмотрите тройку случайных величин, равномерно распределённую в множестве A, и воспользуйтесь тем, что энтропия любой случайной величины с k значениями не превосходит $\log k$.]

281 Докажите то же неравенство непосредственно, без использования сложностей или энтропий. [Указание. Его можно вывести из неравенства теоремы 164.]

Аналогичное рассуждение можно применить к любому линейному неравенству для сложностей, имеющему в левой части (слева от знака $\leqslant$) лишь одно слагаемое с положительным коэффициентом, а в правой части произвольное количество слагаемых с неотрицательными коэффициентами. При этом можно разрешить в правой части не только безусловные сложности, но и условные сложности. Например, неравенству

$$KS(x_1, x_2) \leqslant KS(x_1) + KS(x_2 \mid x_1)$$

соответствует (очевидное) неравенство

$$m(1,2) \leqslant m(1) \cdot m(2 \mid 1),$$

выполненное для произвольного множества $A \subset X_1 \times X_2$, если под $m(1,2)$ понимать число элементов в A, под $m(1)$ понимать число элементов в проекции A на первую координату, а под $m(2 \mid 1)$ понимать максимальный размер сечения множества A, получаемого фиксацией первой координаты.

[Объясним, почему именно это неравенство естественно считать комбинаторным аналогом неравенства для сложностей. «Комбинаторное количество информации» в x_1 есть $\log m(1)$; при фиксированном x_1 у нас имеется не более $m(2 \mid 1)$ возможных сообщений x_2, и потому количество информации в x_2 при известном x_1 (согласно комбинаторному подходу) не превосходит $\log m(2 \mid 1)$. А количество информации в паре $\langle x_1, x_2 \rangle \in A$ считается равным $\log A = \log m(1,2)$.]

$\boxed{\mathbf{282}}$ Покажите, что любому линейному неравенству вида $L \leqslant R$ с неотрицательными коэффициентами, справедливому для сложностей (безусловных и условных), в котором в левой части L стоит только одно слагаемое, соответствует (описанным образом) истинное комбинаторное неравенство.

Неравенства такого вида, в которые не входят условные сложности, можно полностью описать. Рассмотрим неравенство

$$KS(x_1, \ldots, x_n) \leqslant \sum_I \lambda_I \, KS(x_I) + O(\log N),$$

где в правой части все коэффициенты неотрицательны, а сумма берётся по непустым множествам, не совпадающим с полным множеством индексов $\{1, 2, \ldots, n\}$. (Ясно, что только такие неравенства представляют интерес: если в левой части отсутствует некоторое слово, то это слово можно удалить и из правой части: при его замене на пустое слово сложность правой части только уменьшится.)

Теорема 204. *Это неравенство выполнено тогда и только тогда, когда для любого $i \in \{1, 2, \ldots, n\}$ сумма коэффициентов в правой части при членах, содержащих i, не меньше единицы.*

◀ Если сумма коэффициентов при x_i меньше единицы, то неравенство не выполняется даже в случае, когда все остальные слова (кроме x_i) пусты.

С другой стороны, пусть при всех i сумма коэффициентов в правой части не меньше единицы. Разложим каждое слагаемое в сумму, заменив, скажем,

$$KS(x_1, x_2, x_3, \ldots, x_n)$$

на

$$KS(x_1) + KS(x_2 \,|\, x_1) + KS(x_3 \,|\, x_1, x_2) + \ldots + KS(x_n \,|\, x_1, \ldots, x_{n-1}),$$

при этом во всех случаях будем использовать один и тот же порядок (возрастание индексов). Посмотрим отдельно на члены вида $KS(x_i \,|\, \ldots)$ с различными условиями. В левой части в качестве условия используются все предыдущие слова $x_1, \ldots, x_{i-1}$, а в правой части могут быть разные подмножества этого условия, но от уменьшения условия сложность лишь возрастает. Остаётся вспомнить, что по предположению сумма коэффициентов не меньше единицы. ▶

$\boxed{\mathbf{283}}$ Покажите, что для префиксной сложности неравенства, о которых идёт речь в только что доказанной теореме, выполнены с точностью до $O(1)$ (без логарифма длин). [Указание. Приведённое рассуждение показывает, что это неравенство является линейной комбинацией базисных неравенств, которые верны для префиксной сложности с точностью до $O(1)$ (теорема 70, с. 128). В самом деле, если временно переопределить условную сложность, положив $KP(u \,|\, v)$ равным $KP(u, v) - KP(v)$, то неравенство типа $KP(z \,|\, x, y) \leqslant KP(z)$ сводится к базисному неравенству.]

Нам, однако, хочется понять комбинаторный смысл произвольных линейных неравенств для сложностей (энтропий), а не только тех, у которых лишь один член в левой части. Тут мы сталкиваемся с некоторой трудностью.

Посмотрим на базисное неравенство

$$KS(x_1) + KS(x_1, x_2, x_3) \leqslant KS(x_1, x_2) + KS(x_1, x_3) + O(\log N).$$

По аналогии с предыдущим можно было бы предположить, что для произвольного множества $A \subset X_1 \times X_2 \times X_3$ выполнено неравенство

$$m(1) \cdot m(1,2,3) \leqslant m(1,2) \cdot m(1,3).$$

Однако это не так. Это неравенство верно (и обращается в равенство) для любого «параллелепипеда» $a \times b \times c$: в этом случае $m(1) = a$, $m(1,2,3) = abc$, $m(1,2) = ab$ и $m(1,3) = ac$. Но если мы добавим к параллелепипеду $a \times b \times c$ с больши́ми a, b, c ещё и параллелепипед $a' \times 1 \times 1$, взяв a' много больше a, но много меньше ab и ac, то от такого добавления $m(1,2)$, $m(1,3)$ и $m(1,2,3)$ изменятся мало, но $m(1)$ сильно возрастёт, и неравенство нарушится.

Другой пример. Рассмотрим обратное неравенство для сложности пары:

$$KS(x_1) + KS(x_2 \,|\, x_1) \leqslant KS(x_1, x_2) + O(1).$$

Как перевести его на комбинаторный язык? Неравенство

$$m(1) \cdot m(2 \,|\, 1) \leqslant m(1,2)$$

(которое могло бы быть переводом) неверно: $m(1,2)/m(1)$ есть *средний* размер (непустого) сечения, и этот средний размер может быть существенно меньше *максимального*, который мы обозначаем через $m(2\,|\,1)$.

В чём тут дело и каков выход из положения? Есть несколько вариантов. Можно ограничиться некоторыми специальными множествами (однородными или почти однородными), для которых этой проблемы не возникает. Другой вариант состоит в том, чтобы попытаться лучше понять, какое именно комбинаторное утверждение соответствует неравенству. Оба подхода будут рассмотрены ниже. Начнём с первого — однородных множеств.

10.2. Однородные множества

Напомним использованные нами обозначения. Пусть $A \subset X_1 \times \ldots \times X_n$ — некоторое непустое подмножество декартова произведения n конечных множеств $X_1, \ldots$ $\ldots, X_n$. Для каждого множества индексов $I \subset \{1, \ldots, n\}$ можно рассмотреть проекцию A на соответствующие координаты. Она является подмножеством произведения $\prod_{i \in I} X_i$. Число элементов в этой проекции мы будем обозначать $m_A(I)$. Помимо проекций, можно рассматривать и их сечения. Пусть I и J — два непересекающихся множества индексов. Фиксируем произвольным образом I-координаты (выбрав некоторую точку в $\prod_{i \in I} X_i$) и рассмотрим множество всех J-координат точек из A с заданными I-координатами. Таким образом каждой точке множества $\prod_{i \in I} X_i$ соответствует некоторое подмножество произведения $\prod_{j \in J} X_j$. Максимальный размер такого подмножества мы будем обозначать $m_A(J \,|\, I)$. (Если множество A ясно из контекста, мы будем опускать индекс A в этих обозначениях.)

Естественно считать, что $m(\varnothing) = 1$, а также $m(\varnothing \,|\, J) = 1$ при любом J. С другой стороны, $m(I \,|\, \varnothing)$ естественно положить равным $m(I)$.

Пусть, например, имеется некоторое множество $A \subset X_1 \times X_2$ (рис. 10.2).

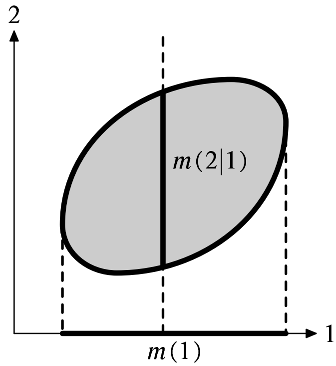


Рис. 10.2. Плоское множество и его характеристики.

Тогда $m_A(\{1\})$ — число элементов в проекции множества A на горизонтальную ось, $m_A(\{2\}$ — число элементов в проекции на вертикальную ось, $m_A(\{2\} \,|\, \{1\})$ — максимальное число элементов в вертикальных сечениях, а $m_A(\{1\} \,|\, \{2\})$ — в горизонтальных. Общее число элементов в множестве есть $m_A(\{1, 2\})$.

Имеет место очевидное неравенство:

$$m(1, 2) \leqslant m(1) \cdot m(2 \,|\, 1)$$

(для краткости мы опускаем индекс A и фигурные скобки в множествах индексов). В самом деле, каждое из $m(1)$ вертикальных сечений содержит не более $m(2 \,|\, 1)$ элементов.

Для n-мерного множества аналогичное неравенство выглядит так:

$$m(1, 2, \ldots, n) \leqslant m(1) \cdot m(2 \,|\, 1) \cdot m(3 \,|\, 1, 2) \cdot \ldots \cdot m(n \,|\, 1, 2, \ldots, n-1).$$

В самом деле, для каждого из $m(1)$ возможных значений первой координаты есть не более $m(2 \,|\, 1)$ значений второй, для каждого из которых есть не более $m(3 \,|\, 1, 2)$ значений третьей и так далее. Порядок координат роли не играет:

$$m(k_1, \ldots, k_n) \leqslant m(k_1) \cdot m(k_2 \,|\, k_1) \cdot m(k_3 \,|\, k_1, k_2) \cdot \ldots \cdot m(k_n \,|\, k_1, \ldots, l_{n-1})$$

для любой перестановки $k_1, k_2, \ldots, k_n$ чисел $1, 2, \ldots, n$ (в левой части так или иначе записывается общее число элементов в множестве A).

Будем называть множество A *однородным*, если все эти неравенства (для любой перестановки $k_1, \ldots, k_n$) обращаются в равенства. Простейший пример однородного множества — «параллелепипед», то есть произведение подмножеств $A_i \subset X_i$. Однако бывают и другие однородные множества. Например, шестиэлементное дву-

мерное множество на рис. 10.3 является однородным (все ненулевые сечения состоят из двух элементов, а проекции на обе оси — из трёх).

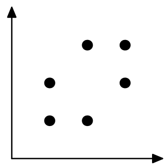


Рис. 10.3. Однородное множество.

Пусть I, J, K — непересекающиеся множества индексов. Тогда для произвольного множества A выполнено неравенство

$$m(J \cup K \mid I) \leqslant m(J \mid I) \cdot m(K \mid I \cup J)$$

(при фиксированных I-координатах есть не более $m(J \mid I)$ возможных значений J-координат, для каждого из которых есть не более $m(K \mid I \cup J)$ значений K-координат).

С помощью этого неравенства можно доказать ранее упоминавшее неравенство

$$m(k_1, \ldots, k_n) \leqslant m(k_1) \cdot m(k_2 \mid k_1) \cdot m(k_3 \mid k_1, k_2) \cdot \ldots \cdot m(k_n \mid k_1, \ldots, l_{n-1}),$$

группируя сомножители в правой части. Например, произведение

$$m(k_3 \mid k_1, k_2) \cdot m(k_4 \mid k_1, k_2, k_3)$$

не меньше

$$m(k_3, k_4 \mid k_1, k_2),$$

после чего произведение

$$m(k_2 \mid k_1) \cdot m(k_3, k_4 \mid k_1, k_2)$$

можно (не увеличивая) заменить на

$$m(k_2, k_3, k_4 \mid k_1)$$

и так далее, пока не получится левая часть. Для однородного множества все эти неравенства обращаются в равенства (так как два крайних члена в цепочке неравенств равны). Из этого рассуждения видно, что для однородных множеств неравенство

$$m(J \cup K \mid I) \leqslant m(J \mid I) \cdot m(K \mid I \cup J)$$

обращается в равенство для любых I, J, K, поскольку можно подобрать цепочку неравенств, в которой оно встречается. Равенство

$$m(J \cup K \mid I) = m(J \mid I) \cdot m(K \mid I \cup J)$$

можно считать определением однородных множеств (требуя его выполнения для любых непересекающихся множеств индексов I, J, K).

$\boxed{\textbf{284}}$ Докажите, что это свойство действительно равносильно однородности.

$\boxed{\textbf{285}}$ Докажите, что проекция однородного множества на любое множество индексов является однородным множеством.

$\boxed{\textbf{286}}$ Докажите, что сечение однородного множества (мы фиксируем одну координату и рассматриваем множество возможных значений остальных координат) является однородным множеством.

Однородные множества важны как источники случайных величин. Пусть $A \subset \subset X_1 \times \ldots \times X_n$ — произвольное множество. Рассмотрим случайную точку в A, принимающую все значения в A с равной вероятностью. Её проекция на i-ю координату есть случайная величина ξ_i со значениями в X_i.

Теорема 205. *Множество A однородно тогда и только тогда, когда для любого $I = \{i_1, \ldots, i_k\}$ величина $\xi_I = \langle \xi_{i_1}, \ldots, \xi_{i_k} \rangle$ принимает все свои значения с равными вероятностями.*

◀ Пусть I — некоторое множество индексов, а J — его дополнение до полного множества индексов $\{1, 2, \ldots, n\}$. Тогда равенство

$$m(1, 2, \ldots, n) = m(I) \cdot m(J \mid I)$$

означает, что средний размер (непустого) сечения, получающегося фиксацией I-координат, то есть $m(1, 2, \ldots, n)/m(I)$, равен его максимальному размеру $m(J \mid I)$, то есть что все сечения одинаковы. А это и значит, что все значения случайной величины ξ_I равновероятны.

Напротив, пусть для некоторого множества A для любого множества индексов I все значения случайной величины ξ_I равновероятны. Взяв $I = \{1, \ldots, n-1\}$, мы получаем, что все (непустые) сечения, получаемые фиксацией первых $n-1$ координат, имеют одинаковый размер, и потому

$$m(1, 2, \ldots, n) = m(n \mid 1, 2, \ldots, n-1) \cdot m(1, 2, \ldots, n-1). \qquad (*)$$

Кроме того, поскольку совместное распределение величин $\xi_1, \xi_2, \ldots, \xi_{n-1}$ равномерно в проекции множества A на координаты $1, 2, \ldots, n-1$, то в этой проекции мы имеем ту же картину: $\xi_1, \ldots, \xi_{n-1}$ суть случайные величины, получающиеся проектированием на разные оси случайной точки некоторого множества, равномерно распределённой в этом множестве. Рассуждая по индукции, можно предполагать, что это множество однородно, и тогда равенство $(*)$ можно продолжить:

$$\begin{aligned} &m(1, 2, \ldots, n) = \\ &= m(n \mid 1, 2, \ldots, n-1) \cdot m(n-1 \mid 1, 2, \ldots, n-2) \cdot \ldots \cdot m(3 \mid 1, 2) \cdot m(2 \mid 1) \cdot m(1). \end{aligned}$$

Поскольку порядок координат мог быть любым, мы заключаем, что множество A является однородным в смысле нашего определения.▶

Следствие. *Для построенных таким образом величин $\xi_1, \ldots, \xi_n$ энтропия набора ξ_I равна $\log m(I)$ (при любом $I \subset \{1, 2, \ldots, n\}$).*

Отсюда мы заключаем, что

Теорема 206. *Каждому неравенству для энтропий случайных величин соответствует неравенство, выполненное для размеров проекций однородных множеств.*

Например, для однородного множества $A \subset X_1 \times X_2 \times X_3$ выполнено неравенство

$$m(1) \cdot m(1, 2, 3) \leqslant m(1, 2) \cdot m(1, 3),$$

соответствующее базисному неравенству для сложностей (теорема 24) и, вообще говоря, неверное для произвольных (не однородных) множеств.

В следующем разделе мы, следуя [27], докажем обратное к теореме 206 утверждение, для чего по данному набору случайных величин построим однородное множество, имеющее соответствующие размеры проекций.

10.3. Построение однородного множества

Пусть имеется некоторый набор случайных величин $\eta_1, \ldots, \eta_n$ с конечным числом значений. Мы хотим построить однородное множество A, для которого размеры проекций соответствуют энтропиям величин η_I. Идеально было бы, чтобы

$$\log m_A(I) = H(\eta_I)$$

для всех $I \subset \{1, \ldots, n\}$. Тогда можно было бы сразу заключить, что неравенство, верное для (логарифмов) размеров проекций однородных множеств, верно и для любых случайных величин.

Однако это, как легко понять, невозможно: энтропия вовсе не обязана быть логарифмом целого числа. Но если нас интересуют линейные неравенства, достаточно, чтобы энтропии были бы *пропорциональны* логарифмам проекций, и даже чтобы они были *приблизительно* (с малой ошибкой) пропорциональны им. Сейчас мы построим однородное множество для случая, когда вероятности всех значений для набора $\langle \eta_1, \ldots, \eta_n \rangle$ рациональны. Ясно, что для неравенств рассматриваемого нами вида этого достаточно: по непрерывности любое неравенство, выполненное для случайных величин с рациональными вероятностями, верно и для любых случайных величин.

Итак, пусть имеется набор случайных величин $\eta_1, \ldots, \eta_n$. Каждое его значение будем изображать столбцом высоты n, в котором сверху вниз записаны значения величин. Каждый столбец имеет некоторую (по предположению рациональную) вероятность. Приведя все эти вероятности к общему знаменателю N, можно составить таблицу из N столбцов, в которой частота (доля) каждого столбца равна его вероятности как значения случайной величины $\langle \eta_1, \ldots, \eta_n \rangle$.

Такую таблицу можно читать и «по строкам», тогда i-я строка есть слово длины N в алфавите, буквами которого являются возможные значения величины η_i. Множество таких слов мы обозначим X_i (длина N у нас фиксирована, поэтому

мы её не указываем в обозначении). А вся таблица есть набор из n слов длины N (каждое в соответствующем алфавите), то есть элемент множества $X_1 \times \ldots \times X_n$.

Рассмотрим теперь всевозможные таблицы, получающиеся из данной перестановкой столбцов. Другими словами, рассмотрим все таблицы ширины N, в которых частоты столбцов такие же, то есть соответствуют распределению вероятности для случайной величины. Получится некоторое подмножество U множества $X_1 \times \ldots \times X_n$. Оно и будет интересующим нас однородным подмножеством.

Чтобы убедиться в этом, заметим, что любой элемент множества U можно получить, применив к исходной таблице перестановку столбцов. Если эту перестановку выбирать случайно, считая все $N!$ перестановок равновероятными, то вероятность получить данный элемент множества U не зависит от выбора этого элемента. (В самом деле, число перестановок, дающих этот элемент, равно числу перестановок, оставляющих его на месте, и определяется лишь количествами равных столбцов в таблице, а не тем, где они расположены.)

Это свойство останется верным, если из таблицы удалить некоторые строки. Следовательно, проекция случайной равномерно распределённой в U точки на любой набор координат также равномерно распределена, так что множество является однородным.

Теперь надо найти размеры проекций. Для начала найдём размер самого множества. Пусть в нём имеется m различных типов столбцов, которые встречаются с частотами $q_1, \ldots, q_m$. Тогда общее число вариантов, которые можно получить перестановками столбцов, равно

$$\frac{N!}{(q_1 N)!\,(q_2 N)! \ldots (q_m N)!},$$

и его логарифм по формуле Стирлинга оценивается как

$$Nh(q_1, \ldots, q_m) + O(\log N),$$

где $h(q_1, \ldots, q_m) = \sum q_i(-\log q_i)$ — шенноновская энтропия величины, принимающей m значений с вероятностями $q_1, \ldots, q_m$, то есть — в нашем случае — величины $\langle \eta_1, \ldots, \eta_n \rangle$. Так что логарифм размера множества примерно в N раз больше энтропии набора случайных величин. То же самое рассуждение применимо и к любой проекции этого множества и показывает, что логарифм размера проекции на множество индексов I примерно в N раз больше энтропии величины η_I.

Если теперь у нас есть линейное неравенство, справедливое для логарифмов проекций однородных множеств, то оно будет справедливо для множества U. Увеличивая N (умножая его на большое целое число) и переходя к пределу, мы видим, что $O(\log N)/N$ стремится к нулю, и заключаем, что то же линейное неравенство верно и для энтропий любых n случайных величин с рациональным распределением вероятностей. По непрерывности это верно и для любого (не обязательно рационального) распределения вероятностей. Приходим к такому утверждению [27]:

Теорема 207 (Чен – Янг). *Всякому неравенству для размеров проекций однородных множеств соответствует неравенство, выполненное для произвольных случайных величин.*

10.4. Однородные множества и орбиты

Возникает вопрос, за счёт чего нам удалось построить однородное множество в предыдущем разделе. Обычно такого рода однородность достигается за счёт какой-либо алгебраической структуры на рассматриваемых объектах. Нетрудно обнаружить такую структуру и в данном случае.

А именно, у нас имеется группа перестановок S_N, которая действует на столбцах таблицы. Однородное множество — это орбита некоторой точки (таблицы с заданными частотами) при этом действии. В общем случае конструкция выглядит следующим образом.

Пусть имеется произвольная конечная группа G и конечные множества $X_1, \ldots$ $\ldots, X_n$. Пусть заданы действия группы G на каждом из X_i. Они вместе задают действие G на множестве $X_1 \times \ldots \times X_n$. Рассмотрим произвольную точку $\langle x_1, \ldots, x_n\rangle \in X_1 \times \ldots \times X_n$ и орбиту U этой точки.

Теорема 208. *Орбита U является однородным подмножеством множества $X_1 \times \ldots \times X_n$.*

◀ Подействуем на точку $x = \langle x_1, \ldots, x_m\rangle$ случайным элементом группы G (считая все элементы равновероятными). Результат будет случайной величиной, значениями которой являются точки из U. При этом все точки равновероятны. В самом деле, элементы группы, переводящие x в заданную точку y, образуют смежный класс при факторизации по стабилизатору точки x (подгруппе, состоящей из тех элементов G, которые оставляют точку x на месте), а все смежные классы имеют один и тот же размер.

Аналогичное утверждение верно и для любой части индексов, поэтому случайный элемент множества U имеет равную вероятность спроектироваться в любую точку проекции, и потому множество U является однородным. ▶

Во что превращаются при этом неравенства для размеров проекций однородных множеств? Размер орбиты U равен отношению размера G и размера стабилизатора точки x. Этот стабилизатор есть пересечение стабилизаторов точек $x_1, \ldots, x_n$ при соответствующих действиях. Аналогичным образом размер проекции на индексы из I равен отношению размера G и размера пересечения стабилизаторов для x_i при $i \in I$. Заметим, что любая подгруппа H группы G может быть стабилизатором некоторого элемента при некотором действии (достаточно рассмотреть действие G на множестве смежных классов по H). Поэтому любое неравенство для размеров проекций однородных множеств превращается в неравенство, справедливое для любых подгрупп произвольной конечной группы.

Например, неравенство $m(1,2) \leqslant m(1) \cdot m(2)$ превращается в неравенство, верное для любых подгрупп H_1 и H_2 произвольной конечной группы G:

$$\frac{|G|}{|H_1 \cap H_2|} \leqslant \frac{|G|}{|H_1|} \cdot \frac{|G|}{|H_2|}$$

или $|H_1 \cap H_2| \geqslant |H_1| \cdot |H_2| / |G|$.

А неравенство $m(1,2,3)^2 \leqslant m(1,2)m(1,3)m(2,3)$ даёт неравенство

$$|H_1 \cap H_2 \cap H_3|^2 \geqslant |H_1 \cap H_2| \cdot |H_1 \cap H_3| \cdot |H_2 \cap H_3| / |G|.$$

Доказательство теоремы 207 показывает, что верно и обратное: всякому неравенству для размеров пересечений подгрупп (в котором они перемножаются в некоторых степенях) соответствует неравенство для случайных величин (поскольку мы можем приближать случайные величины с помощью орбит), а потому и для проекций однородных множеств. Получаем следующий удивительный результат [27]:

Теорема 209 (Чен – Янг). *Всякому линейному неравенству для энтропий случайных величин соответствует неравенство для размеров подгрупп конечной группы и их пересечений, и наоборот.*

10.5. Почти однородные множества

Мы называли множество $A \subset X_1 \times \ldots \times X_n$ однородным, если неравенство

$$m(k_1, \ldots, k_n) \leqslant m(k_1) \cdot m(k_2 \mid k_1) \cdot m(k_3 \mid k_1, k_2) \cdot \ldots \cdot m(k_n \mid k_1, \ldots, k_{n-1})$$

обращается в равенство для любой перестановки $k_1, \ldots, k_n$ чисел $1, \ldots, n$. Будем говорить, что множество c-однородно, если правая часть этого неравенства не более чем в c раз превосходит левую (для любой перестановки).

Таким образом, 1-однородные множества — это однородные множества в смысле старого определения.

Многие свойства однородных множеств переносятся (с «потерей точности») на c-однородные множества.

Теорема 210. *Пусть множество A является c-однородным.*

(а) *Если I, J, K — непересекающиеся множества индексов, то в неравенстве*

$$m(J \cup K \mid I) \leqslant m(J \mid I) \cdot m(K \mid I \cup J)$$

(*верном для любого множества A*) *правая часть превосходит левую не более чем в c раз.*

(б) *Проекция c-однородного множества на любой набор координат c-однородна.*

(в) *Пусть A' — подмножество множества A, составляющее в нём долю не менее ε. Тогда множество A' является c/ε-однородным.*

(г) *Пусть ξ — случайная величина, равномерно распределённая в A. Тогда её проекция ξ_I на любое множество индексов I имеет энтропию не больше* $\log m(I)$ *и не меньше* $\log m(I) - \log c$.

(д) *Пусть I и J — непересекающиеся множества индексов. Тогда $H(\xi_J \mid \xi_I)$ не больше* $\log m(J \mid I)$ *и не меньше* $\log m(J \mid I) - \log c$.

◀ (а) В правой части неравенства

$$m(k_1, \ldots, k_n) \leqslant m(k_1) \cdot m(k_2 \mid k_1) \cdot m(k_3 \mid k_1, k_2) \cdot \ldots \cdot m(k_n \mid k_1, \ldots, k_{n-1})$$

можно группировать сомножители, пользуясь неравенством

$$m(J \mid I) \cdot m(K \mid I \cup J) \geqslant m(J \cup K \mid I).$$

При этом произведение будет постепенно уменьшаться, пока правая часть не превратится в левую. Если исходное неравенство отличалось от равенства не более чем в c раз, то и на каждом шаге уменьшение будет не более чем в c раз, а порядок действий можно выбрать так, чтобы пройти через любую тройку I, J, K.

(б) По предположению

$$m(n\,|\,1,\ldots,n-1)\cdot m(n-1\,|\,1,\ldots,n-2)\cdot\ldots\cdot m(2\,|\,1)\cdot m(1)\leqslant cm(1,\ldots,n)$$

и это неравенство можно продолжить:

$$cm(1,\ldots,n)\leqslant cm(n\,|\,1,\ldots,n-1)\cdot m(1,\ldots,n-1).$$

Сокращая теперь на $m(n\,|\,1,\ldots,n-1)$, получаем условие c-однородности для проекции на координаты $1,2,\ldots,n-1$ (порядок координат может быть любым, а не только $1,2,\ldots,n$, как в нашем примере).

(в) От перехода к подмножеству максимальные размеры сечений увеличиться не могут, а размер всего множества уменьшается не более чем в $1/\varepsilon$ раз.

(г) Случайная величина ξ_I принимает $m(I)$ значений, поэтому её энтропия не превосходит $\log m(I)$. С другой стороны, если $J=\{1,2,\ldots,n\}\setminus I$, то

$$m(I)\cdot m(J\,|\,I)\leqslant cm(1,2,\ldots,n),$$

поэтому каждое значение величины ξ_I имеет вероятность не больше

$$m(J\,|\,I)/m(1,2,\ldots,n)\leqslant c/m(I),$$

и потому её энтропия не меньше $\log m(I)-\log c$.

(д) Сравним равенство

$$H(\xi_1,\ldots,\xi_n)=H(\xi_n\,|\,\xi_1,\ldots,\xi_{n-1})+H(\xi_{n-1}\,|\,\xi_1,\ldots,\xi_{n-2})+\ldots$$
$$\ldots+H(\xi_2\,|\,\xi_1)+H(\xi_1)$$

с неравенством

$$\log m(1,\ldots n)\leqslant\log m(n\,|\,1,\ldots,n-1)+\log m(n-1\,|\,1,\ldots,n-2)+\ldots$$
$$\ldots+\log m(2\,|\,1)+\log m(1).$$

Левые части у них совпадают, поскольку ξ равномерно распределена в множестве A. Каждая энтропия в правой части не превосходит соответствующего логарифма (условная энтропия есть математическое ожидание энтропии при данном значении условия, которая не больше логарифма соответствующего сечения). Кроме того, в силу однородности неравенство отличается от равенства не более чем на $\log c$. Отсюда следует, что каждая энтропия может отличаться от соответствующего логарифма не более чем на $\log c$. Переставляя и группируя члены подходящим образом, мы можем тем же способом доказать неравенство

$$\log m(J\,|\,I)-\log c\leqslant H(\xi_J\,|\,\xi_I)\leqslant\log m(J\,|\,I)$$

для произвольных непересекающихся множеств I, J. ▶

Непосредственное следствие утверждений (г) и (д): если некоторое линейное неравенство выполнено для энтропий, то оно выполнено для логарифмов проекций и сечений c-однородных множеств с ошибкой не более $\lambda \log c$, где λ — сумма абсолютных величин всех коэффициентов неравенства. (Чем больше c, тем менее однородно множество, и тем больше возможная ошибка.)

Это свойство будет использовано нами в следующем разделе, когда мы по набору слов построим почти однородное множество, а затем случайные величины.

10.6. Метод типизации

Следующая теорема по любому набору слов $x_1, \ldots, x_n$ строит почти однородное подмножество A некоторого декартова произведения $X_1 \times X_2 \times \ldots \times X_n$ конечных множеств, для которого $\log m(J \mid I) \approx KS(x_J \mid x_I)$ с логарифмической (от сложности слов) погрешностью. Вот точная формулировка.

Теорема 211 (Ромащенко). *Для всякого n существует число d, при котором верно следующее утверждение: для любого числа $N > 1$ и любого набора слов $x_1, \ldots, x_n$ сложности не более N найдутся конечные множества $X_1, \ldots, X_n$ и N^d-однородное подмножество $A \subset X_1 \times \ldots \times X_n$, для которых*

$$|\log m(J \mid I) - KS(x_J \mid x_I)| \leqslant d \log N$$

для любых непересекающихся множеств $I, J \subset \{1, \ldots, n\}$.

Заметим, что множества X_i в этой теореме не очень по существу: можно говорить о конечном множестве n-ок произвольной природы с нужными размерами проекций. Отметим ещё, что странное условие $N > 1$ объясняется тем, что при $N = 1$ оценка N^c не растёт с ростом c.

В доказательстве используется понятие *сложностного вектора* набора слов. А именно, сложностным вектором слов $x_1, \ldots, x_n$ называется список всех сложностей $KS(x_I \mid x_J)$ для всех пар (I, J) непересекающихся подмножеств множества индексов $\{1, \ldots, n\}$. (Заметим, что этот вектор имеет экспоненциальное по n число компонент: одних безусловных сложностей (при $J = \varnothing$) уже набирается $2^n - 1$, не считая пустого множества.) Будем обозначать сложностной вектор $\varkappa(x_1, \ldots, x_n)$.

◀ Для каждого набора слов $x_1, \ldots, x_n$ рассмотрим множество $A(x_1, \ldots, x_n)$ всех наборов $y_1, \ldots, y_n$, у которых

$$\varkappa(y_1, \ldots, y_n) \leqslant \varkappa(x_1, \ldots, x_n)$$

при покомпонентном сравнении векторов. При $n = 1$ это будет множество всех слов не большей сложности, чем x_1. При $n = 2$ мы рассматриваем все пары слов y_1, y_2, для которых

$$KS(y_1) \leqslant K(x_1), \quad KS(y_2) \leqslant K(x_2),$$
$$KS(y_1, y_2) \leqslant K(x_1, x_2), \quad KS(y_1 \mid y_2) \leqslant K(x_1 \mid x_2), \quad KS(y_2 \mid y_1) \leqslant K(x_2 \mid x_1).$$

Множество $A(x_1, \ldots, x_n)$ заведомо непусто: оно содержит набор $\langle x_1, \ldots, x_n \rangle$. Мы хотим показать, что оно содержит достаточно много (примерно $2^{KS(x_1, \ldots, x_n)}$) эле-

ментов — больше оно содержать и не может, поскольку все его элементы по построению имеют сложность не больше $KS(x_1, \ldots, x_n)$.

В самом деле, зная сложностной вектор $\varkappa(x_1, \ldots, x_n)$, мы можем перечислять множество $A(x_1, \ldots, x_n)$. Задание сложностного вектора требует $O(\log N)$ битов (заметим, что число компонент в этом векторе, хотя и экспоненциально большое, зависит лишь от n и потому может считаться постоянным). Поэтому любой элемент $A(x_1, \ldots, x_n)$ можно задать, указав (помимо сложностного вектора) его порядковый номер в перечислении, и сложность любого элемента в $A(x_1, \ldots, x_n)$ не превосходит

$$\log|A(x_1, \ldots, x_n)| + O(\log N).$$

В частности, это относится и к исходному набору $\langle x_1, \ldots, x_n\rangle$, откуда и следует требуемая оценка.

Покажем, что множество $A(x_1, \ldots, x_n)$ является c-однородным для некоторой константы c, полиномиально зависящей от N. Для этого надо сравнить обе части неравенства

$$m(1, 2, \ldots, n) \leqslant m(1) \cdot m(2\,|\,1) \cdot m(3\,|\,1, 2) \cdot \ldots \cdot m(n\,|\,1, 2, \ldots, n-1).$$

Логарифмы сомножителей в правой части можно оценить сверху соответствующими сложностями: $m(1) \leqslant 2^{KS(x_1)}$, поскольку по построению $KS(y_1) \leqslant KS(x_1)$ для любого набора $\langle y_1, \ldots, y_n\rangle \in A(x_1, \ldots, x_n)$ (формально говоря, следовало бы написать $2^{KS(x_1)+1}$ в оценке, но с нашей логарифмической точностью это не имеет значения). Аналогично $m(2\,|\,1) \leqslant 2^{KS(x_2\,|\,x_1)}$ и так далее. В итоге получается, что логарифм правой части не превосходит

$$KS(x_1) + KS(x_2\,|\,x_1) + KS(x_3\,|\,x_1, x_2) + \ldots + KS(x_n\,|\,x_1, \ldots, x_{n-1}) + O(1),$$

что равно $KS(x_1, \ldots, x_n) + O(\log N)$. Но и для левой части мы знаем, что её логарифм не меньше $KS(x_1, \ldots, x_n) - O(\log N)$, так что разница между логарифмами есть $O(\log N)$, а отношение самих значений ограничено многочленом от N, что и требовалось. Одновременно мы устанавливаем, что

$$KS(x_i\,|\,x_1, \ldots, x_{i-1}) = \log m(i\,|\,1, \ldots, i-1) + O(\log N),$$

а аналогичное рассуждение с группировкой слагаемых даёт

$$KS(x_J\,|\,x_I) = \log m(J\,|\,I) + O(\log N)$$

для любых непересекающихся I и J. ▶

Использованный при доказательстве приём можно назвать *типизацией*: от одного слова мы переходим к целому множеству, в котором это слово является «типичным представителем» (с точки зрения сложностей и мощностей).

Теперь уже легко завершить доказательство обещанной теоремы:

Теорема 212 (Ромащенко). *Всякое линейное неравенство*

$$\sum_I \lambda_I H(\xi_I) \leqslant 0,$$

выполненное для произвольных случайных величин $\xi_1, \ldots, \xi_n$, выполнено для колмогоровских сложностей произвольных слов сложности не более N с погрешностью $O(\log N)$:

$$\sum_I \lambda_I K(\xi_I) \leqslant O(\log N).$$

Заметим, что константа в $O(\log N)$ зависит от n (и быстро растёт с ростом n), но не от слов $x_1, \ldots, x_n$.

◀ В одну сторону (если неравенство верно для сложностей, то оно же верно и для энтропий) мы это уже доказали в разделе 10.1.

Для обратного перехода у нас тоже всё готово. Пусть неравенство верно для энтропий. Рассмотрим произвольные слова $x_1, \ldots, x_n$ и множество $A = A(x_1, \ldots, x_n)$ из только что доказанной теоремы. Рассмотрим случайную величину $\langle \xi_1, \ldots, \xi_n \rangle$, равномерно распределённую в множестве A. Поскольку множество A является N^c-однородным, то энтропии будут отличаться от логарифмов размеров сечений не более чем на $O(\log N)$ по теореме 210. С другой стороны, величины $\log m(I\,|\,J)$ совпадают с соответствующими сложностями с точностью до $O(\log N)$ по теореме 211, что и завершает доказательство. ▶

Неравенства рассмотренного нами вида можно являются простейшими универсальными формулами некоторого формального языка. Возникает естественный вопрос — сохранится ли теорема Ромащенко, если от универсальных формул перейти к $\forall\exists$-формулам того же языка? Оказывается, что нет, более подробно об этом можно прочитать в [122].

10.7. Комбинаторная интерпретация: примеры

Вернёмся теперь к комбинаторной интерпретации. Напомним основную идею: «слово x имеет сложность меньше n» надо понимать как «слово x принадлежит множеству из менее чем 2^n элементов». (Поскольку сложность определена с точностью до $O(1)$, мы не будем аккуратно различать строгие и нестрогие неравенства.)

Строго говоря, таким образом мы получаем комбинаторную интерпретацию не для функции KS, а для двухместного предиката $KS(x) < n$ с аргументами x и n, и все утверждения о сложности (в частности, неравенства) надо предварительно записать в этих терминах.

Разберём несколько примеров.

- Неравенство $KS(x) \leqslant KS(y)$ записывается так: для всех n из $KS(y) < n$ следует $KS(x) < n$.
- Неравенство $KS(x) \leqslant 2\,KS(y)$ записывается так: для всех n из $KS(y) < n$ следует $KS(x) < 2n$.
- Неравенство $KS(z) \leqslant KS(x) + KS(y)$ записывается так: для всех u и v из $KS(x) < u$ и $KS(y) < v$ следует $KS(z) < u + v$.

Используя последний пример в качестве образца, попытаемся перевести на комбинаторный язык неравенство для сложности пары: $KS(x,y) \leqslant KS(x) + KS(y)$.

Первый шаг: если $KS(x) < u$ и $KS(y) < v$, то $KS(x, y) < u + v$. Далее естественно переводить так: если x принадлежит заданному множеству из менее чем 2^u элементов, а y принадлежит заданному множеству из менее чем 2^v элементов, то можно указать множество из менее чем 2^{u+v} элементов, которому принадлежит пара $\langle x, y\rangle$ (а именно, произведение множеств). Так что ничего особенно нового не получается.

Аналогичный перевод утверждения $KS(y\,|\,x) < v$ выглядел бы так: пара $\langle x, y\rangle$ принадлежит известному множеству, у которого все сечения (при фиксированном x) имеют размер менее 2^v. Так что и неравенство $KS(x, y) \leqslant KS(x) + KS(y\,|\,x)$ тоже ясно, как переводить.

Ситуация существенно меняется, если обратить знак неравенства. Неравенство $KS(z) \geqslant KS(x) + KS(y)$ можно переписать так: если $KS(x) \geqslant u$ и $KS(y) \geqslant v$, то $KS(z) \geqslant u + v$. Но в нашем предикате KS *меньше* границы, а не больше (и эта асимметрия существенна: мы можем перечислять слова малой сложности, но не слова большой сложности!) Поэтому надо перейти к отрицаниям: если неверно, что $KS(x) < u$ и неверно, что $KS(y) < v$, то неверно, что $KS(z) < u + v$. Другими словами: если $KS(z) < u + v$, то $KS(x) < u$ *или* $KS(y) < v$.

Попробуем применить тот же метод перевода для неравенства

$$KS(x, y) \geqslant KS(x) + KS(y\,|\,x),$$

с которым у нас были трудности. Получаем вот что: если пара $\langle x, y\rangle$ принадлежит заданному множеству из менее чем 2^{u+v} элементов, то либо x принадлежит какому-то множеству из менее чем 2^u элементов, либо $\langle x, y\rangle$ принадлежит какому-то множеству, у которого любое сечение содержит не более 2^v элементов.

Более точно: для любых u, v и всякого множества пар A из менее чем 2^{u+v} элементов можно указать

- множество B из менее чем 2^u элементов;
- множество C пар, которое содержит менее 2^v элементов с одним и тем же первым членом,

причём так, что для всякого $\langle x, y\rangle \in A$ выполнено по крайней мере одно из двух: либо $x \in B$, либо $\langle x, y\rangle \in C$.

Если теперь вспомнить доказательство теоремы о сложности пары, то там как раз по существу использовались такие множества: для данных x, y мы смотрели, много ли пар с тем же x имеют малую сложность. Если немного, то $KS(y\,|\,x)$ оказывалось малым (в наших теперешних обозначениях: такие пары входят в C), а если много, то тогда $KS(x)$ мало, поскольку это бывает для немногих x. (В наших теперешних обозначениях такие x попадают в B.)

287 Повторите это рассуждение и формально докажите сформулированное комбинаторное утверждение.

Теперь рассмотрим неравенство, в котором и в левой, и в правой части несколько слагаемых:

$$KS(x_1) + KS(x_1, x_2, x_3) \leqslant KS(x_1, x_2) + KS(x_2, x_3)$$

(базисное неравенство, выполненное с точностью до $O(\log N)$ для слов сложности не выше N).

Переходя к двухместному предикату $KS(x) < n$, это неравенство можно переписать так:

> если $KS(x_1, x_2) < a$, $KS(x_1, x_3) < b$ и $a + b = p + q$, то выполнено по крайней мере одно из неравенств $KS(x_1) < p$ и $KS(x_1, x_2, x_3) < q$.

Естественно предположить, что соответствующее комбинаторное утверждение выглядит так:

> если $A \subset X_1 \times X_2 \times X_3$, $m_A(1,2) \leqslant 2^a$, $m_A(1,3) \leqslant 2^b$ и $a + b = p + q$, то существуют такие $B, C \subset X_1 \times X_2 \times X_3$, что $A \subset B \cup C$, $m_B(1) \leqslant 2^p$ и $m_C(1,2,3) \leqslant 2^q$.

Или, переходя к мультипликативной записи и исключая лишние переменные:

> если $A \subset X_1 \times X_2 \times X_3$ и $m_A(1,2) \cdot m_A(1,3) = l \cdot V$ для некоторых чисел $l, V > 0$, то множество A можно покрыть множествами B и C, для которых $m_B(1) \leqslant l$ и $m_C(1,2,3) \leqslant V$.

Геометрически это можно прочесть так: если у множества A проекции на плоскости $1,2$ и $1,3$ малы, то это ещё не мешает его длине в направлении 1 (проекции на первую координату) и объёму (общему числу точек в A) быть большими. Однако A можно разбить на две части B и C, у первой из которых мала длина по первой координате, а у второго мал объём.

(Вспоминая пример с двумя параллелепипедами, мы видим, что в том случае толстый параллелепипед мог бы составить B, а тонкий — C.)

Всё это не более чем аналогии, которые не заменяют доказательств. Но последнее утверждение действительно оказывается верным, хотя и не вполне тривиальным. Вот как его можно доказать.

Рассмотрим проекцию множества A на плоскость $1,2$; она является некоторым подмножеством $X_1 \times X_2$, которое мы обозначим A_{12}. Для каждого $x \in X_1$ рассмотрим сечение этой проекции; пусть это сечение содержит $n_2(x)$ элементов. Тогда

$$m(1,2) = |A_{12}| = \sum_x n_2(x)$$

(размер множества равен сумме всех его сечений). Аналогичным образом

$$m(1,3) = |A_{13}| = \sum_x n_3(x).$$

Длина $m_A(1)$ есть число ненулевых слагаемых в этих суммах, а $m(1,2,3)$ (общее число элементов в A) можно оценить сверху:

$$m(1,2,3) = |A| \leqslant \sum_x n_2(x) n_3(x).$$

(Без ограничения общности можно считать, что это неравенство обращается в равенство, добавив в A недостающие элементы без изменения его проекций.)

Нам надо разбить A на части B и C, при этом у нас есть ограничения на длину B (в направлении первой координаты) и объём C. Естественно, что при данной длине B нужно забрать в это множество как можно больше точек, поэтому мы включим в B самые большие сечения (у которых $n_2(x)n_3(x)$ максимально) в количестве l штук, а остальное отправим в C. После этого надо доказать лишь, что число элементов в C не больше $|A_{12}|\cdot|A_{13}|/l$.

Как это сделать? Для множества C мы можем оценить размеры двух его плоских проекций; они не превосходят $|A_{12}|$ и $|A_{13}|$. Мы знаем также, что все сечения множества C (при любом $x \in X_1$) по площади не превосходят S_l, где S_l — l-е по величине (в порядке убывания) сечение множества A. Рассмотрим неравенство

$$2\,KS(x_1,x_2,x_3) \leqslant KS(x_1,x_2) + KS(x_1,x_3) + KS(x_2,x_3\,|\,x_1)$$

(которое легко доказать, перейдя к условным сложностям относительно x_1). Оно содержит в левой части только один член, поэтому мы уже знаем, что из него вытекает комбинаторное утверждение

$$m(1,2,3)^2 \leqslant m(1,2)\cdot m(1,3)\cdot m(2,3\,|\,1)$$

и потому

$$|C|^2 \leqslant |A_{12}|\cdot|A_{13}|\cdot S_l.$$

Осталось, таким образом, доказать, что

$$S_l \leqslant \frac{|A_{12}|\cdot|A_{13}|}{l^2}.$$

Вспомним, что множество B состоит из l прямоугольников, каждый из которых имеет площадь не меньше S_l. Сумма «оснований» $n_2(x)$ этих прямоугольников не превосходит $|A_{12}|$, а сумма их «высот» $n_3(x)$ не превосходит $|A_{13}|$, так что среднее (арифметическое) основание не больше $|A_{12}|/l$, а средняя (арифметическая) высота не больше $|A_{13}|/l$. Остаётся заметить, что если $S \leqslant a_i b_i$ при всех $i = 1,2,\ldots,l$, то

$$S \leqslant \frac{a_1+\ldots+a_l}{l}\cdot\frac{b_1+\ldots+b_l}{l}$$

(что легко следует из выпуклости логарифма или неравенства о среднем арифметическом и геометрическом).

Таким образом, в нашем случае сформулированное по аналогии с базисным неравенством утверждение о размерах множеств действительно оказалось верным.

10.8. Комбинаторная интерпретация: общий случай

Переходя от примеров к общему утверждению, рассмотрим произвольное линейное неравенство для сложностей и разделим его на две части с положительными коэффициентами

$$\sum \lambda_I\,KS(x_I) \leqslant \sum \mu_J\,KS(x_J)$$

(суммы в левой и правой части берутся по непересекающимся множествам индексов $I, J \subset \{1, 2, \ldots, n\}$; никакое множество индексов не входит одновременно в обе части; коэффициенты λ_I и μ_J положительны).

Как выглядит соответствующее комбинаторное утверждение? По аналогии с рассмотренными примерами его можно сформулировать так:

Пусть $A \subset X_1 \times \ldots \times X_n$ и пусть n_I — произвольный набор чисел, для которого
$$\prod_I (n_I)^{\lambda_I} = \prod_J m_A(J)^{\mu_J}.$$

Тогда множество A можно покрыть множествами B_I, для которых
$$m_{B_I}(I) \leqslant n_I.$$

К сожалению, про сформулированное таким образом утверждение не удаётся доказать, что оно выполнено одновременно с соответствующим неравенством для сложностей. Нам придётся несколько ослабить утверждение, включив в него множитель, соответствующий $O(\log N)$ в неравенстве. Вот этот ослабленный вариант:

Существует такая константа d, что для произвольных $X_1, \ldots, X_n$, для произвольного $A \subset X_1 \times \ldots \times X_n$ и для произвольного набора чисел n_I, для которого
$$\prod_I (n_I)^{\lambda_I} = \prod_J m_A(J)^{\mu_J},$$
существует покрытие множества A множествами B_I, для которых
$$m_{B_I}(I) \leqslant n_I \cdot (\log |A|)^d.$$

Теорема 213. *Сформулированное утверждение верно для данных наборов λ_I и μ_J тогда и только тогда, когда*
$$\sum \lambda_I \, KS(x_I) \leqslant \sum \mu_J \, KS(x_J) + O(\log N)$$
для любых слов $x_1, \ldots, x_n$ сложности не больше N.

◄ Предположим, что неравенство верно и покажем, как можно покрыть произвольное множество A частями требуемого размера. Прежде всего заметим, что без ограничения общности можно считать элементы множества A наборами $\langle x_1, \ldots, x_n \rangle$ слов длины не больше $N = \log |A|$ (слов хватит, а конкретная природа элементов роли не играет).

Предположим временно, что множество A является простым (имеет сложность $O(\log N)$). В этом случае простыми являются и все его проекции, и потому сложности их элементов не превосходят логарифма размера (с точностью до $O(\log N)$). Поэтому для произвольного элемента $\langle x_1, \ldots, x_n \rangle \in A$ выполнены неравенства
$$K(x_J) \leqslant \log m_A(J) + O(\log N)$$

для всех J; складывая их с коэффициентами μ_J, получаем, что

$$\sum_J \mu_J K(x_J) \leqslant \log\Bigl(\prod_J m_A(J)^{\mu_J}\Bigr) + O(\log N).$$

Пользуясь неравенством (которое мы считаем верным) и вспоминая свойство чисел n_I, заключаем, что

$$\sum_I \lambda_I K(x_I) \leqslant \log\Bigl(\prod_I n_I^{\lambda_I}\Bigr) + O(\log N) = \sum_I \lambda_I \log n_I + O(\log N)$$

для любого элемента $\langle x_1, \ldots, x_n\rangle \in A$. А отсюда следует, что (для каждого элемента множества A) хотя бы одно слагаемое в левой части меньше соответствующего слагаемого в правой: для всякого $\langle x_1, \ldots, x_n\rangle$ найдётся I, при котором

$$KS(x_I) \leqslant n_I + O(\log N),$$

то есть x_I принадлежит множеству всех объектов сложности меньше $n_I + O(\log N)$. Соответствующие элементы A и образуют искомое множество B_I.

Это завершает рассуждение для случая простого множества A. Общий случай сводится к этому с помощью стандартного приёма: для каждого N рассмотрим всевозможные множества $A \subset X_1 \times \ldots \times X_n$, где все X_i состоят из слов длины не больше N, и рассмотрим то из них, которое (при данных λ_I и μ_J) покрывается хуже всего (с наибольшим частным левой и правой частей для оптимального покрытия). Это множество является простым (сложности $O(\log N)$) поскольку его можно найти перебором, для которого достаточно знать N и коэффициенты λ_I, μ_J, и поэтому к нему применимо изложенное выше рассуждение. Поскольку это множество было самым трудным для покрытия, для остальных множеств утверждение тем более верно.

(Сказанное требует некоторого уточнения. Вообще говоря, мы не предполагаем, что коэффициенты λ_I и μ_J являются рациональными или даже вычислимыми. Но их достаточно рассматривать с точностью до $1/N$, поскольку логарифмы всех мощностей не превосходят N и общая ошибка не больше $O(1)$. А задание их с такой точностью требует $O(\log N)$ битов.)

В одну сторону утверждение теоремы доказано. Осталось показать, что если верно утверждение о покрытии, то верно и неравенство для сложностей. Это делается с помощью того же метода типизации.

Рассмотрим произвольный набор слов $x_1, \ldots, x_n$, каждое из которых имеет сложность не более N. Пусть для него неравенство не выполняется и левая часть заметно (более чем на $O(\log N)$) больше правой. Включим наш набор в почти однородное множество $A = A(x_1, \ldots, x_n)$. Рассмотрим в качестве n_I величины $KS(x_I)$, уменьшенные поровну настолько, чтобы сравнять левую часть неравенства с правой; поскольку $\log m_A(J)$ не превосходит $KS(x_J)$, можно применить комбинаторное утверждение. В результате множество A будет покрыто множествами B_I, у которых I-проекции содержат не более $2^{n_I+O(\log N)}$ элементов, что заметно меньше соответствующей проекции множества A, содержащей примерно $2^{KS(x_I)}$ элементов (вспомним, что мы заметно уменьшали n_I). В силу почти однородности множества A

всякая его часть, составляющая небольшую долю в одной из проекций, составляет небольшую долю по отношению ко всему A, и потому фиксированное число множеств B_I не может покрыть множества A. ►

288 Проведите аккуратно все требуемые в этом доказательстве оценки (скрытые за словами «заметно», «небольшая доля» и т. п.)

289 Покажите, что теорема 213 и её доказательство переносятся на неравенства, содержащие не только безусловные, но и условные сложности.

10.9. Комбинаторная интерпретация: другой вариант

Проведённое рассуждение подсказывает другой вариант комбинаторной интерпретации неравенств — менее естественный, но проще формулируемый, поскольку не требуется разбивать неравенство на две части с положительными коэффициентами и по-разному их рассматривать.

А именно, рассмотрим набор чисел λ_I любого знака и следующее комбинаторное утверждение:

> Существует такая константа d, что для произвольных конечных множеств $X_1, \ldots, X_n$ и для произвольного $A \subset X_1 \times \ldots \times X_n$ можно представить A в виде объединения не более чем $(\log|A|)^d$ множеств, для каждого из которых выполнено неравенство
>
> $$\prod m(I)^{\lambda_I} \leqslant (\log|A|)^d$$
>
> для размеров его проекций. (Части могут пересекаться.)

Теорема 214. *Сформулированное утверждение верно для некоторого набора коэффициентов λ_I тогда и только тогда, когда для любого N*

$$\sum \lambda_I\, KS(x_I) \leqslant O(\log N)$$

для любых слов $x_1, \ldots, x_n$ сложности не более N.

◄ Пусть верно комбинаторное утверждение, и мы хотим доказать неравенство для сложностей. Рассмотрим произвольные слова x_i сложности не больше N. Применим метод типизации и получим множество $A = A(x_1, \ldots, x_n)$. При этом $\log|A|$ будет ограничен многочленом от N.

По предположению, множество A можно представить в виде объединения полиномиального (от $\log|A|$, а потому и от N) числа множеств. Рассмотрим то из них, которое содержит больше всего элементов. Обозначим его B. Множество B включает в себя полиномиальную долю элементов множества A, которое является c-однородным с полиномиальным значением c, а потому и B является таковым (с бóльшим, но по-прежнему полиномиальным значением c), причём логарифмы размеров проекций у A и B отличаются на $O(\log N)$. Следовательно, из неравенства для B можно вывести неравенство для A (с точностью $O(\log N)$), а потому, как мы видели, и для сложностей.

В другую сторону доказываемая теорема вытекает из такой леммы:

Лемма. *Любое множество $A \subset X_1 \times \ldots \times X_n$ можно представить в виде объединения полиномиального от $N = \log|A|$ числа частей, каждая из которых является c-однородным множеством с полиномиальным (от N) значением c. (Частям разрешается пересекаться.)*

Заметим, что в этой лемме ничего не говорится ни о колмогоровской сложности, ни о неравенствах. Но из неё следует требуемое нам утверждение. В самом деле, как мы доказали, исходное неравенство верно и для шенноновских энтропий проекций любой n-мерной случайной величины. В частности, оно верно и для случайной величины, равномерно распределённой в любой из частей множества A. Поскольку части однородны, для любой из них логарифмы размеров её проекций отличаются от шенноновской энтропии проекций случайной величины не более чем на $O(\log c) = O(\log N)$.

Докажем теперь сформулированную лемму. Интересно, что (пожалуй, самое) простое доказательство этой леммы использует колмогоровскую сложность. Без ограничения общности можно предполагать, что элементы $\langle x_1, \ldots, x_n\rangle \in A$ состоят из двоичных слов. Для каждого элемента $x \in A$ рассмотрим множество всех $\langle y_1, \ldots, y_n\rangle \in A$, для которых сложностной вектор, составленный из условных (относительно A) сложностей, не превосходит такого же вектора для x.

Эта конструкция отличается от ранее рассмотренной «типизации» в двух отношениях: во-первых, мы рассматриваем лишь элементы $y \in A$ (а раньше никакого A не было); во-вторых, мы используем сложности при условии A.

Заметим, что полученное множество определяется сложностями слов $x_1, \ldots, x_n$ и их комбинаций, а не самими этими словами, и потому получится лишь полиномиальное число таких множеств. Осталось проверить, что каждое из них является c-однородным для полиномиального (от N) значения c.

Это делается так же, как и раньше: число элементов в таком множестве не сильно меньше $2^{KS(x_1,\ldots,x_n\,|\,A)}$, а логарифмы сечений оцениваются условными сложностями, так что эти два изменения оставляют доказательство однородности в силе. ▶

Интересно найти чисто комбинаторное доказательство только что рассмотренной леммы, не использующее колмогоровскую сложность. Это не вполне тривиально даже для двумерных множеств. Пусть мы имеем некоторое конечное множество $A \subset \mathbb{N}^2$. «Почти однородность» в этом случае означает, что

$$m(1,2) \approx m(1)m(2\,|\,1), \qquad m(1,2) \approx m(2)m(1\,|\,2).$$

Другими словами, среднее (непустое) вертикальное сечение $m(1,2)/m(1)$ должно не слишком сильно отличаться от максимального сечения $m(2\,|\,1)$ — и то же самое для горизонтальных сечений.

Естественная идея — разбить множество на части в зависимости от вертикальных сечений, при этом в каждой части размеры сечений отличаются друг от друга не более чем (скажем) вдвое. В этом случае для каждой части максимальное и среднее сечение тоже будет отличаться не более чем вдвое. Проблема в том, что это нужно сделать не только для вертикальных сечений, но и для горизонтальных — после чего вертикальные сечения могут снова испортиться.

Как преодолеть эту трудность? Во-первых, можно заметить, что достаточно найти почти однородное множество, составляющее не очень малую (полиномиальную) долю в исходном множестве. В самом деле, после этого можно повторить то же рассуждение с остатком множества, получив вторую почти однородную часть (непересекающуюся) и так далее: если каждая очередная часть забирает ε-долю элементов, то после $1/\varepsilon$ шагов число оставшихся элементов уменьшится примерно в e раз. Повторяя это полиномиальное число раз, можно добиться, чтобы осталось менее одного элемента (то есть, чтобы вообще ничего не осталось).

Как же построить не очень малую часть, почти однородную в обоих направлениях? После деления на части по вертикали выберем наибольшую часть (остальные не используются), поделим её по горизонтали и снова выберем наибольшую часть (выбросив остальные). Нужно только заметить, что она останется почти однородной по вертикали (по аналогии с пунктом (в) теоремы 210).

Это рассуждение проходит для любой размерности и одно преимущество по сравнению с предыдущим (использующим колмогоровскую сложность). Оно даёт части, неоднородность которых есть полином от логарифма мощности не всего множества, а этой части (более сильное условие).

Ещё немного усложнив рассуждение, можно достичь ещё одного улучшения: неоднородность каждой части ограничена константой (зависящей от размерности задачи n, но не от размера разбиваемого на части множества). Это сделано в работе [3] с помощью следующего приёма.

Для каждого разбиения на части определим его *вес* таким образом, что разбиение наименьшего веса (которое существует, поскольку число разбиений конечно) будет удовлетворять всем нужным условиям.

Вес разбиения будет суммой весов всех элементов в этом разбиении. А вес элемента x, попавшего в часть X, определяется формулой

$$\sum_{A,B} \log m_X(B \mid A) - d \log |X|,$$

где сумма берётся по парам непересекающихся множеств $A, B \subset \{1, 2, \ldots, n\}$, а d — некоторый постоянный коэффициент (заметим, кстати, что в сумме также есть член $\log|X|$ — при $A = \varnothing$, $B = \{1, 2, \ldots, n\}$). Подчеркнём, что веса всех элементов внутри одной части одинаковы.

Покажем, что если выбрать d достаточно большим, то разбиение наименьшего веса содержит немного частей. А именно, мы покажем, что выгодно соединить две части, у которых все параметры $\log m_X(B \mid A)$ достаточно близки (отличаются не более чем на единицу). В самом деле, при соединении каждое $m_X(B \mid A)$ возрастёт не более чем втрое (по сравнению с обеими частями), а $|X|$ возрастёт не менее чем в полтора раза (также по сравнению с обеими частями). При достаточно большом d второе перевесит первое, и при слиянии частей веса всех элементов уменьшатся. Заметим, что нужное значение d определяется лишь числом слагаемых (которое, в свою очередь, зависит лишь от n и не зависит от размера множества).

Пусть d выбрано таким образом. Будем классифицировать части по целым частям значений $\log m_X(B \mid A)$ при всех A и B. Как мы видели, никакие две части не попадут в один класс, а число классов ограничено полиномом от максимального

значения $\log|X|$, которое не превосходит логарифма числа элементов в разбиваемом множестве. Таким образом мы получаем желаемую оценку на число частей.

Осталось доказать, что в разбиении наименьшего веса все части почти однородны. Другими словами, надо показать, что достаточно неоднородная часть может быть разбита на две таким образом, что вес разбиения уменьшится. При разбиении части на две все остальные части и веса входящих в них точек никак не затрагиваются, поэтому можно изучать лишь изменение веса в пределах одной части. В формуле

$$\sum_{A,B} \log m_X(B\,|\,A) - d\log|X|$$

для веса точки все члены (и слагаемые, и вычитаемое) уменьшаются. Нам надо разбить неоднородную часть на две так, чтобы уменьшение слагаемых (для обеих частей) пересилило бы уменьшение вычитаемого. Последнее легко подсчитать: если часть из m элементов разбивается на две части из pm и qm элементов (так что $p+q=1$), то вычитаемое уменьшится на $dmh(p,q)$, где

$$h(p,q) = p(-\log p) + q(-\log q)$$

не превосходит единицы (это энтропия случайной величины с двумя значениями). Таким образом, прибавка в весе (удельном, в расчёте на точку) за счёт уменьшения вычитаемого не больше d.

Если множество (рассматриваемая нами часть) сильно неоднородно, то найдутся такие множества индексов A и B, что $m(A\cup B)$ много больше $m(A)\cdot m(B\,|\,A)$. Это означает, что если мы рассмотрим проекцию на $A\cup B$, то у этой проекции сечения, параллельные B, сильно различаются по размеру, и максимальное из них значительно (скажем, в l раз) превосходит среднее.

Эту проекцию (и тем самым всё множество) можно разбить на две части, считая в качестве граничного размера сечения среднее геометрическое между максимальным и средним размером. У части с «малыми» сечениями тогда уменьшится максимальный размер сечения в $\sqrt{l}$ раз. С другой стороны, большие сечения по неравенству Чебышёва составляют не более доли $1/\sqrt{l}$ (иначе бы среднее было больше), поэтому у части с большими сечениями уменьшится (также в $\sqrt{l}$ раз) размер A-проекции.

Таким образом, после разбиения для каждой из новых частей хотя бы одно слагаемое в формуле для весов уменьшится на $\log\sqrt{l}$ (а остальные, как мы уже говорили, не возрастут). Поэтому при достаточно большом l (когда $\log\sqrt{l} > d$) общий вес заведомо уменьшится. Значит, в разбиении минимального веса все части c-однородны для некоторой константы c, которая (вслед за d и l) определяется лишь числом n (хотя и быстро растёт с ростом n), что и завершает доказательство леммы с использованием весов.

10.10. Неравенства для двух и трёх слов

Мы убедились, что имеется некоторый класс неравенств, которые можно определять самыми разными способами: неравенства для энтропий, для сложностей, для

размеров проекций однородных множества, для размеров подгрупп и т. д. и т. п. Но хорошо бы понять всё-таки, каковы эти неравенства.

Это удалось сделать лишь для простейших случаев (при $n \leqslant 3$). Для $n = 1$ ситуация вообще тривиальна. При $n = 2$ имеются неравенства

$$0 \leqslant H(\xi_1) \leqslant H(\xi_1, \xi_2), \quad 0 \leqslant H(\xi_2) \leqslant H(\xi_1, \xi_2), \quad H(\xi_1, \xi_2) \leqslant H(\xi_1) + H(\xi_2),$$

которые означают, что три величины

$$H(\xi_2 \mid \xi_1),\ H(\xi_1 \mid \xi_2),\ I(\xi_1 : \xi_2)$$

неотрицательны. С другой стороны, ясно, что значения этих трёх величин могут быть любыми неотрицательными числами: возьмём три независимые величины α, β, γ с такими энтропиями и положим

$$\xi_1 = \langle \alpha, \beta \rangle, \xi_2 = \langle \beta, \gamma \rangle,$$

тогда, как легко проверить, $H(\xi_1 \mid \xi_2) = H(\alpha)$, $I(\xi_1 : \xi_2) = H(\beta)$, $H(\xi_2 \mid \xi_1) = H(\gamma)$. Таким образом, в данном случае выясняется, что указанные неравенства необходимы и достаточны, чтобы набор из трёх чисел мог быть равен

$$H(\xi_1), H(\xi_2), H(\xi_1, \xi_2),$$

и потому никаких других неравенств не нужно (любое другое неравенство будет следствием этих; заметим также, что линейное программирование учит, что всякое следствие есть линейная комбинация с неотрицательными коэффициентами).

Прежде чем переходить к случаю $n = 3$, отметим, что для $n = 2$ мы сделали нечто большее, чем просто описали все верные линейные неравенства для энтропий: мы доказали также, что все тройки чисел, удовлетворяющие этим неравенствам, могут быть значениями энтропий.

Геометрически это можно описать так. Для всякого набора случайных величин есть точка в (трёхмерном при $n = 2$) линейном пространстве, состоящая из энтропий комбинаций этих величин. Рассматривая разные наборы случайных величин, мы получаем некоторое множество $\mathscr{E}$ в этом линейном пространстве.

Линейные неравенства для энтропий — это замкнутые полупространства, содержащие целиком множество $\mathscr{E}$. В данном случае (при $n = 2$) мы установили, что $\mathscr{E}$ есть в точности многогранный конус, заданный нашими неравенствами.

В общем случае пересечение всех полупространств, содержащих некоторое множество, может быть и больше самого множества (например, если оно невыпукло). Поэтому, даже зная эти подпространства (зная все неравенства), мы имеем лишь частичную информацию о множестве возможных значений энтропий.

С другой стороны, именно неравенства (то есть полупространства, содержащие $\mathscr{E}$) представляют наибольший интерес, поскольку результат о совпадении классов неравенств для сложностей, энтропий, размеров проекций однородных множеств и т. д. относится именно к ним. Сами множества возможных значений там

другие: например, размеры проекций однородных множеств обязательно являются логарифмами целых чисел; для колмогоровских сложностей вообще всё определено лишь с точностью до $O(1)$ и так далее.

После этих комментариев перейдём к случаю $n = 3$. В этом случае имеется семимерное пространство (соответствующее семи непустым подмножествам трёхэлементного множества). Удобнее перейти к новым координатам $a_1, \ldots, a_7$ в этом векторном пространстве, описанным на с. 60 (рис. 2.4) для случая колмогоровских сложностей. В этих координатах известные нам базисные неравенства означают, что все a_i, кроме «центральной части» a_5, которую мы также обозначали $I(\xi_1 : \xi_2 : \xi_3)$, неотрицательны[1], а эта центральная часть неотрицательна в сумме с любой из трёх частей a_2, a_4, a_6. Другими словами, можно сказать, что множество $\mathscr{E} \subset \mathbb{R}^7$ всех возможных наборов энтропий содержится в множестве $\mathscr{F}$ всех семёрок, удовлетворяющих этим неравенствам.

Множество $\mathscr{F}$, как легко понять, состоит из неотрицательных линейных комбинаций конечного числа векторов (образующих этого выпуклого конуса). А именно, образующими будут векторы, в которых одно из a_i равно единице, а остальные нулю. Кроме того, есть ещё один особый вектор e: у него $a_5 = -1$, $a_2 = a_4 = a_6 = 1$, остальные a_i равны нулю. Этого достаточно: хотя a_5 может быть отрицательно, но по модулю не может превосходить a_2, a_4 и a_6, поэтому надо взять особый вектор e с коэффициентом $|a_5|$ и добавить остальные по мере необходимости.

Теперь ясно, что других неравенств нет. В самом деле, все образующие реализуются как наборы энтропий; особый вектор соответствует независимым величинам ξ_1, ξ_2 с двумя равновероятными значениями 0,1 и $\xi_3 = \xi_1 + \xi_2 \pmod 2$. Любое верное неравенство верно для образующих, и потому для всего $\mathscr{F}$, то есть вытекает из базисных неравенств.

$\boxed{\mathbf{290}}$ Покажите, что множество $\mathscr{E}$ (напомним, что мы рассматриваем случай $n = 3$) не выпукло. Например, для особого вектора e вектор λe принадлежит $\mathscr{E}$ тогда и только тогда, когда λ есть логарифм целого числа.

$\boxed{\mathbf{291}}$ Докажите, что множество $\mathscr{E}$ (для n случайных величин) замкнуто относительно сложения: если два вектора $e, e' \in \mathbb{R}^{2^n-1}$ принадлежат $\mathscr{E}$, то и их сумма $e + e'$ принадлежит $\mathscr{E}$. [Указание. Рассмотрите два независимых набора случайных величин, задающих векторы e и e', и соедините их в один.]

$\boxed{\mathbf{292}}$ Докажите, что замыкание множества $\mathscr{E}$ (для любого числа случайных величин) выпукло. [Указание. Если e и e' принадлежат множеству $\mathscr{E}$, то при больших целых k и l вектор $ke + le'$ также принадлежит множеству $\mathscr{E}$, осталось лишь научиться умножать вектор на число. На целые числа мы его умножать умеем, а приближённо делить на большое число можно так: с малой вероятностью ε берём данные случайные величины, в противном случае тривиальные (имеющие только одно значение).]

[1] На самом деле мы определяли величину $I(x_1 : x_2 : x_3)$ для слов x_1, x_2, x_3, а не для случайных величин; определение для энтропий аналогично.

10.11. Размерности и неравенство Инглетона

В случае двух и трёх случайных величин (или слов) мы описали все истинные линейные неравенства для энтропий (и тем самым для сложностей); при этом для случая $n = 2$ мы описали даже множество $\mathscr{E}$, а не только двойственное к нему множество всех неравенств, выполненных для всех элементов из $\mathscr{E}$.

Для случая четырёх и более слов такого сделать уже не удаётся, и полный набор неравенств до сих пор не известен. Сообщим, что про это известно.

Напомним, что мы рассматриваем набор случайных величин $\xi = \xi_1, \ldots, \xi_n$ и через ξ_I (для данного множества индексов $I \subset \{1, \ldots, n\}$) обозначаем часть этого набора, состоящую из величин с индексами из I. Через $H(\xi_I)$ обозначается энтропия этой части. Условные энтропии $H(\xi_I \,|\, \xi_J)$ (которые можно рассматривать только для непересекающихся множеств I и J, так как общие элементы можно удалить из I) выражаются через безусловные и потому их нет необходимости рассматривать отдельно.

Каждому набору случайных величин $\langle \xi_1, \ldots, \xi_n \rangle$ соответствует точка в пространстве $\mathbb{R}^{2^n-1}$, составленная из чисел $H(\xi_I)$ при всех непустых I. Эти точки (для всех наборов) образуют множество, которое мы обозначали через $\mathscr{E}$. Как мы видели в задачах 290, 291 и 292, это множество не обязательно выпукло, но его замыкание является выпуклым конусом (вместе с любыми двумя точками содержит их положительные линейные комбинации).

От множества $\mathscr{E}$ можно перейти к двойственному множеству всех линейных неравенств, выполненных для энтропий случайных величин (то есть для элементов $\mathscr{E}$). Геометрически это означает, что мы рассматриваем все полупространства, содержащие $\mathscr{E}$. Обратный переход даёт множество всех точек, для которых выполнены все эти неравенства (пересечение всех упомянутых полупространств); линейное программирование учит, что это пересечение есть минимальный замкнутый выпуклый конус, содержащий $\mathscr{E}$.

Среди неравенств заведомо есть следующие:

$$\begin{aligned} H(\xi_I) &\geqslant 0 && \text{для любого } I, \\ H(\xi_I) &\leqslant H(\xi_J) && \text{для любых } I \subset J, \\ H(\xi_{I\cap J}) + H(\xi_{I\cup J}) &\leqslant H(\xi_I) + H(\xi_J) && \text{для любых } I, J. \end{aligned}$$

Будем называть эти неравенства *базисными.* Этот термин раньше употреблялся по отношению к неравенству

$$H(\xi_1) + H(\xi_1, \xi_2, \xi_3) \leqslant H(\xi_1, \xi_2) + H(\xi_1, \xi_3),$$

которое соответствует случаю $I = \{1, 2\}$, $J = \{1, 3\}$, и к которому сводится неравенство для произвольных I и J, если объединить случайные величины в группы. Для удобства мы включаем в базисные неравенства и неравенства первых двух типов.

Таким образом, множество $\mathscr{E}$ содержится в многогранном конусе, состоящем из всех наборов, удовлетворяющих базисным неравенствам. В случае $n = 2$ имеет

место совпадение; при $n = 3$, как мы говорили, совпадения нет, но $\mathscr{E}$ плотно в этом конусе. При $n = 4$ это, как мы увидим, уже не так.

Выяснение ситуации естественно начать с описания структуры конуса. Всякое множество, задаваемое системой однородных линейных неравенств, есть множество всех положительных линейных комбинаций своих крайних точек (рёбер конуса). Если бы (как это происходит при $n = 3$) все эти крайние точки попали в $\mathscr{E}$, то мы бы установили, что все неравенства для энтропий следуют из базисных (потому что всякое неравенство, верное для рёбер, верно и для их положительных линейных комбинаций).

Для случая $n = 4$ такие рёбра можно найти вручную или с помощью компьютерной программы; это было сделано (см. статью [51], где приведён список всех рёбер). Выяснилось, что большинство рёбер действительно принадлежит $\mathscr{E}$ (и соответствующие случайные величины легко найти, см. [51]), но есть и несколько особых рёбер. Эти особые рёбра по существу все одинаковы (отличаются перестановкой переменных), и мы приведём одно из таких особых рёбер:

$$
\begin{gathered}
H(\xi_1) = H(\xi_2) = H(\xi_3) = H(\xi_4) = 2n,\\
H(\xi_1, \xi_2) = 4n,\\
H(\xi_1, \xi_3) = H(\xi_1, \xi_4) = H(\xi_2, \xi_3) = H(\xi_2, \xi_4) = H(\xi_3, \xi_4) = 3n,\\
H(\xi_1, \xi_2, \xi_3) = H(\xi_1, \xi_2, \xi_4) = H(\xi_1, \xi_3, \xi_4) = H(\xi_2, \xi_3, \xi_4) = 4n,\\
H(\xi_1, \xi_2, \xi_3, \xi_4) = 4n.
\end{gathered}
$$

Другими словами, каждое слово имеет сложность $2n$, все слова вместе и все тройки слов имеют сложность $4n$, а все пары имеют сложность $3n$, за исключением одной, которая имеет сложность $4n$. (Множитель n указан, поскольку ребро конуса есть луч, и его точки определены с точностью до пропорциональности.)

Довольно трудно представить себе смысл этих условий; картинки для четырёх слов довольно запутаны. Можно заметить, что величины ξ_1 и ξ_2 входят симметрично; аналогично для величин ξ_3 и ξ_4. Можно нарисовать схемы распределения сложности для троек случайных величин (рис. 10.4). Пытаясь придумать случай-

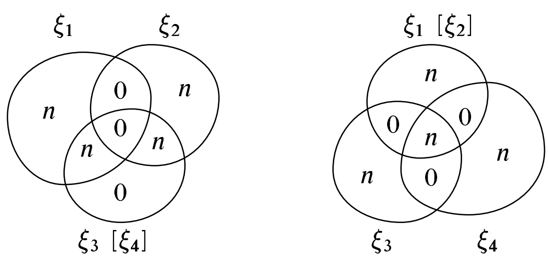


Рис. 10.4. Распределение сложностей для троек слов.

ные величины (или двоичные слова) с такой энтропией (сложностью), мы сталкиваемся со следующей трудностью: правая картинка показывает, что ξ_3 и ξ_4 хорошо бы иметь n битов общей информации, и вся эта информация входит как в ξ_1, так и

в ξ_2. С другой стороны, левая картинка показывает, что ξ_1 и ξ_2 не должны иметь общей информации (в пересечении стоят нули). Конечно, это всего лишь разговоры, поскольку слова «n битов общей информации» пока ничего определённого не означают. Однако, как мы увидим ниже, действительно таких величин $\xi_1, \xi_2, \xi_3, \xi_4$ не существует.

Если особые рёбра оказываются недостижимыми, можно предположить, что выполняются некоторые пока неизвестные нам неравенства. Как их найти? Простейшая гипотеза (как мы увидим позже, неверная, но пока мы про это не знаем) состоит в том, что мы уже знаем все рёбра конуса (неособые рёбра из числа найденных) и нужно лишь найти неравенства, образующие его грани. Это также можно сделать с помощью компьютера, и мы получаем новые неравенства

$$I(\xi_3:\xi_4) \leqslant I(\xi_3:\xi_4\,|\,\xi_1) + I(\xi_3:\xi_4\,|\,\xi_2) + I(\xi_1:\xi_2)$$

(и аналогичные, получаемые перестановкой переменных). Для наглядности мы записали это неравенство с условными сложностями. Выражая их через безусловные, мы получаем

$$12 + 3 + 4 + 134 + 234 \leqslant 13 + 23 + 14 + 24 + 34$$

(для краткости мы пишем лишь индексы: скажем, 134 обозначает $H(\xi_1, \xi_3, \xi_4)$), но в таком виде его смысл ещё менее ясен. Оказывается, что это неравенство хорошо известно в теории матроидов и называется неравенством Инглетона:

Теорема 215. *Для любых конечномерных подпространств H_1, H_2, H_3, H_4 произвольного векторного пространства выполнено неравенство*

$$\begin{gathered}\dim(H_1 + H_2) + \dim H_3 + \dim H_4 + \dim(H_1 + H_3 + H_4) + \dim(H_2 + H_3 + H_4) \leqslant \\ \leqslant \dim(H_1 + H_3) + \dim(H_2 + H_3) + \dim(H_1 + H_4) + \dim(H_2 + H_4) + \dim(H_3 + H_4).\end{gathered}$$

Прежде чем доказывать эту теорему, полезно понять связи между неравенствами для энтропий и для размерностей.

Рассмотрим конечномерное пространство X над конечным полем $\mathbb{F}$ и его подпространства. Этим подпространствам соответствуют случайные величины следующим образом. Вероятностным пространством будет множество всех линейных функционалов $X \to \mathbb{F}$, и каждому подпространству $Y \subset X$ соответствует случайная величина ξ_Y, которая сопоставляет с каждым функционалом его ограничение на Y. (Менее формально: случайная величина, соответствующая Y, есть ограничение случайного линейного функционала на подпространство Y.) Значениями такой случайной величины являются элементы пространства Y^*, сопряжённого к Y, и все они равновероятны, их число равно $\mathbb{F}$ в степени $\dim Y^* = \dim Y$, так что энтропия такой случайной величины равна $\dim Y \cdot \log|\mathbb{F}|$.

Заметим, что у нас не просто для каждого подпространства есть случайная величина, а есть некоторое их совместное распределение. Поэтому можно задать такой вопрос. Пусть Y и Z — два подпространства. Какова будет энтропия пары величин $\langle \xi_Y, \xi_Z \rangle$? Заметим, что ограничения случайного функционала на подпространства

Y и Z однозначно определяют его ограничение на $Y + Z = \{y + z \mid y \in Y, z \in Z\}$ и наоборот. Поэтому энтропия пары равна $\dim(Y + Z) \cdot \log |\mathbb{F}|$.

Из этого наблюдения сразу же следует такая

Теорема 216. *Всякому неравенству, выполненному для энтропий случайных величин и их наборов, соответствует неравенство для размерностей конечномерных подпространств векторного пространства над конечным полем (в котором энтропии группы величин соответствует размерность суммы подпространств).*

$\boxed{293}$ Покажите, что аналогичное утверждение верно для размерностей конечномерных подпространств векторного пространства над $\mathbb{R}$ и над $\mathbb{C}$. [Указание. Вещественная размерность пространства вдвое больше его комплексной размерности, так что достаточно рассмотреть вещественный случай. Будем считать пространство евклидовым и рассмотрим случайные величины, являющиеся проекциями случайной точки единичного шара на каждое из подпространств, причём в каждом пространстве значения округляются с точностью ε. При этом возникают разные погрешности (проекция случайной точки шара не равномерна в круге; из-за округления проекции на X и Y не определяют проекцию на $X + Y$ однозначно, а лишь с точностью до конечного числа вариантов), но главные члены при $\varepsilon \to 0$ всё-таки пропорциональны размерностям.]

Эти рассуждения, однако, не применимы непосредственно к произвольным бесконечным полям, и необходимо более сложное рассуждение. Во-первых, тот факт, что над некоторым полем существует пространство и набор подпространств с заданными размерностями (их самих и их сумм), можно сформулировать на языке матриц: существует матрица определённого размера, в которой определённые миноры равны нулю, а некоторые другие не равны нулю. Поэтому, если набор размерностей возможен для некоторого поля, то он возможен и для его расширений. Следовательно, можно ограничиться алгебраически замкнутыми полями. Поскольку алгебраически замкнутые поля одной характеристики элементарно эквивалентны, достаточно рассмотреть для каждой характеристики какое-то одно поле. Характеристику 0 (комплексные числа) мы уже обсуждали. Для конечной характеристики рассмотрим алгебраическое замыкание поля вычетов по модулю p. Если в нём набор размерностей реализуется матрицей, то все элементы этой матрицы алгебраичны, и можно построить конечное расширение поля вычетов по модулю p, в которое они все войдут. Тем самым мы свели дело к конечному полю, где мы умеем переходить от подпространств к случайным величинам.

$\boxed{294}$ Проведите это рассуждение подробно.

Заметим ещё, что переход от подпространств к соответствующим случайным величинам является довольно общим приёмом построения примеров точек из $\mathscr{E}$. В частности, все до сих пор встречавшиеся примеры элементов $\mathscr{E}$ могут быть получены как раз таким способом; существенно другие примеры появятся лишь в следующем разделе, когда будет идти речь об условно независимых случайных величинах.

◀ Перейдём теперь к доказательству неравенства Инглетона. Его нельзя непосредственно вывести из теоремы 216 (поскольку для энтропий оно не выполняется, как мы впоследствии увидим). Чтобы его доказать, надо установить ещё некоторые связи между энтропиями и размерностями.

В наших неравенствах для энтропий можно считать условную энтропию $H(\alpha\,|\,\beta)$ сокращением для $H(\alpha,\beta)-H(\beta)$. В соответствующем неравенстве для размерностей условная энтропия переводится как $\dim(A+B)-\dim B$, что равно размерности образа A при линейном отображении, ядро которого равно B (образа A при факторизации по B). Как учит линейная алгебра, это можно переписать как $\dim A-\dim(A\cap B)$. Другое сокращение, $I(\alpha:\beta)$, расшифровывается как $H(\alpha)+H(\beta)-H(\alpha,\beta)$, что соответствует $\dim A+\dim B-\dim(A+B)$. Последнее выражение равно $\dim(A\cap B)$. Наконец, $I(\alpha:\beta\,|\,\gamma)$ можно развернуть (не до конца) в

$$H(\alpha\,|\,\gamma)+H(\beta\,|\,\gamma)-H(\alpha,\beta\,|\,\gamma),$$

и ему соответствует

$$\dim A/C+\dim B/C-\dim(A+B)/C,$$

если через X/C обозначить пространство, получаемое из X при линейном отображении с ядром C. Заметим, что последнее выражение нельзя переписать в виде $\dim(A\cap B)/C$: это выражение равно размерности пересечения образов подпространств A и B при факторизации по C, а это пересечение содержит образ $A\cap B$ при факторизации по C, но не обязано с ним совпадать. (Пусть, например, A, B, C — три различных одномерных подпространства двумерного пространства.)

Возвращаясь к неравенству Инглетона для размерностей подпространств, мы можем переписать его так:

$$\dim(A\cap B)\leqslant I(A:B\,|\,C)+I(A:B\,|\,D)+\dim(C\cap D)$$

(где $I(A:B\,|\,C)$ обозначает размерность пересечения образов A и B при факторизации по C). Обозначая $A\cap B$ через X, мы видим, что достаточно доказать

$$\dim X\leqslant\dim X/C+\dim X/D+\dim(C\cap D),$$

поскольку $\dim X/C\leqslant I(A:B\,|\,C)$ (образ пересечения при факторизации содержится в пересечении образов, хотя может быть и меньше). А это неравенство соответствует легко проверяемому неравенству

$$H(\xi)\leqslant H(\xi\,|\,\gamma)+H(\xi\,|\,\delta)+I(\gamma:\delta)$$

для энтропий, и остаётся применить теорему 216. ▶

295 Докажите неравенство $H(\xi)\leqslant H(\xi\,|\,\gamma)+H(\xi\,|\,\delta)+I(\gamma:\delta)$ для энтропий. [Указание. Как легко заметить по картинке и проверить прямым вычислением,

$$H(\xi)+H(\xi\,|\,\gamma,\delta)+I(\gamma:\delta\,|\,\xi)=H(\xi\,|\,\gamma)+H(\xi\,|\,\delta)+I(\gamma:\delta),$$

так что рассматриваемое неравенство является суммой базисных.]

296 Строго говоря, наше доказательство неравенства Инглетона годится для пространств над конечным полем (или над полем $\mathbb{R}$, если воспользоваться соответствующей задачей). Как перенести его на случай произвольного поля? [Указание. Поле было существенно, когда мы переходили от неравенства для энтропий к неравенству для размерностей. Но это неравенство для энтропий является комбинацией базисных, а базисные неравенства верны для размерностей над произвольным полем.]

297 Мы знаем, что неравенствам для энтропий соответствуют неравенства для размеров подгрупп и их пересечений. Покажите, что неравенству Инглетона соответствует неравенство, верное для подгрупп *коммутативной* группы. [Указание: Действуйте как при доказательстве теоремы 215, роль пересечения подпространств играет сумма подгрупп (которая будет подгруппой в коммутативном случае).]

В качестве побочного продукта наших рассмотрений получаем такое любопытное утверждение:

298 Докажите, что любое линейное неравенство, выполненное для произвольных четвёрок подпространств, следует из базисных неравенств и неравенства Инглетона. [Указание. Как мы уже говорили, все рёбра конуса решений базисных неравенств, кроме упоминавшихся особых, реализуемы не только случайными величинами, но и подпространствами, а выбросив из числа образующих особые рёбра, мы получим конус, гранями которого являются базисные неравенства и неравенства Инглетона (для разных порядков переменных). Как обойтись здесь без длинных (ручных или компьютерных) вычислений, неясно.]

299 Сформулируйте и докажите аналогичное утверждение для четырёх конечных подгрупп коммутативной группы.

10.12. Условно независимые случайные величины

Покажем теперь (как мы давно обещали), что неравенство Инглетона для произвольных случайных величин может нарушаться. А именно, иногда правая его часть обращается в нуль, а левая нет. Для этого мы используем технику условной независимости; дальнейшие результаты и ссылки можно найти в [131, 94].

Говорят, что случайные величины α и β *независимы при условии* γ (где γ — третья случайная величина на том же вероятностном пространстве), если $I(\alpha:\beta\,|\,\gamma)=0$. Легко проверить, что это равносильно следующему: для любого значения γ_0 величины γ, принимаемого с ненулевой вероятностью, условные распределения α и β внутри события $\gamma=\gamma_0$ независимы.

300 Проверьте это. [Указание: $I(\alpha:\beta\,|\,\gamma)$ есть среднее значение (по всем γ_0 с весами, равными их вероятностям) взаимной информации соответствующих распределений.]

Будем говорить, что случайные величины α и β *условно независимы*, если найдутся (на том же вероятностном пространстве или на его измельчении) случайные

величины γ и δ с такими свойствами:

- γ и δ независимы;
- α и β независимы при условии γ;
- α и β независимы при условии δ.

Благодаря нашей оговорке (разрешению измельчать вероятностное пространство, разбивая элементарные события на несколько событий нужной суммарной вероятности) свойство условной независимости становится свойством совместного распределения величин α и β и не зависит от выбора вероятностного пространства. (Напомним, что здесь, как и всюду, мы рассматриваем лишь случайные величины с конечным числом значений.)

Три требования, входящие в определение условной независимости, означают, что три слагаемых в правой части неравенства Инглетона равны нулю. Осталось показать, что отсюда не следует, что левая его часть равна нулю:

Теорема 217. *Существуют условно независимые случайные величины, не являющиеся независимыми.*

◄ Для доказательства достаточно предъявить четвёрку случайных величин α, β, γ, δ, для которой бы выполнялись требования из определения условной независимости, но α и β были бы зависимыми. Укажем такой пример, следуя А. Ромащенко. Каждая из четырёх величин принимает значения 0 и 1 с вероятностью $1/2$. Величины γ и δ независимы, и каждая из четырёх комбинаций имеет вероятность $1/4$.

Величины α и β определяются так: если $\gamma = \delta$, то это общее значение будет одновременно значением величин α и β (которые в этом случае равны). Если же $\gamma \neq \delta$, то условное распределение вероятностей величин α и β (в каждом из случаев $\gamma = 1, \delta = 0$ и $\gamma = 0, \delta = 1$) задаётся таблицей

	0	1
0	1/8	3/8
1	3/8	1/8

Легко подсчитать, что (скажем) при фиксированном $\gamma = 0$ условное распределение вероятностей для α и β будет полусуммой этой матрицы и матрицы

	0	1
0	1	0
1	0	0

Оно равно

	0	1
0	9/16	3/16
1	3/16	1/16

что есть распределение двух независимых величин, равных нулю с вероятностью $3/4$. Вместе с тем совместное распределение величин α и β есть среднее арифме-

тическое всех четырёх матриц условных распределений и равно

	0	1
0	5/16	3/16
1	3/16	5/16

так что величины α и β зависимы. ▶

Следствием этой теоремы является такое утверждение: при $n = 4$ неотрицательные линейные комбинации векторов, соответствующих размерностям пространств, не исчерпывают всего $\mathscr{E}$ (поскольку для них выполнено неравенство Инглетона, которое верно не всегда). Наша нижняя оценка для замыкания множества $\mathscr{E}$ при $n = 4$ (неособые рёбра и их комбинации) оказывается, таким образом, не наилучшей.

Как мы увидим в следующем разделе, и верхняя оценка (конус решений базисных неравенств) тоже не является точной: при $n \geqslant 4$ существуют неравенства, не вытекающие из базисных, но верные для всех элементов $\mathscr{E}$.

10.13. Неравенства, не сводящиеся к базисным

Неравенства, не сводящиеся к базисным (их иногда называют *нешенноновскими*) были обнаружены в [193, 194], подробнее см. в [95]. Сейчас их известно довольно много; мы рассмотрим, пожалуй наиболее понятное (точнее, наименее непонятное) из них, приведённое в [95].

Теорема 218. *Для любых случайных величин $\alpha, \beta, \gamma, \delta, \varepsilon$ выполняется неравенство*

$$I(\alpha:\beta) \leqslant I(\alpha:\beta\,|\,\gamma) + I(\alpha:\beta\,|\,\delta) + I(\gamma:\delta) + I(\alpha:\beta\,|\,\varepsilon) + I(\alpha:\varepsilon\,|\,\beta) + I(\beta:\varepsilon\,|\,\alpha).$$

Смысл и природа этого неравенства, увы, довольно загадочны. Но некоторые комментарии к нему (хотя бы для его запоминания) всё-таки сделать можно.

Его правая часть делится на две группы, без ε и с ε (по три члена в каждой). Первая группа знакома нам по неравенству Инглетона (если бы второй группы не было, то оно бы и получилось). Таким образом, наше неравенство является ослаблением неравенства Инглетона за счёт дополнительного (неотрицательного) слагаемого

$$W(\alpha,\beta,\varepsilon) = I(\alpha:\beta\,|\,\varepsilon) + I(\alpha:\varepsilon\,|\,\beta) + I(\beta:\varepsilon\,|\,\alpha).$$

Это слагаемое (символически показанное на рис. 10.5) содержит величину ε, которая в оставшуюся часть не входит. Так что можно сказать так: неравенство Инглетона справедливо с погрешностью, которая не превосходит $\inf_\varepsilon W(\alpha,\beta,\varepsilon)$.

Другое наблюдение: если взять в качестве α, β и ε одну и ту же случайную величину ξ (не одинаково распределённые величины, а именно одну и ту же), то $W(\alpha,\beta,\varepsilon)$ обращается в нуль, и мы получаем неравенство

$$H(\xi) \leqslant H(\xi\,|\,\gamma) + H(\xi\,|\,\delta) + I(\gamma:\delta), \qquad (*)$$

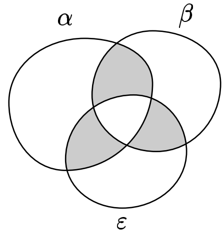


Рис. 10.5. Величина $W(\alpha,\beta,\varepsilon)$.

которое уже встречалось нам при доказательстве неравенства Инглетона для размерностей линейных пространств.

Ещё одно замечание: частным случаем теоремы 218 является такое «условное неравенство»: если $W(\alpha,\beta,\varepsilon)=0$ для некоторого ε, то

$$I(\alpha:\beta) \leqslant I(\alpha:\beta\,|\,\gamma) + I(\alpha:\beta\,|\,\delta) + I(\gamma:\delta)$$

для любых γ и δ. Из него видно, что особое ребро из раздела 10.11 не принадлежит $\mathscr{E}$ (достаточно положить $\alpha=\xi_3$, $\beta=\xi_4$, $\gamma=\varepsilon=\xi_1$, $\delta=\xi_2$), хотя удовлетворяет всем базисным неравенствам, так что неравенство теоремы 218 действительно не следует из базисных.

Кстати, это условное неравенство можно доказать и непосредственно, применив неравенство $(*)$ к величине ξ из следующей теоремы (которую нужно применять к величинам α, β и ε).

Теорема 219. *Если $W(\alpha,\beta,\gamma)=0$, то у тройки величин α,β,γ выделяется общая информация в следующем смысле: найдётся такая величина ξ, что*

$$H(\xi\,|\,\alpha)=H(\xi\,|\,\beta)=H(\xi\,|\,\gamma)=0,$$
$$I(\alpha:\beta\,|\,\xi)=I(\beta:\gamma\,|\,\xi)=I(\alpha:\gamma\,|\,\xi)=0.$$

◄ Рассмотрим трёхмерную таблицу распределений вероятностей для величин α,β,γ. Условие теоремы означает, что любое двумерное сечение этой таблицы (параллельное осям координат) имеет ранг 1 (если оно имеет ранг нуль, то соответствующее значение вообще не встречается и эту плоскость можно выбросить).

Предположим сначала, что все элементы таблицы не равны нулю и покажем, что в этом случае величины α,β,γ независимы. Рассмотрим все одномерные сечения, параллельные первой координате, как векторы. Получится двумерная таблица, в каждой клетке которой стоит ненулевой вектор. По условию в каждом столбце все векторы пропорциональны, и то же самое для строк. Поскольку векторы ненулевые, то и все они пропорциональны, то есть все плоскости таблицы (в перпендикулярном векторам направлении) пропорциональны и имеют ранг 1, что и требовалось доказать.

Рассмотрим теперь общий случай. Покажем, что наша трёхмерная таблица — блочно-диагональная. Это означает, что можно (при некотором k) представить таблицу как соединение k блоков, каждый из которых заполняет «параллелепипед» (произведение множеств по каждой координате), и проекции этих k параллелепипедов на каждую координату не пересекаются, а вне блоков стоят нули.

Из ранее доказанного следует, что внутри каждого блока имеет место независимость, то есть это как раз соответствует утверждению теоремы (величина ξ, номер блока, является функцией от любой из трёх величин α, β, γ и при известном значении ξ величины α, β, γ независимы).

Чтобы доказать, что таблица блочно-диагональная, достаточно использовать такое свойство множества ненулевых позиций (мест таблицы, где стоят положительные числа): если в двух противоположных углах прямоугольника, параллельного осям координат, стоят не-нули, то и в других двух стоят не-нули. (Иначе определитель соответствующего минора 2×2 не обращался бы в нуль и ранг был бы по крайней мере 2.)

В самом деле, будем постепенно увеличивать блок (начав с одноэлементного). Если какая-то ненулевая точка совпадает по одной из координат с точками блока, то из сформулированного свойства вытекает (как нетрудно проверить), что блок можно расширить (оставляя его параллелепипедом). Так будем расширять пока можно, после чего применим то же рассуждение к оставшейся части матрицы. ►

$\boxed{\textbf{301}}$ Проведите эти рассуждения подробно.

$\boxed{\textbf{302}}$ (**а**) Докажите следующее утверждение (иногда называемое *леммой о двойном марковском свойстве*): если $I(\beta : \gamma \mid \alpha) = 0$ и $I(\alpha : \gamma \mid \beta) = 0$, то найдётся такая величина ξ, что $H(\xi \mid \alpha) = 0$, $H(\xi \mid \beta) = 0$ и $I(\langle \alpha, \beta \rangle : \gamma \mid \xi) = 0$. (**б**) Выведите отсюда утверждение теоремы 219. [Указание к пункту (б). В качестве случайной величины ξ в теореме 219 можно взять ту же случайную величину, что и в пункте (а). Независимость α и γ при известном ξ следует из независимости пары $\alpha\beta$ и γ (и аналогично для β и γ при известном ξ). А из дополнительного условия независимости α и β при известном γ следует, что α и β независимы при известном ξ и что ξ есть функция γ. Для установления первого можно сначала нарисовать диаграмму тройки случайных величин $\alpha\beta, \gamma, \xi$ и отметить нулевую область; из диаграммы видно, что ξ имеет не меньше взаимной информации с парой $\alpha\beta$, чем имеет γ. С другой стороны, из условия $W(\alpha, \beta, \gamma) = 0$ следует, что взаимная информация $\alpha\beta$ и γ равна $I(\alpha : \beta : \gamma) = I(\alpha : \beta)$. А из условий $H(\xi \mid \alpha) = H(\xi \mid \beta) = 0$ следует, что и взаимная информация $\alpha\beta$ и ξ равна $I(\alpha : \beta : \xi)$ (что опять же видно из диаграмм). Поэтому $I(\alpha : \beta) \geqslant I(\alpha : \beta : \xi) \geqslant I(\alpha : \beta : \gamma) = I(\alpha : \beta)$ и все неравенства здесь обращаются в равенства. Равенство первых двух членов как раз и означает независимость α и β при известном ξ. Наконец, энтропия ξ при известном γ не превосходит суммы энтропии ξ при известном α, энтропии ξ при известном β и общей информации α и β при известном γ. В этой сумме все три слагаемых равны нулю.]

В [132] аналогичное теореме 219 утверждение было доказано и для колмогоровской сложности: если для тройки слов a, b, c величины $I(a : b \mid c)$, $I(a : c \mid b)$,

$I(b:c\,|\,a)$ малы (скажем, ограничены величиной $O(\log(|a|+|b|+|c|))$), то существует такое слово d, что условные сложности $KS(d\,|\,a)$, $KS(d\,|\,b)$, $KS(c\,|\,a)$, и величины относительной взаимной информации $I(a:b\,|\,d)$, $I(b:c\,|\,d)$, $I(a:c\,|\,d)$ также малы (тоже не превосходят $O(\log(|a|+|b|+|c|))$).

Независимизация

Всё сказанное, однако, позволяет доказать теорему 218 лишь в частных случаях. Вот общее доказательство:

◀ Мы используем (до конца ещё, видимо, не понятый) приём «независимизации», который состоит в следующем.

Переменные, входящие в неравенство, делятся на три группы:

$$(1)\ \alpha,\beta;\quad (2)\ \gamma,\delta;\quad (3)\ \varepsilon.$$

Заметим, что переменные второй и третьей групп никогда не встречаются в неравенстве вместе (хотя и те, и другие по отдельности сочетаются с переменными первой группы). Поэтому *без ограничения общности можно предполагать, что пара* $\langle\gamma,\delta\rangle$ *независима с* ε *при известном* $\langle\alpha,\beta\rangle$. В самом деле, если мы изменим совместное распределение, выполнив (для каждого значения пары $\langle\alpha,\beta\rangle$) «принудительную независимизацию» величин $\langle\gamma,\delta\rangle$ и ε, то совместное распределение величин первой и второй групп, а также совместное распределение величин первой и третьей групп, останутся неизменными, и все величины, входящие в неравенство, не изменятся.

Раз так, то достаточно доказать (вообще говоря, более слабое) неравенство

$$\begin{aligned}I(\alpha:\beta)\leqslant I(\alpha:\beta\,|\,\gamma)+I(\alpha:\beta\,|\,\delta)+I(\gamma:\delta)+W(\alpha,\beta,\varepsilon)&+\\ +I(\langle\gamma,\delta\rangle:\varepsilon\,|\,\langle\alpha,\beta\rangle)&+\\ +I(\gamma:\varepsilon\,|\,\langle\alpha,\beta\rangle)&+\\ +I(\delta:\varepsilon\,|\,\langle\alpha,\beta\rangle)&.\end{aligned}$$

В самом деле, если пара $\langle\gamma,\delta\rangle$ независима с ε при известных α,β, то каждая из величин γ и δ по отдельности также независима с ε при известных α,β. Последнее неравенство удивительным образом является суммой восьми базисных неравенств

$$\begin{aligned}I(\langle\alpha,\beta\rangle:\varepsilon\,|\,\gamma,\delta)&\geqslant 0,\\ I(\alpha:\beta\,|\,\varepsilon,\gamma)&\geqslant 0,\\ I(\alpha:\beta\,|\,\varepsilon,\delta)&\geqslant 0,\\ I(\gamma:\delta\,|\,\varepsilon)&\geqslant 0,\\ I(\gamma:\varepsilon\,|\,\alpha)&\geqslant 0,\\ I(\gamma:\varepsilon\,|\,\beta)&\geqslant 0,\\ I(\delta:\varepsilon\,|\,\alpha)&\geqslant 0,\\ I(\delta:\varepsilon\,|\,\beta)&\geqslant 0.\end{aligned}$$

Это несложно проверить: если раскрыть все взаимные информации и сложить, то все лишние члены замечательным образом сокращаются. (Но почему так получается и как до этого можно было догадаться, неясно.) ▶

Таким образом, наше неравенство, хотя и не в прямом смысле, является следствием базисных. Можно сказать, что мы обнаружили (помимо сложения неравенств с неотрицательными коэффициентами) ещё одно «правило вывода»: если удаётся разбить переменные в каком-то неравенстве на три группы, причём переменные второй и третьей групп не входят одновременно ни в одно слагаемое, то можно вывести это неравенство из (вообще говоря) более слабого неравенства, куда добавлена ещё и общая информация переменных второй и третьей групп при известных переменных первой группы.

Пользуясь данным методом, можно доказать и другие «неклассические» информационные неравенства. Более того, известно, что таким способом можно вывести бесконечно много линейно независимых неравенств (известно бесконечно много независимых неравенств даже для четвёрки случайных величин).

Удаление уникальной информации

Есть и другой приём вывода новых информационных неравенств. Он основан на следующей теореме Алсведе – Кёрнера [2], которую мы приводим без доказательства (оно приведено в [95], лемма 5).

Теорема 220. *Пусть заданы совместно распределённые случайные величины $\alpha, \beta, \varepsilon$. Рассмотрим n независимых копий этого распределения $\alpha_i, \beta_i, \varepsilon_i$ $(i = 1, \ldots, n)$ и обозначим*

$$A = (\alpha_1, \ldots, \alpha_n), B = (\beta_1, \ldots, \beta_n), E = (\varepsilon_1, \ldots, \varepsilon_n).$$

Тогда на одном вероятностном пространстве с A, B, E можно определить случайную величину E', для которой размеры всех семи областей на диаграмме тройки A, B, E' (рис. 2.4) такие же, как у тройки A, B, E (с точностью $o(n)$), за исключением области, содержащей информацию о E', не входящую в A, B, размер которой теперь равен $o(n)$. (Другими словами, $H(E' \mid A, B) = o(n)$.)

Неформально говоря, случайная величина E' получается из E удалением всей информации, которая не входит в пару A, B. Аналогичное утверждение верно для любого количества совместно распределенных случайных величин $\alpha_1, \ldots, \alpha_k$: проведём серию из n независимых испытаний, и обозначим через $A_1, \ldots, A_k$ полученные случайные величины. Из случайной величины A_k можно удалить «уникальную информацию», не входящую в $A_1, \ldots, A_{k-1}$ (точнее, сделать её равной $o(n)$), изменив размеры всех остальных областей на диаграмме для $A_1, \ldots, A_k$ всего лишь на $o(n)$.

$\boxed{\textbf{303}}$ Докажите теорему Алсведе – Кёрнера для $k = 2$.

Заметим, что переход к независимым испытаниям существен: если α и ε — равномерно распределённые зависимые величины с двумя значениями, то построить ε', которое было бы функцией от α и имело бы нужную энтропию, нельзя.

Теперь объясним, как из теоремы Алсведе – Кёрнера следует теорема 218. Пусть заданы некоторые совместно распределённые случайные величины $\alpha, \beta, \gamma, \delta, \varepsilon$. Сначала, не используя этой теоремы, докажем неравенство теоремы 218 с добавочным членом $3H(\varepsilon\,|\,\alpha,\beta)$ в правой части. Нетрудно убедиться, что оно получается сложением неравенств

$$\begin{aligned} H(\varepsilon\,|\,\gamma) &\leqslant H(\varepsilon\,|\,\alpha) + H(\varepsilon\,|\,\beta) + I(\alpha:\beta\,|\,\gamma),\\ H(\varepsilon\,|\,\delta) &\leqslant H(\varepsilon\,|\,\alpha) + H(\varepsilon\,|\,\beta) + I(\alpha:\beta\,|\,\delta),\\ H(\varepsilon) &\leqslant H(\varepsilon\,|\,\gamma) + H(\varepsilon\,|\,\delta) + I(\gamma:\delta) \end{aligned}$$

с равенством

$$I(\alpha:\beta) + 2H(\varepsilon\,|\,\alpha) + 2H(\varepsilon\,|\,\beta) = H(\varepsilon) + W(\alpha,\beta,\varepsilon) + 3H(\varepsilon\,|\,\alpha,\beta).$$

Третье неравенство нам уже встречалось (задача 295 на с. 382), а два предыдущих следуют из его условной версии. Все они являются линейными комбинациями базисных неравенств. Равенство же легко проверить, нарисовав диаграмму для случайных величин $\alpha, \beta, \varepsilon$.

Теорема Алсведе – Кёрнера позволит нам избавиться от «паразитного» члена $3H(\varepsilon\,|\,\alpha,\beta)$. Для этого рассмотрим n независимых копий $\alpha_i, \beta_i, \gamma_i, \delta_i, \varepsilon_i$ (при $i=1,\ldots,n$) нашего распределения и положим $A=(\alpha_1,\ldots,\alpha_n)$, $B=(\beta_1,\ldots,\beta_n)$, $C=(\gamma_1,\ldots,\gamma_n)$, $D=(\delta_1,\ldots,\delta_n)$ $E=(\varepsilon_1,\ldots,\varepsilon_n)$. Энтропии A, B, C, D, E в n раз больше энтропий $\alpha, \beta, \gamma, \delta, \varepsilon$; то же самое верно для пар, троек и так далее. Применим к случайным величинам $\alpha, \beta, \varepsilon$ теорему Алсведе – Кёрнера и получим новую случайную величину E', определённую на том же вероятностном пространстве, что и A, B, C, D, E. Затем применим к пятерке A, B, C, D, E' только что доказанное неравенство с дополнительным членом $3H(E'\,|\,A,B)$ в правой части. Заменим в этом неравенстве дополнительный член на $o(n)$. Затем заменим всюду E' на E. Все члены неравенства, содержащие C или D, не содержат E', поэтому они замены не почувствуют. (В этом месте мы используем ту же особенность нашего неравенства, что в методе независимизации.) А каждый из членов неравенства, содержащих только A, B, E', является суммой размеров областей диаграммы A, B, E', отличных от области $H(E'\,|\,A,B)$. Поэтому в результате замен неравенство сохранится с точностью $o(n)$. Наконец, сокращая на n и переходя к пределу, получаем неравенство теоремы 218.

Полезно сравнить методы независимизации и удаления уникальной информации. И в том, и в другом случае мы модифицировали совместное распределение случайных величин A, B, C, D, E (при независимизации мы тоже могли перейти к серии испытаний — это бы не помешало, но и не помогло). В первом случае мы сохраняли совместные распределения A, B, E и A, B, C, D. Во втором случае мы сохраняли только совместное распределение A, B, C, D. В первом случае мы убивали член $I(CD:E\,|\,AB)$, а во втором — член $H(E\,|\,AB)$, не изменяя (или изменяя несущественно) все остальные члены в неравенстве.

11. Общая информация

11.1. Представление слов в несжимаемом виде

В какой степени можно представлять себе «информацию, содержащуюся в данном слове», как нечто материальное? Допустим, у нас есть слово x, «содержащее n битов информации», то есть имеющее сложность n. Можно ли — как если бы эти биты были камешками — разложить их на две равные кучки? Этому вопросу можно придать формальный смысл: можно ли найти такие слова x_1 и x_2 сложности $n/2$, для которых $KS(x_1 \mid x) \approx 0$, $KS(x_2 \mid x) \approx 0$ (слова x_1, x_2 «не содержат новой информации по сравнению с x») и $KS(x \mid x_1, x_2) \approx 0$ («никакая информация не потеряна»)? Если, как обычно, понимать приближённое равенство как совпадение с точностью до $O(\log n)$, где n — максимальная длина (или сложность) рассматриваемых слов, то ответ на этот вопрос положительный, как мы сейчас увидим.

Будем говорить, что слова x и y *эквивалентны с точностью c*, или *c-эквивалентны*, если $KS(x \mid y) \leqslant c$ и $KS(y \mid x) \leqslant c$. Это отношение не является, конечно, настоящим отношением эквивалентности: если x эквивалентно y, а y эквивалентно z с точностью c, то мы можем утверждать лишь, что x эквивалентно z с точностью $2c + O(\log c)$. (Можно было бы рассматривать не слова, а последовательности слов полиномиально растущей длины и говорить, что последовательность $x_0, x_1, \ldots$ эквивалентна $y_0, y_1, \ldots$, если $KS(x_i \mid y_i) = O(\log i)$ и $KS(y_i \mid x_i) = O(\log i)$; тогда бы это было настоящее отношение эквивалентности.)

Легко проверить, что сложности c-эквивалентных слов мало отличаются (отличаются не более чем на $O(c)$ и даже $c + O(\log c)$). Более общо, при замене слова на другое, c-эквивалентное первому, все сложности с его участием меняются на $O(c)$, так что, скажем, величина $I(x:y \mid z)$ меняется не более чем на $O(c)$ при замене любого из слов x, y, z (или всех трёх) на c-эквивалентное. Теперь мы можем сделать следующее (почти очевидное) наблюдение:

Теорема 221. *Для всякого слова x существует слово x' длины $KS(x)$, эквивалентное x с точностью до $O(\log KS(x))$. Это слово является «несжимаемым» (сложность близка к длине) с той же точностью.*

◀ В самом деле, возьмём в качестве x' любое кратчайшее описание слова x, которое как раз имеет длину $KS(x)$ и сложность, отличающуюся от $KS(x)$ не более, чем на константу. Ясно, что $KS(x \mid x') = O(1)$, поскольку x восстанавливается по x'. С другой стороны, цепочка неравенств

$$KS(x) \leqslant KS(x, x') \leqslant KS(x') \leqslant l(x') = KS(x)$$

(выполненных с точностью до константы), показывает, что все эти неравенства обращаются в равенства (с той же точностью), и потому $KS(x,x') \approx KS(x)$ и по теореме о сложности пары $KS(x'\,|\,x) \approx 0$ с точностью до $O(\log KS(x))$. ►

Из этого утверждения очевидно следует ответ на поставленный нами вопрос о разбиении слова на части: можно заменить слово на эквивалентное несжимаемое, и в качестве x_1 и x_2 взять просто-напросто левую и правую половины.

$\boxed{\textbf{304}}$ Проверьте, что выполнены все требуемые свойства.

$\boxed{\textbf{305}}$ Докажите, что если $KS(y\,|\,x) = n$, то существует такое («промежуточное») z, что $KS(z\,|\,x) \approx n/2$ и $KS(y\,|\,z) \approx n/2$ с точностью до $O(\log KS(x,y))$.

Так что пока биты информации ведут себя вполне материальным образом. Более того, это верно и в более общем случае, когда одновременно рассматривается информация в слове и её часть.

Пусть даны слова x и y, причём $KS(y\,|\,x) \approx 0$ («информация в слове y является частью информации в слове x»). Тогда существует несжимаемое слово x', эквивалентное x, некоторое начало y' которого эквивалентно y. (Такое начало автоматически будет несжимаемым и иметь длину $KS(y)$.) Более точно это утверждение формулируется так:

Теорема 222. *Для любых слов x и y существуют слова x' и y', эквивалентные x и y (соответственно) с точностью до $O(KS(y\,|\,x) + \log KS(x,y))$ и несжимаемые с той же точностью, причём y' является началом x'.*

◄ Рассмотрим в качестве y' кратчайшее описание слова y, оно несжимаемо и имеет длину $KS(y)$. Кроме того, оно эквивалентно y.

Теперь рассмотрим кратчайшее описание z' слова x при известном слове y; оно также будет несжимаемым, а длина его равна $KS(x\,|\,y)$. Зная y' и z', можно восстановить сначала y, а потом и x, поэтому сложность пары y', z' не меньше $KS(x,y)$. С другой стороны, суммарная длина слов y' и z' равна $KS(y) + KS(x\,|\,y) \approx KS(x,y)$, так что слово $x' = y'z'$ несжимаемо.

Как мы видели, $KS(x\,|\,x') \approx 0$. Осталось доказать, что $KS(x'\,|\,x) \approx 0$. Из равенства $KS(x\,|\,x') \approx 0$ следует $KS(x,x') \approx KS(x') \approx KS(x,y)$. Кроме того, $KS(x,y)$ примерно равно $KS(x)$ (с точностью $O(KS(y\,|\,x))$). Следовательно, $KS(x,x') \approx KS(x)$, откуда по теореме Колмогорова–Левина отсюда следует $KS(x'\,|\,x) \approx 0$. ►

Эта теорема показывает, что пару слов x, y с $KS(y\,|\,x) \approx 0$ действительно можно себе представлять так: слово x состоит из $KS(x)$ почти что материальных битов, из которых первые $KS(y)$ битов образуют y.

11.2. Выделение общей информации

Можно пойти ещё дальше и интересоваться аналогичным утверждением для двух произвольных слов. Напомним, что три основных сложностных характеристики пары слов x, y — это их сложности $KS(x)$, $KS(y)$ и сложность пары $KS(x,y)$; через них, как мы говорили в разделе 2.3, выражаются (с логарифмической точ-

ностью) условные сложности и взаимная информация:

$$\begin{aligned} KS(x\,|\,y) &= KS(x,y) - KS(y),\\ KS(y\,|\,x) &= KS(x,y) - KS(x),\\ I(x:y) &= KS(x) + KS(y) - KS(x,y) \end{aligned}$$

(см. рис. 2.2 на с. 56). Там же говорилось, что иногда этот рисунок можно понимать буквально, взяв в качестве x и y пересекающиеся подслова длинного случайного слова. Можно предположить, что это общий случай, то есть верна следующая гипотеза:

> для любых слов x и y существует несжимаемое слово u длины $KS(x,y)$, эквивалентное (с логарифмической точностью) паре $\langle x,y\rangle$, причём начало этого слова длины $KS(x)$ эквивалентно x, а конец длины $KS(y)$ эквивалентен y (с той же точностью).

Но это, как выясняется, не так. Чтобы убедиться в этом, заметим, что из нашей гипотезы (для слов x и y) следует, что существует слово z (средняя часть), для которой

$$\begin{aligned} KS(z\,|\,x) &= 0,\\ KS(z\,|\,y) &= 0,\\ KS(z) &= I(x:y) \end{aligned}$$

(все равенства понимаются с логарифмической точностью). Можно сказать, что в этом случае «из слов x и y выделяется общая информация z». Как мы вскоре увидим, это возможно далеко не для всех пар слов x и y.

$\boxed{\mathbf{306}}$ Покажите, что если общая информация выделяется (найдётся слово z с указанными тремя свойствами), то для пары слов x, y выполнена сформулированная выше гипотеза (найдётся слово u с указанными свойствами).

Построим опровергающий нашу гипотезу пример. Это можно сделать самыми разными способами, о которых мы ещё будем говорить, но, видимо, самый простой из них (предложенный Ан. А. Мучником в [115]) таков. Для определённости договоримся, что строимые нами слова x и y с невыделяемой общей информацией будут иметь сложность $2n$, а общая информация будет равна n (с логарифмической точностью), то есть пара $\langle x,y\rangle$ должна иметь сложность $3n$.

Мы хотим, чтобы из этой пары не выделялась общая информация, то есть чтобы не существовало слова z сложности n, для которого $KS(z\,|\,x)$ и $KS(z\,|\,y)$ обращаются в нуль (все равенства — с точностью $O(\log n)$). Нам удобнее будет записать эти условия иначе: $KS(x\,|\,z) = n$ и $KS(y\,|\,z) = n$. Это утверждение будет выполняться с некоторым запасом (мы разрешим сложностям быть даже немного больше n):

Теорема 223. *Для любого n существуют слова x и y, для которых $KS(x) = 2n + O(\log n)$, $KS(y) = 2n + O(\log n)$, $I(x:y) = n + O(\log n)$, но не существует такого слова z сложности меньше $1{,}1n$, что $KS(x\,|\,z) < 1{,}1n$ и $KS(y\,|\,z) < 1{,}1n$.*

◀ Как мы знаем, список всех слов сложности меньше k сам имеет сложность не больше $k + O(\log k)$ (чтобы его задать, достаточно, помимо числа k, указать число таких слов, которое имеет порядок $O(2^k)$ и потому требует k битов).

Аналогичное рассуждение позволяет установить, что список всех пар слов $\langle u, v\rangle$, для которых $KS(u) < 1{,}1n$ и $KS(v \mid u) < 1{,}1n$, имеет сложность не больше $2{,}2n + O(\log n)$: помимо n, достаточно указать число таких пар, а оно есть $O(2^{2,2n})$.

По тем же причинам список всех слов сложности меньше $2n$ имеет сложность не больше $2n + O(\log n)$, а список слов сложности меньше $3n$ имеет сложность не больше $3n + O(\log n)$.

Что можно сказать о сложности всех этих списков вместе взятых? Её можно оценить суммой сложностей. Но оказывается, что есть лучшая оценка: сложность всех списков вместе *не превосходит максимальной из наших оценок* (с точностью до $O(\log n)$). В самом деле, при известном n каждый из этих списков задаётся числом своих элементов. Более того, достаточно знать суммарное число элементов в этих списках, поскольку их можно перечислять и ждать появления нужного числа элементов: достигнув нужной суммы, мы автоматически исчерпаем все списки. А сумма нескольких чисел требует для записи лишь на несколько битов больше, чем самое длинное из них.

Итак, все вместе эти списки имеют сложность $3n + O(\log n)$. Пусть эти списки у нас есть. Используя их, мы перебором находим пару слов x, y длины $2n + 2$ каждое, для которых:

- $KS(x) \geqslant 2n$;
- $KS(y) \geqslant 2n$;
- $KS(x, y) \geqslant 3n$;
- не существует слова z, для которого одновременно $KS(z) < 1{,}1n$, $KS(x \mid z) < 1{,}1n$ и $KS(y \mid z) < 1{,}1n$.

Все эти условия проверяются по нашим спискам, так что надо только убедиться, что такая пара существует. Заметим, что первое условие бракует не более четверти всех пар (всего слов x имеется 2^{2n+2}, из них негодных не больше 2^{2n}). Аналогично со вторым условием. Третье отбрасывает не более 2^{3n} пар, что ещё меньше. Наконец, четвертое условие бракует не более $2^{1,1n} \times 2^{1,1n}$ пар для каждого из не более чем $2^{1,1n}$ значений z, то есть всего не более $2^{3,3n}$ пар, что тоже очень мало по сравнению с общим числом пар.

Сложность найденной таким образом (первой в порядке перебора) пары будет не меньше $3n$ и не больше сложности наших списков, то есть $3n + O(\log n)$; сложность каждого слова будет не меньше $2n$ и не больше его длины, то есть $2n + O(1)$, и построение гарантирует, что общая информация не выделяется. ▶

Видно, что в нашем рассуждении есть «запас»; проведём подсчёты более аккуратно. Пусть мы хотим построить слова x и y сложности $2n$ с общей информацией n, но так, чтобы нельзя было найти слово z, для которого одновременно

$$KS(z) < \alpha, \quad KS(x \mid z) < \beta \quad \text{и} \quad KS(y \mid z) < \gamma.$$

Список всех пар u, v с $KS(u) < \alpha$ и $KS(v\,|\,u) < \beta$ имеет сложность $\alpha + \beta$, поэтому возникает условие $\alpha + \beta < 3n$. Аналогично появляется условие $\alpha + \gamma < 3n$. Наконец, для существования пары нужно, чтобы запрещённых пар было меньше 2^{4n}; это даёт условие $\alpha + \beta + \gamma < 4n$. Если все эти условия выполнены, то искомую пару x, y построить можно.

Более того, мы можем параллельно проводить отбраковку пар для всех троек целых чисел α, β, γ, удовлетворяющих указанным неравенствам. Таких троек $O(n^3)$, поэтому если неравенства выполняются с логарифмическим запасом, то останутся неотбракованные пары; сложность объединения полиномиального числа перечислимых списков также увеличится на $O(\log n)$. Получаем такой результат:

Теорема 224. *Для любого n существуют слова x и y сложности $2n + O(\log n)$, для которых $KS(x, y) = 3n + O(\log n)$, и при этом для любого слова z выполнено хотя бы одно из трёх неравенств*:

(а) $KS(z) + KS(x\,|\,z) \geqslant 3n - O(\log n)$;

(б) $KS(z) + KS(y\,|\,z) \geqslant 3n - O(\log n)$;

(в) $KS(z) + KS(x\,|\,z) + KS(y\,|\,z) \geqslant 4n - O(\log n)$.

Более подробно утверждение теоремы формулируется так: для некоторой константы c и для всех n существуют слова x и y, сложности которых отличаются от $2n$ не более чем на $c \log n$, сложность пары отличается от $3n$ не более чем на $c \log n$ и для любого слова z выполнено одно из трёх неравенств с заменой $O(\log n)$ в правой части на $c \log n$.

Построенная в этой теореме пара слов является «наихудшей» с точки зрения возможности выделения общей информации. Это можно уточнить следующим образом. Для данной пары x, y можно рассмотреть множество $C(x, y) \subset \mathbb{N}^3$, состоящее из всех троек α, β, γ, для которых существует слово z с

$$KS(z) < \alpha, \quad KS(x\,|\,z) < \beta \quad \text{и} \quad KS(y\,|\,z) < \gamma.$$

Множество $C(x, y)$ «наследственно вверх» (вместе с каждой тройкой содержит и все покоординатно бóльшие) и зависит от пары x, y (а не только от сложностей слов x, y и сложности пары); как мы увидим на примере пар с $KS(x) = KS(y) = 2n$ и $KS(x, y) = 3n$, соответствующие множества могут сильно различаться. При этом для пары пересекающихся подслов несжимаемого слова множество $C(x, y)$ будет максимально возможным, а для построенной в теореме 224 — минимально возможным.

Рассмотрим этот вопрос более детально. Пусть даны слова x, y сложности $2n$ с общей информацией n (как обычно, мы допускаем отклонения порядка $O(\log n)$, не оговаривая этого особо). Любая тройка $\langle \alpha, \beta, \gamma \rangle$, принадлежащая $C(x, y)$, должна удовлетворять (с логарифмической точностью) очевидным неравенствам

$$\alpha + \beta \geqslant 2n, \quad \alpha + \gamma \geqslant 2n, \quad \alpha + \beta + \gamma \geqslant 3n$$

(поскольку $KS(x) \leqslant KS(z) + KS(x\,|\,z)$ и так далее). Это означает, что множество $C(x, y)$ содержится в множестве C_M всех троек, удовлетворяющих этим неравенствам. (И здесь аккуратная формулировка требовала бы оговорок о логарифмических погрешностях, которые мы для краткости опускаем.)

Как изобразить это множество наглядно? При каждых β и γ есть некоторое минимальное значение $\alpha_0(\beta,\gamma)$, начиная с которого $\langle\alpha,\beta,\gamma\rangle$ попадает в C_M. График функции $\langle\beta,\gamma\rangle\mapsto\alpha_0(\beta,\gamma)$, представляет собой поверхность, состоящую из трёх плоских частей (соответствующих трём неравенствам), которую можно изобразить с помощью линий уровня (рис. 11.1). Каждая линия уровня соответствует некоторому значению α и отделяет на плоскости β,γ удовлетворяющие неравенствам пары от не удовлетворяющих.

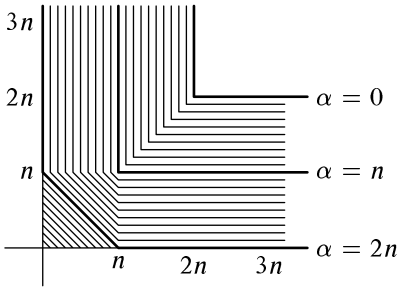


Рис. 11.1. Множество C_M.

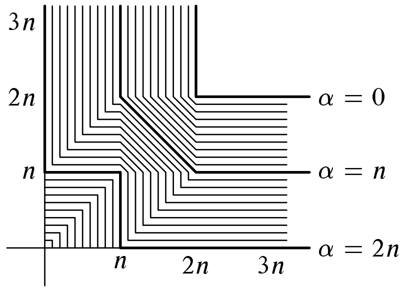


Рис. 11.2. Множество C_m.

Глядя на эту картинку, можно убедиться, что для случая пары $\langle x,y\rangle$ с полностью выделяемой общей информацией (пересекающиеся куски несжимаемого слова) множество $C(x,y)$ равно C_M, так что верхняя оценка C_M для этого множества достижима. Несложно понять, как нужно выбирать слово z для данных α, β и γ: при $\alpha<n$ выбираем z внутри общей части, тем самым сокращая условную сложность обоих слов x и y на α. При $\alpha>n$, помимо общей части, можно присоединить к z куски из остатков слов x и y (в некоторой пропорции; изменение пропорции соответствует сдвигу по наклонному отрезку на рисунке).

Теорема 224 даёт пример пары x,y с меньшим множеством $C(x,y)$. В самом деле, для построенной в этой теореме пары множество $C(x,y)$ содержится не только в C_M, но и в объединении множеств решений неравенств

$$\alpha+\beta\geqslant 3n,\quad \alpha+\gamma\geqslant 3n,\quad \alpha+\beta+\gamma\geqslant 4n$$

(соответствующих неравенствам (а) – (в) этой теоремы). Пересекая C_M с этим объединением, мы получаем меньшее множество, которое мы назовём C_m; оно изображено на рисунке 11.2.

Таким образом, для построенной в теореме пары x,y множество $C(x,y)$ содержится в C_m. Оказывается, что оно совпадает с C_m; более того, для любой пары x,y (а не только для пары из теоремы) множество C_m содержится в $C(x,y)$.

Чтобы проверить это, надо для каждой точки $\langle\alpha,\beta,\gamma\rangle$ из C_m указать подходящее z. Достаточно делать это для минимальных троек (поскольку увеличение не выводит из $C(x,y)$). Точки на наклонных линиях на рисунке соответствуют словам z, которые соединяют часть слова x с частью слова y в некоторой пропорции. Например, точка $(1{,}5n,1{,}5n)$ при $\alpha=n$ соответствует слову z длины n, состоящему из двух половин; левая половина содержит $n/2$ битов из кратчайшего описания

слова x, а правая содержит $n/2$ битов кратчайшего описания слова y. Аналогичным образом точка (n, n) соответствует слову длины $2n$, содержащему по n битов того и другого описаний. Точка $(n+h, h)$ (для h от 0 до n) соответствует слову z длины $2n-h$, являющемуся началом кратчайшего описания слова y. Тогда $KS(y\,|\,z)$ равно числу битов этого описания, не вошедших в z, то есть h. С другой стороны, $KS(x\,|\,z)$ не превосходит суммы сложности x относительно y, равной n, и сложности y относительно z, равной h. Наконец, точки вида $(h, 0)$ при h от 0 до n соответствуют словам z сложности $3n-h$, которые содержат полностью всё слово y и $n-h$ битов кратчайшего описания слова x при известном y.

В итоге получаем такое утверждение:

Теорема 225. *Для любой пары слов x, y множество $C(x, y)$ заключено (с обычными оговорками о логарифмической точности) между нижней оценкой C_m и верхней C_M; обе оценки достигаются для некоторых пар.*

Зная множество $C(x, y)$ для данных слов x, y, можно уже чисто формально получать разные следствия. Вот один пример:

Теорема 226. *Пусть $C(x, y) = C_m$ (пара из теоремы 224). Тогда для любого слова z выполняется неравенство*

$$KS(z) \leqslant 2\,KS(z\,|\,x) + 2\,KS(z\,|\,y).$$

Прежде, чем доказывать теорему, поясним её смысл. В ней утверждается, что только слова z малой сложности могут быть просты относительно x и y одновременно. Заметим, что если у x и y выделяется общая информация, то существует такое z большой сложности, например, выделенная общая информация x, y. Поэтому неудивительно, что в условии теоремы требуется невыделяемость (в сильном смысле) общей информации у x, y.

◀ Можно считать, что $KS(z) = O(n)$ (для слов большой сложности $KS(z\,|\,x)$ и $KS(z\,|\,y)$ близки к $KS(z)$ и уж заведомо больше половины $KS(z)$).

Перепишем неравенство, выразив $KS(z\,|\,x)$ и $KS(z\,|\,y)$ через величины, входящие в определение $C(x, y)$:

$$KS(z) \leqslant 2\,KS(x, z) - 2\,KS(x) + 2\,KS(y, z) - 2\,KS(y)$$

и далее

$$KS(z) \leqslant 2\,KS(z) + 2\,KS(x\,|\,z) - 2\,KS(x) + 2\,KS(z) + 2\,KS(y\,|\,z) - 2\,KS(y),$$

то есть

$$2\,KS(x) + 2\,KS(y) \leqslant 3\,KS(z) + 2\,KS(x\,|\,z) + 2\,KS(y\,|\,z).$$

Левая часть равна $8n$, а правая не может быть меньше $8n$, в чём легко убедиться, рассматривая каждую линию уровня (причём на ней достаточно проверить точки с минимальной суммой координат). ▶

307 Покажите, что при малых значениях $KS(z)$ верна лучшая оценка

$$KS(z) \leqslant KS(z\,|\,x) + KS(z\,|\,y),$$

но если не ограничивать z, то константу 2 улучшить нельзя.

11.3. Комбинаторный смысл общей информации

Построенный в предыдущем пример пары слов, из которых не выделяется общая информация, всё же не даёт наглядного объяснения, чем именно отличаются такие слова. Полностью удовлетворительного ответа на этот (неформальный) вопрос мы не знаем, но некоторые наблюдения сделать можно.

Что означает, что для данных слов x и y найдётся слово z, при котором $KS(z) < \alpha$, $KS(x\,|\,z) < \beta$ и $KS(y\,|\,z) < \gamma$? Обозначим через $U_n(z)$ множество всех слов, сложность которых при известном z меньше n. Таких слов (примерно) 2^n. Наше условие означает, что пара $\langle x, y\rangle$ принадлежит одному из множеств $U_\beta(z) \times U_\gamma(z)$ при $KS(z) < \alpha$; всего таких множеств имеется (примерно) 2^α.

Таким образом, условие $\langle \alpha, \beta, \gamma\rangle \in C(x, y)$ означает, что пара $\langle x, y\rangle$ принадлежит объединению перечислимого семейства из 2^α «прямоугольников» размера $2^\beta \times 2^\gamma$ (прямоугольником мы называем декартово произведение любых двух множеств слов). Верно и обратное: если пара $\langle x, y\rangle$ покрывается объединением перечислимого семейства из 2^α прямоугольников размера $2^\beta \times 2^\gamma$, то тройка $\langle \alpha, \beta, \gamma\rangle$ (плюс сложность порождающего это семейство алгоритма и логарифмическая добавка) принадлежит $C(x, y)$. Таким образом, можно сказать, что множество $C(x, y)$ для данной пары зависит от того, какие (порождаемые простыми алгоритмами) перечислимые семейства прямоугольников её покрывают.

Отсюда ясен план построения другого примера пары слов, из которых не выделяется информация: надо взять множество пар, плохо покрываемое прямоугольниками, а потом взять его случайный элемент.

Удобно представлять себе множества пар как бинарные отношения, или двудольные графы, изображая пару $\langle x, y\rangle$ как ребро, соединяющее точку x в левой доле графа с точкой y в его правой доле. Прямоугольник, таким образом, задаётся множествами вершин слева и справа и покрывает все рёбра, соединяющие вершины этих множеств.

Свойство, которое будет гарантировать, что граф плохо покрывается прямоугольниками, совсем простое: мы будем рассматривать графы, в которых нет четырёхугольников (не существует таких вершин a, b в одной доле и c, d в другой, что все четыре ребра ac, ad, bc, bd входят в граф). Имеет место следующая простая комбинаторная

Лемма. *Рассмотрим произвольный двудольный граф с l вершинами слева и L вершинами справа (считая для определённости, что $l \leqslant L$). Если этот граф не содержит четырёхугольников, то плотность рёбер в нём (число рёбер, делённое на Ll) не превосходит большей из оценок $O(1/\sqrt{L})$ и $O(1/l)$.*

(Другими словами, если в прямоугольной таблице расставлены звёздочки так, что никакие четыре из них не стоят в вершинах прямоугольника, параллельного сторонам таблицы, то доля звёздочек не превосходит $O(1)$, делённого либо на меньшую сторону, либо на квадратный корень из большей стороны.)

Доказательство. Каждой из l вершин с левой стороны соответствует множество её соседей справа. Условие леммы (отсутствие четырёхугольников) означает, что пересечение любых двух таких множеств содержит не более одного элемента. Формула включений и исключений позволяет в таком случае оценить число эле-

ментов в объединении: оно не меньше суммы размеров всех множеств (эта сумма есть общее число рёбер графа) минус число пар множеств (которое не больше l^2). С другой стороны, объединение содержит не более L элементов.

Отсюда мы заключаем, что общее число рёбер графа не превосходит $L + l^2$, а их плотность не больше $1/l + l/L$. Это даёт необходимую оценку при $l \leqslant \sqrt{L}$, когда первый член больше второго. Если же $l \geqslant \sqrt{L}$, то мы получаем оценку $O(l/L)$, что недостаточно (мы хотим $O(1/\sqrt{L})$). Чтобы получить требуемое, заметим, что мы можем оставить от графа лишь $\sqrt{L}$ вершин в левой доле, для которых число соседей максимально. От этого его плотность лишь возрастёт, а после этого наше рассуждение даёт необходимую оценку. Лемма доказана.

Теперь построим граф без четырёхугольников и установим с помощью только что доказанной леммы, что он плохо покрывается прямоугольниками.

Такой граф проще всего построить геометрически. Рассмотрим произвольное конечное поле $\mathbb{F}$ и плоскость (двумерное векторное пространство) над этим полем. Элементами левой доли будут точки плоскости, элементами правой доли — прямые. Четырёхугольников нет, потому что (как учил ещё Евклид) через две различные точки проходит не более одной прямой.

Оценим число вершин и рёбер такого графа. Если поле содержит (примерно) 2^n элементов, то вершин с каждой стороны будет примерно 2^{2n}, а рёбер — примерно 2^{3n} (на каждой прямой примерно 2^n точек, и через каждую точку проходит примерно 2^n прямых). Таким образом, для большинства ребер (x, y) этого графа сложности $KS(x)$ и $KS(y)$ близки к $2n$, сложность пары $KS(x, y)$ близка к $3n$, а $I(x:y)$ примерно равно n.

308 Покажите, что $I(x:y) = n + O(\log n)$ для всех рёбер этого графа, сложность которых больше $3n - O(\log n)$.

Чтобы дать новое доказательство теоремы 223, посмотрим, какую часть построенного графа (множества пар) можно покрыть $2^{1,1n}$ прямоугольниками размера $2^{1,1n} \times 2^{1,1n}$ (мы используем те же) значения параметров α, β и γ, что и в теореме 223). К каждому такому прямоугольнику можно применить доказанную лемму, по которой плотность рёбер в нём не превосходит $2^{-0,55n}$, так что общее число рёбер, покрытых всеми прямоугольниками, не больше

$$2^{1,1n} \times 2^{1,1n} \times 2^{1,1n} \times 2^{-0,55n} = 2^{2,75n} \ll 2^{3n}.$$

Поэтому подавляющее большинство рёбер не покрыто (а также имеет нужные сложности, как мы говорили) и любое из таких рёбер является примером, существование которого утверждает теорема 223. Таким образом, мы (следуя Ан. А. Мучнику) получаем новое доказательство этой теоремы.

309 Покажите, что любое ребро (x, y), сложность которого близка к $3n$, не допускает выделение общей информации (в смысле теоремы 223). [Указание: множество покрытых рёбер перечисляется простым алгоритмом.]

В этом построении (если рассматривать его как альтернативное доказательство теоремы 223) есть один тонкий момент: нам нужно иметь поле из 2^n элементов (или хотя бы из близкого числа элементов). Поле из 2^n элементов может быть построено

как расширение степени n поля из двух элементов (поле разложения многочлена $x^{2^n} - x$); кроме того, можно воспользоваться известным из теории чисел фактом, что между m и $2m$ всегда есть простое число («постулат Бертрана», который для бо́льших m следует, например, из теоремы о плотности распределения простых чисел).

Тем самым пара слов, из которых не выделяется общая информация, приобретает конкретные очертания: первое из них — это точка конечной плоскости над конечным полем, а второе — проходящая через неё прямая, причём берётся случайная из таких пар.

Ради симметрии можно было бы перейти к проективной плоскости (что мало влияет на сложность, так как бесконечно удалённые точки составляют небольшую долю). Или можно говорить о случайной паре ортогональных (относительно скалярного произведения произведения $x_1y_1 + x_2y_2 + x_3y_3$) прямых в трёхмерном векторном пространстве над конечным полем.

Было бы интересно построить аналогичный пример на основе точек на вещественной сфере (прямых в трёхмерном пространстве). При естественном подходе (дискретизировать с какой-то точностью) возникает проблема: когда x и y близки, то им сразу много точек почти ортогональны, и надо это как-то оценивать.

Как и раньше, наугад взятый коэффициент (1,1) не оптимален. Ровно то же самое рассуждение можно проводить для любых α, β, γ, и оно окончится успешно (то есть число покрытых рёбер будет меньше $3n$), если числа α, β, γ не слишком велики. Применяя лемму (ровно в той форме, как она сформулирована), мы получаем, что множество $C(x, y)$ для нашего примера (случайной пары точка – прямая) содержится (с точностью до $O(\log n)$) в множестве S, граница которого изображена (в виде линий уровня) на рисунке 11.3. (Линия уровня для $\alpha = 3n$ состоит из начала координат.)

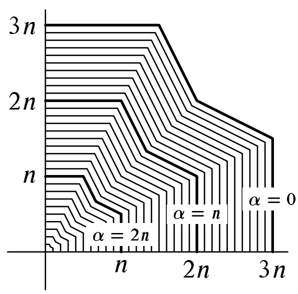


Рис. 11.3. Множество S.

При $\beta \leqslant \gamma$ множество состоит из всех решений неравенства

$$\alpha + \gamma/2 + \max\{\gamma/2, \beta\} \geqslant 3n,$$

а при $\gamma \leqslant \beta$ из симметричного ему множества решений неравенства

$$\alpha + \beta/2 + \max\{\beta/2, \gamma\} \geqslant 3n.$$

Это множество можно ещё пересечь с C_M (поскольку $C(x,y) \subset C_M$ для любой пары (x,y)); получатся более замысловатые линии уровня, которые мы рисовать не будем.

Заметим, что в отличие от ранее приведённых примеров, это лишь верхняя оценка для множества $C(x,y)$; каково это множество в точности, мы не знаем. При исследовании этого вопроса надо иметь в виду следующее. Может оказаться, что множество $C(x,y)$ зависит от выбора поля $\mathbb{F}$. Это кажется вполне правдоподобным, как показывает следующая задача.

310 Пусть мощность поля $\mathbb{F}$ есть полный квадрат: $|\mathbb{F}| = q^2$ (тогда q — степень простого числа). Пусть u — случайная прямая в $\mathbb{F}^2$, а v — случайная точка на прямой u. Докажите, что $C(u,v)$ содержит тройку $(1{,}5n; n; n)$, где $n = \log|\mathbb{F}| = 2\log q$, лежащую на границе множества S на рис. 11.3 (игнорируя логарифмические добавки). (А для произвольного поля мощности примерно 2^n доказать это не удаётся.)

[Указание. Воспользуйтесь тем, что существует подполе $\mathbb{G}$ поля $\mathbb{F}$ мощности q. По этой причине каждый элемент из $\mathbb{F}$ может быть представлен в виде $t + s\alpha$, где $t, s \in \mathbb{G}$, а α — некоторый фиксированный элемент $\mathbb{F}$ (называемый примитивным элементом $\mathbb{F}$ над $\mathbb{G}$). Пользуясь этим, можно множество всех пар (прямая, точка на прямой) разбить на q^3 множеств, каждое из которых состоит из q^3 пар. При этом каждое из этих множеств содержит не более q^2 различных точек и не более q^2 различных прямых. Для построения такого разбиения будем задавать каждую прямую уравнением $y = kx + b$ (игнорируя прямые, не задающиеся таким уравнением — все они не случайны) и представим коэффициенты этого уравнения в виде $k = f + r\alpha$ и $b = h + s\alpha$, где $f, r, h, s \in \mathbb{G}$. А абсциссу x каждой точки (x,y) на этой прямой представим в виде $x = g + t\alpha$ (откуда нетрудно найти, что $y = fg + h + (ft + gr + s)\alpha + rt\alpha^2$). Фиксируя r, t, s (что можно сделать как раз q^3 способами), мы получаем разбиение множества всех пар (прямая, точка на ней). При этом каждое из разбивающих множеств содержит q^2 прямых (что хорошо) и q^3 точек (что плохо). Сократить количество точек можно с помощью следующего трюка. Будем представлять коэффициенты каждой прямой в виде $k = f + r\alpha$, $b = h + (s - ft)\alpha$, а точки на ней в виде $(g + t\alpha, fg + h + (gr + s)\alpha + rt\alpha^2)$ (где по-прежнему все коэффициенты f, g, h, s, r, t принадлежат $\mathbb{G}$). Фиксируя r, t, s, мы получаем уже множество, содержащее как q^2 прямых (однозначно задаваемых парами (f,h)), так и q^2 точек (однозначно задаваемых парами $(g, fg + h)$).]

311 Докажите, что от выбора случайного ребра (x,y) в графе точек и прямых множество $C(x,y)$ не зависит (с логарифмической точностью): принадлежность тройки $\langle\alpha, \beta, \gamma\rangle$ этому множеству определяется тем, какое максимальное число рёбер может быть в прямоугольнике $2^\beta \times 2^\gamma$. А именно, для попадания в $C(x,y)$ необходимо и достаточно, чтобы это число, умноженное на 2^α, было больше общего числа рёбер (умноженного или делённого на многочлен от n).

[Указание (И. Разенштейн, [129]). Граф транзитивен: для любой пары рёбер существует перестановка точек и перестановка прямых, переводящая граф в себя и совмещающая эти рёбра. Поэтому, если есть прямоугольник с бóльшим числом рёбер, то можно покрыть граф случайными копиями этого прямоугольника, взяв их

количество с полиномиальным запасом; если же все прямоугольники маленькие, то в них не хватит рёбер.]

Как и в предыдущем разделе, из доказанных утверждений следует неравенство, оценивающее безусловную сложность z через условные:

Теорема 227. *Пусть x, y — случайная пара точка – прямая на плоскости над конечным полем из* (*примерно*) 2^n *элементов. Тогда для любого слова z выполняется неравенство*

$$KS(z) \leqslant 2\,KS(z\,|\,x) + 2\,KS(z\,|\,y) + O(\log n).$$

◄ Как и раньше, это сводится к проверке неравенства

$$8n \leqslant 3\alpha + 2\beta + 2\gamma$$

для всех $\langle\alpha, \beta, \gamma\rangle \in C(x, y)$. А для этого достаточно проверить, что это неравенство верно для всех точек множества из $S \cap C_M$. Без потери общности можно считать, что $\beta \leqslant \gamma$. Если при этом β больше половины γ, то искомое неравенство получается сложением удвоенного неравенства $\alpha + \gamma/2 + \beta \geqslant 3n$, входящего в определение S, и неравенства $\alpha + \gamma \geqslant 2n$, входящего в определение C_M. А иначе оно получается сложением неравенства $\alpha + \gamma \geqslant 3n$ из определения S и неравенств $\alpha + \beta \geqslant 2n$ и $\alpha + \beta + \gamma \geqslant 3n$ из определения C_M. ►

Итак, у нас есть два разных построения пар с невыделяемой общей информацией. Первое из них даёт лучшие оценки для множества $C(x, y)$; в чём же тогда преимущества второго? Наиболее явное, хотя и неформализуемое преимущество состоит в том, что мы нашли комбинаторную причину невыделяемости общей информации (существование графов, плохо покрываемых прямоугольниками), а также простое достаточное условие для этого (отсутствие четырёхугольников).

Более формально можно сказать, что во втором случае мы построили пример, являющийся «стохастическим»: наша пара является элементом максимальной сложности в некотором простом множестве.

В первом (неконструктивном) доказательстве это, видимо, не так; но можно построить и пример стохастических объектов, достигающих наилучших оценок для множества $C(x, y)$, доказав вероятностно существование графа, который плохо покрывается прямоугольниками (см. [129]).

Ещё один способ объяснить, чем второе доказательство лучше — это рассмотреть сложность с оракулом. Фактически мы доказали, что для любого оракула A существуют слова x, y сложности (обычной, без оракула) $2n$ с общей информацией n, для которых не существует слова z с $KS^A(z) < 1{,}1n$, $KS^A(x\,|\,z) < 1{,}1n$ и $KS^A(y\,|\,z) < 1{,}1n$. В самом деле, для любого оракула получаются также прямоугольники, и потому они покрывают лишь небольшую часть множества пар, и в качестве x, y нужно взять какую-либо непокрытую (и не являющуюся простой) пару. (Естественно, результат будет зависеть от оракула, поскольку любая пара становится простой при подходящем оракуле.)

$\boxed{\mathbf{312}}$ Сформулируйте аналогичное утверждение, используя вместо оракула дополнительное условие u произвольной сложности.

Вопрос о том, что происходит с общей информацией и её выделением при добавлении оракула, рассмотрен в [119]. При этом предполагается, что оракул независим с парой слов (иначе и взаимная информация может измениться). Ясно, что если информация выделялась без оракула, то общая информация сохранится и при добавлении оракула. Оказывается, что верно и обратное (правда, доказать это удаётся не с логарифмической точностью, как хотелось бы, а с худшей).

Комбинаторные оценки размеров множеств в графе применимы и к другим алгебраическим конструкциям. Можно, например, рассмотреть пару случайных перпендикулярных одномерных подпространств в четырёхмерном пространстве или (что то же самое с точностью до бесконечно удалённых точек) случайную пару (точка трёхмерного пространства, проходящая через неё плоскость). Про возникающий граф, правда, уже нельзя утверждать, что в нём нет четырёхугольников: для любых двух одномерных подпространств есть двумерное подпространство, им ортогональное, и любое его одномерное подпространство будет соединено с обоими подпространствами левой доли. Тем не менее, таких одномерных подпространств довольно мало, и можно применить аналогичную оценку с формулой включений и исключений (подробнее это рассуждение проведено в доказательстве теоремы 5 в [28]).

Существует множество аналогичных примеров. Например, можно рассмотреть пару пересекающихся прямых в трёхмерном пространстве над конечным полем. Ещё одна серия примеров: фиксируем число $n \geqslant 3$ и числа k, l, для которых $0 < k < l < n$. Теперь рассмотрим в n-мерном пространстве над конечным полем случайную пару вида (k-мерное подпространство, содержащее его l-мерное подпространство). Как показал А. Ромащенко в [131], в этом случае выделение общей информации также невозможно. Доказательство основано на том, что после нескольких шагов блуждания в возникающем двудольном графе (слева стоят k-мерные подпространства, справа l-мерные, рёбра соответствуют отношению включения) распределение вероятностей приближается к равномерному.

11.4. Условная независимость и общая информация

Существует ещё один способ построения пары слов, из которых не выделяется общая информация, довольно загадочный [94, 131]. Для начала вспомним неравенство задачи 295 (с. 382):

$$H(\xi) \leqslant H(\xi \mid \alpha) + H(\xi \mid \beta) + I(\alpha : \beta),$$

а точнее, соответствующее неравенство для колмогоровских сложностей:

$$KS(z) \leqslant KS(z \mid x) + KS(z \mid y) + I(x : y)$$

(логарифмические члены мы опускаем для краткости). В случае, когда $I(x:y) = 0$, из него следует оценка сложности через условные сложности:

$$KS(z) \leqslant KS(z \mid x) + KS(z \mid y),$$

что не удивительно: если взаимной информации нет, так она и не выделяется. Казалось бы, это ничего дать не может. Но оказывается, что аналогичную оценку можно получить и в случае, когда x и y *условно независимы относительно двух независимых слов*, то есть существуют слова u и v, для которых $I(x:y\,|\,u)=0$, $I(x:y\,|\,v)=0$ и $I(u:v)=0$. А именно, имеет место такое неравенство:

Теорема 228.

$$KS(z)\leqslant 2\,KS(z\,|\,x)+2\,KS(z\,|\,y)+I(x:y\,|\,u)+I(x:y\,|\,v)+I(u:v)$$

для произвольных слов x,y,z,u,v с точностью $O(\log KS(x,y,u,z,v))$.

Заметим, что это неравенство можно было бы вывести из предыдущего и из неравенства Инглетона

$$I(x:y)\leqslant I(x:y\,|\,u)+I(x:y\,|\,v)+I(u:v),$$

но беда в том, что для произвольных слов неравенство Инглетона может не выполняться. Поэтому придётся действовать в другом порядке.

◄ Начнём с уже надоевшего неравенства

$$KS(z)\leqslant KS(z\,|\,u)+KS(z\,|\,v)+I(u:v)$$

и оценим $KS(z\,|\,u)$ и $KS(z\,|\,v)$, пользуясь релятивизованными аналогами того же самого неравенства:

$$KS(z\,|\,u)\leqslant KS(z\,|\,x,u)+KS(z\,|\,y,u)+I(x:y\,|\,u),$$
$$KS(z\,|\,v)\leqslant KS(z\,|\,x,v)+KS(z\,|\,y,v)+I(x:y\,|\,v).$$

После этого остаётся лишь заметить, что добавление нового условия лишь уменьшает сложность: $KS(z\,|\,x,u)\leqslant KS(z\,|\,x)$. ►

Эту теорему можно применить для построения слов с плохо выделяемой общей информацией. Возьмём условно независимые, но не независимые случайные величины (теорема 217, с. 384): пусть величины α и β независимы при известных γ или δ, причём γ и δ независимы.

Искомые слова можно получить в результате последовательности независимых испытаний величин α,β: с большой вероятностью значения α и значения β образуют два слова, из которых не выделяется общая информация. Чтобы доказать это, нам достаточно представить себе, что испытания проводятся параллельно для γ и δ и применить уже известные нам результаты к четырём возникающим словам.

Технически, чтобы получить лучшую оценку погрешности (логарифм вместо квадратного корня, возникающего в сравнении вероятности с частотой), следует говорить не о независимых испытаниях, а о случайных словах с фиксированными частотами. Вот как это делается.

Возьмём четвёрку слов x,y,u,v длины N, частоты в которой совпадают с вероятностями для величин $\langle\alpha,\beta,\gamma,\delta\rangle$. Это значит, что число позиций $i=1,\ldots,N$, в которых в словах x,y,u,v стоят буквы a,b,c,d соответственно, равняется

$$\Pr[\alpha=a,\beta=b,\gamma=c,\delta=d]\cdot N+O(1)$$

(нам пришлось добавить $O(1)$ для округления, поскольку произведение вероятности на длину N может не быть целым).

Как мы знаем из раздела 7.3 (теорема 146), для большинства слов с указанными частотами сложности самих этих слов и их комбинаций равны энтропиям соответствующих случайных величин (умноженным на N) с точностью $O(\log N)$. Таким образом, мы получаем четвёрку слов x, y, u, v, для которых

$$I(x:y\,|\,u) = O(\log N),\quad I(x:y\,|\,v) = O(\log N),\quad I(u:v) = O(\log N),$$

и при этом

$$I(x:y) = NI(\alpha:\beta) + O(\log N),$$

причём, напомним, $I(\alpha:\beta) \neq 0$. Остаётся воспользоваться теоремой 228 и заключить, что для любого z выполняется неравенство

$$KS(z) \leqslant 2\,KS(z\,|\,x) + 2\,KS(z\,|\,y)$$

с точностью до $O(\log N)$, в то время как сложности слов x и y и их взаимная информация с той же точностью пропорциональны N (с ненулевыми коэффициентами пропорциональности)

Таким образом, мы получаем новую конструкцию слов с невыделяемой взаимной информацией. Эти слова, как и в нашем геометрическом примере, являются стохастическими, и удивительным образом никаких комбинаторных оценок нам делать не пришлось.

$\boxed{\textbf{313}}$ Мы уже знаем (теорема 217, с. 384), что существуют условно независимые случайные величины α и β, каждая из которых равномерно распределена на $\{0,1\}$, и которые оказываются равны друг другу с вероятностью $5/8$. Покажите, что это утверждение можно обобщить: для любого c ($3/8 \leqslant c \leqslant 5/8$) можно построить пару условно независимых случайных величин, каждая из которых равномерно распределена на $\{0,1\}$, которые принимают равные значения с вероятностью c.

Если c слишком далеко от $1/2$, то аналогичное утверждение уже неверно. Тем не менее, для любых $0 < c < 1$ можно доказать свойство невыделяемости общей информации у слов, полученных независимыми испытаниями случайных величин, равных с вероятностью c. Мы разделим это утверждение на две задачи:

$\boxed{\textbf{314}}$ Пусть случайные величины α и β равномерно распределена на $\{0,1\}$ и принимают одинаковые значения с некоторой вероятностью $c \in (0,1)$. Докажите, что существуют конечные цепочки случайных величин $\alpha_1, \ldots, \alpha_k$ и $\beta_1, \ldots, \beta_k$, для которых α и β независимы относительно α_1 и относительно β_1; величины α_1 и β_1 в свою очередь независимы относительно α_2 и относительно β_2, которые независимы относительно α_3 и β_3,..., которые независимы относительно α_k и относительно β_k, которые уже независимы безо всяких условий. [Указание. Если утверждение истинно для некоторого $c \geqslant 1/2$, то оно истинно и для $c' = (c^2+1)/2$. Чтобы установить это, нужно применить конструкцию из доказательства теоремы 217; в качестве случайных величин γ и δ надо взять случайные величины, для

которых утверждение выполнено при данном c. При этом условные распределения пары α, β при известных исходах $\gamma = 0$, $\delta = 1$ и $\gamma = 1, \delta = 0$ нужно подправить так, чтобы случайные величины α и β стали независимыми при известном γ и при известном δ. Наконец, нетрудно убедиться, что любое число от $1/2$ до 1 может быть получено из некоторого числа от $1/2$ до $5/8$ конечным числом применений функции $c \mapsto (c^2 + 1)/2$. Для доказательства утверждения для $c < 1/2$ можно инвертировать одну из случайных величин α, β.]

315 Пусть случайные величины α и β удовлетворяют утверждению задачи 314. Возьмём максимально сложную пару двоичных слов x, y длины n, у которых все частоты совпадают с соответствующими вероятностями α и β. Докажите, что у этих слов x и y не выделяется общая информация. Более точно, для любого слова z выполнено неравенство $K(z) \leqslant O(K(z\,|\,x) + K(z\,|\,y) + \log n)$ (константа в $O(\cdot)$ может зависеть от длины цепочки условно независимых случайных величин).

Решение задачи 314 можно обобщить на алфавиты большего размера и на неравномерные распределения вероятностей. Существует простой критерий, показывающий, когда случайные величины α, β удовлетворяют утверждению задачи 314. Это оказывается возможным тогда и только тогда, когда в матрице совместного распределения α и β нельзя так переставить строки и столбцы, чтобы получилась блочная матрица вида

$$\left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array}\right)$$

(в обоих недиагональных блоках все элементы равны нулю). В этом случае общая информация (из типичных слов, получаемых при n испытаниях α и β) не выделяется. А если блочное представление возможно, то некоторая часть информации (указывающая, в какой из блоков мы каждый раз попадаем) выделяется. Это рассуждение [94] даёт новое доказательства критерия, впервые полученного Гачем и Кёрнером [47].

12. Алгоритмическая теория информации для нескольких источников

12.1. Постановка задачи о передаче информации

Рассмотрим ориентированный граф, рёбра которого играют роль «линий связи», а вершины — узлов обработки информации. В некоторые вершины извне поступает какая-то информация; в результате обработки информации (в вершинах) и её передачи (по рёбрам) некоторая другая информация должна попасть в заданные вершины. Такая задача традиционно рассматривается в шенноновской теории информации с несколькими источниками; здесь мы хотим проанализировать её с точки зрения алгоритмической теории информации, следуя [155].

Более формально. Пусть задан некоторый конечный ориентированный ациклический граф. Некоторые его вершины объявлены *входами*, и для каждого входа указано *входное* слово. Другие его вершины объявлены *выходами*, и для каждой из них указано *выходное* слово. Возникает задача: на каждом ребре графа написать некоторое двоичное слово (*передаваемое* по данному ребру), причём для каждой вершины v должно выполняться такое условие:

> если X — любое выходящее из вершины v слово (указанное изначально выходное слово или передаваемое по исходящим из v рёбрам), а $Y_1, \ldots, Y_k$ — все входящие в эту вершину слова (заданные изначально входные слова или передаваемые по входящим в v рёбрам), то
>
> $$KS(X \mid Y_1, \ldots, Y_k) \approx 0.$$

Неформально говоря, это условие означает, что в вершине v происходит лишь переработка информации, а никакой новой информации не возникает.

Приближённое равенство нулю, как обычно, означает, что указанная сложность есть $O(\log N)$, где N — суммарная длина всех входящих и выходящих слов, указанных в формулировке задачи. Тем самым мы фактически рассматриваем не одну такую задачу передачи информации, а семейство задач, соответствующих разным значениям N.

При этом мы будем накладывать ограничения на «пропускные способности» рёбер, то есть длины передаваемых по ним слов. Такие ограничения будут накладываться лишь на некоторые рёбра; пропускная способность остальных не ограничивается. Нас будет интересовать, возможно ли решить задачу при указанных ограничениях.

Начнём с самого простого примера: пусть имеется граф, состоящий из двух вершин и соединяющего их ребра (рис. 12.1; договоримся сразу, что на наших рисунках все рёбра ведут сверху вниз)

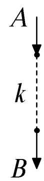


Рис. 12.1. Простейшая задача передачи информации.

В верхнюю вершину входит слово A, а из нижней должно выйти слово B, при этом соединяющий эти вершины канал связи имеет пропускную способность k. Другими словами, мы для данных слов A и B ищем такое слово X, что

$$KS(X \mid A) \approx 0, \quad KS(B \mid X) \approx 0, \quad l(X) \leqslant k.$$

Ясно, что это возможно только при $KS(B \mid A) \approx 0$ и $KS(B) \leqslant k$ (последнее неравенство также понимается «с точностью до логарифма»); с другой стороны, эти необходимые условия являются также и достаточными, так как в качестве X можно взять кратчайшее описание для B.

Более формально это можно сказать так: пусть имеется последовательность слов A_n и B_n, а также последовательность натуральных чисел k_n, причём длины слов A_n и B_n, а также числа k_n, ограничены сверху некоторым многочленом от n. Тогда следующие два свойства равносильны:

(1) существует последовательность слов X_n, для которых $l(X_n) \leqslant k_n + O(\log n)$, а также $KS(X_n \mid A_n) = O(\log n)$ и $KS(B_n \mid X_n) = O(\log n)$;

(2) $KS(B_n \mid A_n) = O(\log n)$ и $KS(B_n) \leqslant k_n + O(\log n)$.

Эквивалентность (1) и (2) вытекает из того, что

$$KS(B \mid A) \leqslant KS(B \mid X) + KS(X \mid A) + O(\log KS(A, B, X)),$$
$$KS(B) \leqslant l(X) + KS(B \mid X) + O(\log KS(B \mid X))$$

при любых A, B, X (отсюда следует импликация $(1) \Rightarrow (2)$), а также того, что

> для любых A, B, k найдётся слово X, при котором $l(X) \leqslant KS(B)$, $KS(X \mid A) \leqslant KS(B \mid A) + O(\log KS(B))$ и $KS(B \mid X) = O(1)$,

что даёт импликацию $(2) \Rightarrow (1)$; во втором случае в качестве X надо взять кратчайшее описание слова B.

При $A = B$ наше утверждение приобретает совсем простой вид: передача слова A по каналу связи с пропускной способностью k возможна, когда сложность слова A не превосходит k.

Перейдём к более содержательным примерам.

12.2. Условное кодирование

Следующая задача может быть названа «задачей передачи слова A при известном слове B» (рис. 12.2). В ней требуется закодировать слово A не более чем k битами, передать и раскодировать обратно; при этом и при кодировании, и при декодировании известно слово B (пропускная способность рёбер, изображённых на рисунке сплошными линиями, не ограничивается, так что по ним можно передать слово B полностью).

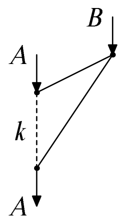


Рис. 12.2. Передача A при известном B.

Эта задача разрешима тогда и только тогда, когда $KS(A\,|\,B) \leqslant k$. В самом деле, в пункте декодирования, помимо слова B (или какого-то слова, из него полученного, и не содержащего новой по сравнению с B информации) известны k битов, поэтому раскодирование возможно, лишь если $KS(A\,|\,B) \leqslant k$. С другой стороны, если $KS(A\,|\,B) \leqslant k$, то в качестве слова X, передаваемого по каналу ограниченной пропускной способности, надо взять кратчайшее описание B при известном A, а по двум другим каналам передавать слово B. Заметим, что сложность кратчайшего описания относительно пары $\langle A, B\rangle$ логарифмическая, так как зная длину этого описания, можно параллельно пробовать все слова этой длины, пока не найдётся одно из кратчайших описаний.

$\boxed{\textbf{316}}$ В этом рассуждении достаточно одного кратчайшего описания; покажите, тем не менее, что *все* кратчайшие описания A при известном B имеют относительно пары $\langle A, B\rangle$ сложность, не превосходящую $O(\log KS(A, B))$.

$\boxed{\textbf{317}}$ Дайте точную формулировку приведённого только что критерия возможности передачи слова A при известном B по аналогии с разобранным ранее примером, рассмотрев последовательности A_n, B_n и k_n, и докажите его.

12.3. Условное кодирование: теорема Мучника

Следующий замечательный результат Ан. А. Мучника [116] является аналогом теоремы Вольфа – Слепяна из шенноновской теории информации [159] и говорит, что в предыдущем примере не обязательно использовать слово B при кодировании. Рассмотрим граф, получающийся из предыдущего удалением одного ребра (рис. 12.3).

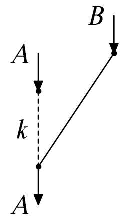


Рис. 12.3. Передача A при известном B: теорема Мучника.

Оказывается, что эта задача разрешима при тех же условиях, что и предыдущая, то есть при $KS(A \mid B) \leqslant k$. Необходимость этих условий очевидна (граф стал только меньше); вот точная формулировка утверждения о достаточности.

Теорема 229. *Пусть A и B — произвольные слова сложности не более n. Тогда найдётся слово X длины не более $KS(A \mid B) + O(\log n)$, при котором $KS(X \mid A) = O(\log n)$ и $KS(A \mid B, X) = O(\log n)$.*

Имеется в виду, что скрытая в $O(\log n)$ константа не зависит от n и от выбора слов A и B.

Теорему Мучника можно переформулировать так: для любых слов A и B существует программа, переводящая B в A, имеющая логарифмическую сложность относительно A и безусловную сложность $KS(A \mid B)$ (плюс логарифмическая добавка). Другими словами, дополнительное ограничение, состоящее в простоте программы относительно A, приводит к росту (безусловной) сложности программы не более чем на $O(\log KS(A, B))$.

◀ Пусть слово A имеет сложность a. Заменим это слово на его (кратчайшее) описание длины a. От этого величины $KS(A \mid B)$, $KS(X \mid A)$ и $KS(A \mid B, X)$ изменятся не более чем на $O(\log n)$, поэтому мы можем далее предполагать, что слово A имеет длину (а не сложность) a.

Пусть сложность $KS(A \mid B)$ равна m. В самом первом приближении идею доказательства можно объяснить так. Рассмотрим хеш-функцию $\chi\colon \mathbb{B}^a \to \mathbb{B}^m$, которая каждому слову длины a ставит в соответствие хеш-значение (отпечаток, fingerprint) длины m.

Для данного слова B имеется 2^m слов Z длины a, для которых $KS(Z \mid B) \leqslant m$ (на самом деле их $O(2^m)$, но мы пока что пренебрегаем такими мелочами). Пусть $S_B \subset \mathbb{B}^a$ — множество этих слов. Слово A — один из элементов множества S_B.

Пусть хеширование прошло исключительно удачно в том смысле, что у всех слов из S_B хеш-значения оказались различными. Тогда любое слово $P \in S_B$ может быть однозначно восстановлено по $\chi(P)$ при известном B (и χ). Поэтому можно в качестве X, существование которого утверждается в теореме, взять $\chi(A)$: оно имеет правильную длину, просто относительно A (если функция χ проста) и вместе с B позволяет восстановить A (перечисляем множество S_B, пока не обнаружится слово с правильным хеш-значением).

Разумеется, так просто всё не выходит. Какова бы ни была хеш-функция χ, при $a > m$ найдётся как минимум 2^{a-m} слов, имеющих одно и то же хеш-значение (а при простой χ можно найти много простых слов с одним хеш-значением, и они будут просты относительно B при любом B), так что хеширование не будет удачным.

Поэтому мы модифицируем этот план и будем сопоставлять с каждым словом $Z \in \mathbb{B}^a$ не одно хеш-значение, а несколько (в количестве $\mathrm{poly}(n)$). Таким образом, вместо хеш-функции мы рассматриваем двудольный граф $E \subset \mathbb{B}^a \times \mathbb{B}^m$, в котором каждая вершина Z в левой доле (из $\mathbb{B}^a$) имеет $\mathrm{poly}(n)$ соседей в правой доле. Этих соседей мы будем называть *отпечатками* вершины x.

Доказывая теорему, мы будем искать слово X среди отпечатков слова A. Это гарантирует нам, что $KS(X \mid A) = O(\log n)$, если сам граф E прост (имеет сложность $O(\log n)$). В самом деле, для задания X при известном A достаточно указать порядковый номер X среди отпечатков слова A.

Если для данного $A \in S_B$ найдётся отпечаток X, уникальный для A внутри S_B (это значит, что в S_B нет других слов, имеющих X своим отпечатком), то слово A можно восстановить, перечисляя S_B и ожидая появления слова, имеющего X среди своих отпечатков. В этом случае $KS(A \mid B, X) = O(\log n)$ (мы, как и раньше, предполагаем, что граф E имеет сложность $O(\log n)$).

Более того, это верно и в случае, когда в S_B имеется $\mathrm{poly}(n)$ слов с отпечатком X, нужно лишь дополнительно указать порядковый номер A среди этих слов (в порядке перечисления множества S_B), что потребует дополнительных $O(\log n)$ битов.

Таким образом, нам хотелось бы следующего: *у слова A имеется правый сосед, у которого мало* ($\mathrm{poly}(n)$) *левых соседей.* Говоря о соседях, мы рассматриваем сужение графа E на S_B как двудольный граф с левой долей S_B и правой долей $\mathbb{B}^m$. Отметим, что в этом уменьшенном графе число вершин слева и справа примерно одинаково — порядка 2^m.

Это же можно переформулировать так: будем говорить, что вершина справа плоха, если у неё много левых соседей, а вершина слева плоха, если все её соседи справа плохи. Нам хотелось бы, чтобы интересующая нас вершина A не оказалась плохой.

Дальнейший план действий таков: справа плохих вершин мало, так как в графе мало рёбер (число рёбер оцениваем как число вершин слева, умноженное на максимальную степень левых вершин, которая есть $\mathrm{poly}(n)$), а каждая плохая вершина справа даёт большой вклад в число рёбер. Так мы добьёмся, чтобы доля плохих вершин справа была не больше $1/p(n)$ для заданного полинома p (при этом придётся полиномиально увеличить границу, при превышении которой вершина объявляется плохой).

Отсюда можно заключить, что и слева плохих вершин мало, если граф E обладает следующим свойством типа экспандера: *для всякого множества T в левой доле мощности не более 2^{m-1} множество $E(T)$ всех соседей всех элементов множества T содержит не меньше элементов, чем само множество T.* В самом деле, возьмём в качестве T множество плохих вершин слева. Все их правые соседи будут плохими, поэтому слева не больше плохих вершин, чем справа. (Заметим, что указанное свойство графа E сохраняется при его сужении на S_B.)

Остаётся объяснить, откуда мы возьмём граф E с нужными свойствами и что делать, если нужное нам слово A попадёт в (пусть и небольшое) число плохих вершин слева.

Как хорошо знают специалисты, существование экспандеров несложно доказать вероятностно (случайный граф обладает нужным свойством с положительной вероятностью); явное их построение — гораздо более сложное и интересное дело, и в последние десятилетия тут есть большие продвижения. Мы, однако, можем обойтись без сложных конструкций с помощью простого трюка. Нужное нам свойство проверяется алгоритмически (за очень большое, но конечное время). Поэтому можно перебирать все графы в каком-либо естественном порядке, пока не обнаружится первый подходящий. Он будет иметь логарифмическую сложность (чтобы организовать перебор, нам достаточно знать размеры множеств). Таким образом из существования *какого-нибудь* графа с нужным свойством автоматически следует существование *простого* графа с тем же свойством.

Последнее: может ли слово A оказаться плохим? Заметим, что плохие слова справа можно алгоритмически перечислять при известных B (и E); по мере обнаружения новых элементов в S_B множество плохих слов справа растёт. Поскольку граф E мы считаем известным, можно перечислять и плохие слова слева. Таким образом, каждое плохое слово можно задать (при известном B) его порядковым номером, что меньше m битов, так как плохих слов заметно меньше 2^m. (Ещё нужно $O(\log n)$ битов, чтобы задать m, a и E, но и с учётом этой добавки получится меньше m, как мы увидим.) А мы с самого начала предполагали, что сложность A при известном B равна m (так выбиралось число m). Поэтому A не может оказаться плохим.

Итак, мы описали схему доказательства теоремы «сверху вниз». Теперь мы разберём более подробно отдельные шаги этого доказательства, двигаясь «снизу вверх». Начнём с существования графов со свойствами типа экспандера.

Лемма. *Пусть a и m — натуральные числа, причём $a \geqslant m$. Тогда существует двудольный граф $E \subset \mathbb{B}^a \times \mathbb{B}^m$, в котором степень вершин в левой доле ($\mathbb{B}^a$) не превосходит $a + m + 2$, обладающий следующим свойством: для любого множества $T \subset \mathbb{B}^a$, содержащего не более 2^{m-1} элементов, множество $E(T)$ всех правых соседей всех элементов из T содержит больше элементов, чем само T.*

Доказательство. Докажем, что случайный граф обладает указанным свойством с положительной вероятностью. Говоря о случайном графе, мы имеем в виду, что для каждой точки в левой доле случайно выбираются $a + m + 2$ соседей, равномерно распределённых в правой доле. При этом этот выбор делается независимо (для разных точек и для разных соседей одной точки).

Если указанное в лемме свойство не выполняется, то существует некоторое (непустое) множество T в левой доле и некоторое множество U в правой доле, для которых $|U| = |T|$, но все соседи элементов из T принадлежат U. Мы подсчитаем вероятность такого события для данных T и U и покажем, что сумма этих вероятностей по всем T и U меньше единицы.

Пусть фиксированы T и U; обозначим число элементов в них через t. По предположению, $t \leqslant 2^{m-1}$, поэтому вероятность того, что случайный элемент правой доли попадёт в U, не больше $1/2$. Вероятность того, что это случится $t(a + m + 2)$

раз при независимых испытаниях (по $a + m + 2$ испытаний для t элементов множества T) не превосходит $2^{-t(a+m+2)}$.

Просуммируем вероятность сначала по всем парам множеств T и U размера t. Число различных T не превосходит $(2^a)^t$ (выбираем t раз один из 2^a элементов; совпадения и возможность перестановки лишь уменьшают число вариантов), число различных U не превосходит $(2^m)^t$. Таким образом, сумма вероятностей по множествам размера t не превосходит

$$2^{mt} \cdot 2^{at} \cdot 2^{-t(m+a+2)} = (1/4)^t.$$

Остаётся заметить, что сумма ряда $\sum(1/4)^t$ (по всем $t \geqslant 1$) меньше 1 (равна 1/3). Лемма доказана.

Пусть $E_{m,a}$ — первый в каком-нибудь естественном порядке граф, удовлетворяющий лемме (для данных m и a). Его сложность не превосходит $2\log a + O(1)$ (достаточно указать числа a и m, отведя для них первую и вторую половины описания длины $2\log a$).

Для данного слова B и для данных m и a рассмотрим множество S_B слов длины a, имеющих относительно B сложность не более m. В нём не более 2^{m+1} слов. Рассмотрим ограничение графа $E_{m,a}$ на S_B. В этом двудольном графе не более $2^{m+1} \cdot (a + m + 2) \leqslant a2^{m+3}$ рёбер (каждая из не более чем 2^{m+1} вершин имеет не более чем $a + m + 2$ соседей). Объявим *плохими* вершины в правой доле, имеющие не менее a^4 соседей слева. Тогда число плохих вершин будет не более $k = 2^{m+3}/a^3$.

Слева мы объявим *плохими* вершины (слова из S_B), у которых все соседи справа плохие. По построению графа $E_{m,a}$ число плохих вершин слева также не превосходит $k = 2^{m+3}/a^3$. В самом деле, мы можем предполагать, что a настолько велико, что $k < 2^{m-1}$. Если бы нашлось $k + 1$ плохих вершин, то по свойству графа количество соседей этих $k + 1$ вершин превосходило бы k и все они были бы плохие. Плохие вершины слева можно перечислять (при известных m, a и B), поэтому плохое слово можно задать его порядковым номером (на что требуется $m - 3\log a + O(1)$ битов), так что

$$KS(P \mid B, m, a) \leqslant m - 3\log a + O(1)$$

для любого плохого слова P. Видно, что запас больше $2\log a$, необходимых для указания m и a, так что все плохие слова имеют условную сложность (относительно B) меньше m и слово A среди них оказаться не может.

Следовательно, у него есть хороший сосед справа. Обозначим его через X. Тогда $KS(X \mid A) \leqslant (3 + \varepsilon)\log a + O(1)$ (надо указать m, a и порядковый номер X среди соседей A в $E_{m,a}$; ε-добавка требуется для кодирования пар). Длина X равна m, то есть $KS(A \mid B)$. Наконец, $KS(A \mid B, X)$ не превосходит $(6 + \varepsilon)\log a$ (из коих $4\log a$ уходит на указание порядкового номера A среди a^4 левых соседей вершины X, а $2\log a$ требуется для указания m и a; ε нужно для кодирования пар).

Теорема Мучника доказана. ►

318 Покажите, что фактически доказано более сильное утверждение, чем объявлено: в условии теоремы требовалось, чтобы слово B имело сложность не более n, но это нигде не использовано.

Напомним, что все неравенства в теореме Мучника верны с точностью $O(\log n)$, где n максимальная из сложностей исходных слов A, B. Можно ли усилить формулировку теоремы, потребовав, чтобы все неравенства были выполнены с точностью $O(\log m)$, где m максимальная из *условных* сложностей $KS(A\,|\,B)$ и $KS(B\,|\,A)$? Оказывается, что нет и это доказано в работе [179, Раздел 5] (мы это доказательство воспроизводить не будем).

12.4. Комбинаторный смысл теоремы Мучника

Многие утверждения о колмогоровской сложности имеют комбинаторный эквивалент — равносильное утверждение чисто комбинаторного характера, в котором о сложности ничего не говорится. Во многих случаях эквивалентом является утверждение о существовании выигрышной стратегии в некоторой игре. (См. об этом для случая общей теории алгоритмов в [113] и в [117] применительно к колмогоровской сложности.)

Для доказанной только что теоремы 229 также имеется комбинаторный эквивалент. Рассмотрим для данных чисел a, b и m (считаем, что $m \leqslant a$) игру двух игроков: Математика (**М**) и его Противника (**П**). Игра имеет также параметр c (соответствующий константе в $O(\log n)$, как мы увидим).

Математик имеет право указать для каждого слова A длины a не более чем $c(a+b)^c$ слов длины m, которые он условно именует *простыми относительно* A. Кроме того, он для каждой пары слов B (длины b) и X (длины m) имеет право указать до $c(a+b)^c$ слов длины a, называя их *простыми относительно пары* B, X.

Противник для каждого слова B длины b может указать до 2^m слов длины a, назвав их m*-простыми относительно* B.

Игрок может сделать очередной ход (объявив простыми ещё какие-то слова) в любой момент (независимо от ходов другого игрока). При этом он видит ходы другого игрока, сделанные к этому моменту. Возникающая игра по существу конечна: поскольку объявленные слова нельзя взять назад, партия достигает некоторого предельного положения. (Однако, наблюдая за игрой, невозможно определить, достигнуто ли это предельное положение или ещё нет — игроки сохраняют право хода до бесконечности, даже если им и не пользуются.) В этом предельном положении определяется победитель:

> **М** выиграл, если для всякого слова B длины b и для всякого слова A длины a, которое **П** объявил m-простым относительно B, найдётся слово X длины m, которое **М** объявил простым относительно A и для которого **М** объявил A простым относительно B и X.

Теперь можно сформулировать комбинаторный эквивалент теоремы 229.

Теорема 230. *Существует такая константа c, что для любых натуральных a, b, m (для которых $m \leqslant a$) в описанной игре с параметрами a, b, m и c выигрывает* **М**.

Объясним, почему это комбинаторное утверждение действительно является комбинаторным эквивалентом теоремы 229. Предположим, что оно верно для некоторого c. Рассмотрим Противника, который делает свои ходы, не обращая внимания на Математика, и объявляет m-простыми слова длины a, которые имеют сложность менее m (относительно соответствующего B). Поведение Противника в этом случае задаётся алгоритмом, для задания которого надо знать значения a, b, m. Выигрышная стратегия Математика (существующая по предположению) может быть найдена перебором (при известных a, b и m). Поэтому слова, которые Математик объявляет простыми, будут и в самом деле иметь малую колмогоровскую сложность (условную) — для их задания, помимо порядкового номера (который записывается $\log c + c\log(a+b)$ битами), достаточно указать a, b, m, что требует ещё $O(\log(a+b))$ битов. Таким образом, мы получаем утверждение теоремы 229. (Техническое пояснение: в оценке $c(a+b)^c$ первый множитель c нужен для малых значений a и b и соответствует слагаемому $O(1)$ в оценках колмогоровской сложности, которое, строго говоря, надо было бы явно добавлять к $O(\log n)$ на случай $n=1$ и $\log n = 0$.)

В обратную сторону: предположим теперь, что утверждение теоремы 229 верно (с некоторой константой в $O(\log n)$, которую мы будем обозначать c') и докажем, что для некоторого достаточно большого c выполнено комбинаторное утверждение. Рассуждая от противного, допустим, что для всех c для некоторых a, b, m Математик проигрывает, а значит Противник имеет выигрышную стратегию. Эту стратегию вместе с тройкой a, b, m можно найти перебором. Поэтому, если эта стратегия объявила некоторое A простым относительно B, то условная сложность A при известном B и в самом деле невелика — она не превышает $m + O(KS(c))$. Мы получим противоречие, дав этой стратегии сыграть против следующей «слепой» стратегией Математика: не обращая внимания на ходы противника называть простыми слова X, для которых $KS(X\,|\,A) < c\log(a+b) + \log c$, а также слова A, для которых $KS(A\,|\,B, X) < c\log(a+b) + \log c$.

Ограничения на количество слов, объявляемых простыми, не нарушены. Поэтому осталось доказать, что эта стратегия Математика обыграет выигрышную стратегию Противника. Пусть даны слова A, B длин a, b для которых Противник объявил A простым относительно B. По теореме 229 существует слово X' длины $KS(A\,|\,B) + c'\log(a+b)$, для которого выполняется утверждение теоремы. Поскольку $KS(A\,|\,B)$ не превосходит $m + O(KS(c))$, длина слова X' ненамного превышает m. Оставим от него первые m битов, полученное слово обозначим X. Сложности $KS(X'\,|\,A)$ и $KS(A\,|\,B, X')$ невелики по условию, а количество отрезанных у X' битов тоже невелико. Поэтому сложности $KS(X\,|\,A)$ и $KS(A\,|\,B, X)$ также малы, и при правильном выборе c меньше $c\log(a+b) + \log c$ (а значит, Математик выигрывает).

Осталось привести оценки. Условная сложность $KS(X\,|\,A)$ не превосходит

$$c'\log(a+b) + O(\log m)$$

(сложность X' относительно A плюс длина префиксного кода m). Условная сложность $KS(A\,|\,B, X)$ не превосходит суммы

$$c'\log(a+b) + O(KS(c))$$

(сложность A относительно пары B, X плюс длина отрезанного конца X'). Видно, что можно подобрать число c вида 2^i так, что обе суммы не превосходили $c\log(a+b)+\log c$ (для чисел $c=2^i$ сложность c, входящая в верхнюю оценку для $KS(A\,|\,B,X)$, значительно меньше $\log c$).

Таким образом, мы установили, что комбинаторное утверждение теоремы 230 действительно эквивалентно сложностному утверждению теоремы 229 (и, в частности, истинно).

На комбинаторный язык можно перевести не только формулировку теоремы 229, но и её доказательство. В ходе доказательства Математик не использует своё право указывать простые относительно A слова по ходу игры, а указывает их все сразу (в соответствии с графом-экспандером). Далее он объявляет слово A, соседнее с X, простым относительно X и B, если к моменту объявления A простым относительно B среди слов, объявленных Противником простыми относительно B, мало слов, соседних с X. В результате ему — для каждого B — удастся обслужить большинство слов, объявленных простыми относительно этого B. Оставшееся меньшинство слов, объявленных простыми относительно B, передаётся на следующий уровень обслуживания, где делается всё то же самое, но с уменьшенным на единицу m. И так далее. В итоге количество слов, объявленных простыми относительно A, вычисляется как сумма убывающей вдвое геометрической прогрессии, и потому вдвое больше первоначально взятого, но это не страшно.

Тот факт, что Математик указывает слова сразу (а не по ходу игры), имеет и некоторые алгоритмические следствия:

$\boxed{\textbf{319}}$ Докажите, что для любого слова A длины n существует слово X длины $KS(A)$, для которого $KS(A\,|\,X)=O(\log n)$ и *тотальная* условная сложность X при известном A (минимальная сложность всюду определённой программы, переводящей A в X) есть $O(\log n)$. Покажите, что заменить обе сложности $KS(A\,|\,X)$ и $KS(X\,|\,A)$ на тотальные нельзя. [Указание. Для первого можно использовать рассуждение Мучника; во втором случае можно использовать существование нестохастических слов (см. главу 14 об алгоритмической статистике).]

Эта задача показывает, что рассуждение Мучника даёт содержательные результаты даже в случае пустого слова B. При непустом B мы можем получить вариант теоремы, в котором $KS(X\,|\,A)$ заменено на тотальную сложность X при известном A (а остальные условные сложности оставлены без изменений); нужно только дополнительно предположить, что длина слова A (а не только его сложность) не превосходит n.

12.5. Отступление: on-line паросочетания

Немного модифицировав комбинаторное доказательство теоремы 229, можно получить доказательство более сильного (и более простого) комбинаторного утверждения.

Рассмотрим двудольный граф. Вершины левой доли образуют множество A, вершины правой доли — множество B, рёбра графа образуют множество $E\subset A\times B$. Задача о паросочетании обычно ставится так: имеется некоторое множество $A'\subset A$,

надо для каждой вершины $a \in A'$ выбрать одного из её соседей в B (то есть вершину из B, соединённую с a ребром), причём выбранные соседи разных вершин должны быть различны.

Рассмотрим более сложный вариант этой задачи, который можно назвать «паросочетанием в режиме on-line»: нам по очереди указывают вершины в левой доле, и мы для них должны указать парную вершину в правой доле (и не имеем возможности впоследствии изменить этот выбор). Будем говорить, что граф $E \subset A \times B$ *допускает on-line паросочетание размера* k, если существует способ, гарантирующий спаривание любых k вершин левой доли (в том порядке, в котором их указывают). Другими словами, рассматривается игра на графе, в которой противник указывает нам по очереди k вершин из A, и после каждого его хода мы обязаны выбрать одного из соседей для только что указанной им вершины (и при этом не повторяться).

Заметим, что это определение несимметрично (мы выбираем вершины справа, а противник — слева). Ещё из определения видно, что свойство «допускать on-line паросочетание данного размера» лежит в PSPACE. (Интересно было бы узнать, есть ли какие-нибудь лучшие верхние оценки сложности.)

Теорема 231. *Для некоторой константы c при любых a и m (где $a \geqslant m$) существует двудольный граф с 2^a вершин в левой доле и $2^m a^c$ вершин в правой доле, в котором каждая вершина слева имеет не более a^c соседей справа и который допускает on-line паросочетание размера 2^m.*

Другими словами, существует граф с заданными размерами долей с небольшой (полиномиальной) степенью вершин левой доли, допускающий on-line паросочетание почти что максимально возможного размера (размер паросочетания ограничен числом вершин справа) — максимального с точностью до полиномиального множителя.

Прежде чем доказывать теорему 231, объясним, как из неё вытекает утверждение теоремы 229. Как и раньше, для начала заменим слово A на его кратчайшее описание длины a. Положим m равным $KS(A\,|\,B)+1$ и применим теорему 231; получится некоторый двудольный граф (2^a вершин слева и $a^c 2^m$ вершин справа, степень левых вершин не более a^c). Раз граф с указанными свойствами существует, то его можно найти перебором. Фиксируем этот граф и алгоритм построения on-line паросочетания. Далее для данного B можно перечислять слова, имеющие сложность меньше m относительно этого B. Их не больше 2^m и среди них есть слово A. По очереди подбираем для них пару в графе. Рассмотрим слово X, парное к слову A. Как и все правые соседи A (которых мало), оно имеет малую сложность относительно A. (Поскольку граф и алгоритм построения on-line паросочетания были найдены перебором, то они задаются параметрами a и m и имеют логарифмическую сложность.) С другой стороны, зная B и X (а также все числовые параметры), можно восстановить процесс построения паросочетания и дождаться, пока слово X будет соединено со некоторым словом, тем самым восстановив A.

◄ Перейдём теперь к доказательству теоремы 231. Заметим, что достаточно доказать более слабое утверждение, разрешив подыскивать пары не для всех 2^m элементов, указываемых в левой доле, а только для половины (по нашему выбору).

В самом деле, если мы умеем это делать, то для необслуженных элементов можно параллельно запустить аналогичный процесс (с уменьшенным на 1 значением m и соответствующим графом, возможно, другим), число необслуженных уменьшится ещё вдвое и так далее. В итоге мы найдём паросочетание в графе, в котором левая доля общая, а правая является объединением правых долей всех использованных графов (для $m, m-1, m-2\ldots$ вплоть до нуля).

Остаётся заметить, что такому (ослабленному) требованию удовлетворяет граф со свойствами экспандера, о котором мы говорили. Алгоритм построения паросочетания будет самым простым: если у вершины есть ещё не использованный сосед, то его и выбираем, а если все соседи уже заняты, то сдаёмся и вершину не обслуживаем. Несложно сообразить, что в результате будет обслужена как минимум половина вершин. В самом деле, если обслужено меньше, то и справа использовано ровно столько же, то есть менее 2^{m-1} вершин. С другой стороны, для каждой необслуженной вершины, каковых существует более 2^{m-1}, все её соседи использованы. (Иначе почему мы её не обслужили?) А это противоречит свойству экспандера: согласно этому свойству множество из 2^{m-1} элементов (и потому любое большее) не может иметь меньше 2^{m-1} соседей. ▶

Естественный вопрос, возникающий в связи с этим доказательством — его соотношение с доказательством теоремы Вольфа – Слепяна. Было бы интересно указать комбинаторный факт, из которого легко следуют как теорема Мучника, так и теорема Вольфа – Слепяна. Интересно было бы также изучить другие конструкции on-line паросочетаний (прямое вероятностное доказательство, не использующее дублирования уровней, или явные конструкции.)

Другое доказательство теоремы Мучника можно получить, используя (хорошо известную) комбинаторную конструкцию, так называемые *экстракторы.* Эта идея была предложена (в другой ситуации и ещё до Мучника) в статье [14], и применена к доказательству теоремы Мучника в [125]. Преимущество такого подхода в том, что можно использовать известные явные конструкции экстракторов и другую технику теории сложности (например, псевдослучайные генераторы, как предложил А. Ромащенко) для доказательства варианта теоремы Мучника, где используется сложность с ограничением на используемую память [123, 124]. (Вообще сложность с ограничением на ресурсы — важная отдельная тема, которую мы в нашей книге не обсуждаем.)

12.6. Относительное кодирование пары слов

В следующей задаче для данной пары слов A и B мы хотим передать по пунктирному каналу слово X длины не более k, которое позволяет получить A из B и одновременно позволяет получить B из A (рис. 12.4).

Сразу же ясно, что это возможно лишь при $KS(A\,|\,B) \leqslant k$ и $KS(B\,|\,A) \leqslant k$. В самом деле, сложности слов, передаваемых по нижним наклонным линиям, не превосходят k (поскольку по определению эти слова просты относительно X), а вместе с тем эти слова позволяют получить B из A (формально говоря, из некоторого слова, простого относительно A) и наоборот.

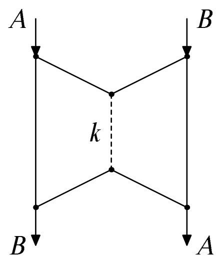


Рис. 12.4. Теорема Беннета – Гача – Ли – Витаньи – Цурека.

Получаем необходимое условие разрешимости задачи:

$$\max(KS(A\,|\,B), KS(B\,|\,A)) \leqslant k$$

(как всегда, все неравенства понимаются с точностью до логарифмических слагаемых). Как показали Беннет, Гач, Ли, Витаньи и Цурек в [5], это необходимое условие является на самом деле и достаточным. Вот точная формулировка:

Теорема 232. *Пусть A, B — произвольные слова, для которых $KS(A\,|\,B) < k$ и $KS(B\,|\,A) < k$. Тогда существует слово X длины k, для которого $KS(A\,|\,B, X) = O(\log k)$, $KS(B\,|\,A, X) = O(\log k)$ и $KS(X\,|\,A, B) = O(\log k)$.*

◀ Рассмотрим все пары слов $\langle A, B\rangle$, для которых $KS(A\,|\,B) < k$ и одновременно $KS(B\,|\,A) < k$. Получаем перечислимое бинарное отношение (на парах слов), все сечения которого (вертикальные, при фиксированном A, и горизонтальные, при фиксированном B) содержат не более 2^k элементов.

Рассматривая это отношение как (бесконечный) двудольный граф, можно сказать, что степень вершин этого графа (в каждой из долей) не превосходит 2^k.

Сейчас мы покажем, как разбить построенное множество пар на не более чем 2^{k+1} классов, каждый из которых является взаимно однозначным соответствием (не содержит двух пар на одной вертикали или горизонтали). В терминах графов: как раскрасить рёбра графа в 2^{k+1} цветов так, чтобы любые два ребра, имеющие общую вершину (слева или справа), имели разные цвета.

А именно, пронумеруем классы от 0 до $2^{k+1} - 1$. По мере появления новых пар в перечислении будем относить их к минимальному допустимому классу (в котором ещё нет пар с тем же первым или вторым элементом). (Другими словами, мы выбираем для вновь появившегося ребра первый цвет, не использованный для рёбер с общим концом.)

Ясно, что классов хватит, поскольку использованных классов меньше $2^k + 2^k$ (пар с тем же первым членом меньше 2^k, равно как и пар с тем же вторым членом). В терминах графа: новое ребро имеет общий левый конец менее чем с 2^k рёбрами и общий правый конец менее чем с 2^k, так что запрещены менее 2^{k+1} цветов.

Теперь в качестве X возьмём номер класса. Он содержит $k + 1$ битов, но отбрасывание одного бита меняет все условные сложности не более на $O(1)$. Ясно, что зная A, число k и номер класса, можно порождать множество пар, классифицировать их описанным способом и подождать появления на A-вертикали пары из класса

с данным номером X, поэтому $KS(A\,|\,B, X, k) = O(1)$ и $KS(A\,|\,B, X) = O(\log k)$. Аналогично и для $KS(B\,|\,A)$. Наконец, $KS(X\,|\,A, B, k) = O(1)$, поскольку зная A, B и k, можно дождаться классификации пары $\langle A, B\rangle$ и найти X. ▶

$\boxed{\mathbf{320}}$ Докажите более сильное утверждение о двудольных графах: если для (конечного) двудольного графа степень каждой вершины в левой и правой доле не превосходит N, то можно так раскрасить его рёбра в N цветов, чтобы рёбра одного цвета не имели общих концов. Объясните, почему в доказательстве теоремы 232 не удаётся сослаться на этот факт, а нужно доказывать его (ослабленный) вариант заново. [Указание. Можно считать, что степень равна N, после чего применить теорему Форда – Фалкерсона или теорему Холла о паросочетаниях. Нам этого недостаточно, так как в нашем случае граф не задан целиком, а строится постепенно, и рёбра надо красить on-line.]

В терминах программ доказанную теорему можно сформулировать так: для любых слов A и B существует программа сложности $\max(KS(A\,|\,B), KS(B\,|\,A))$ (с точностью до логарифмического слагаемого), которая переводит A в B и B в A. В самом деле, эта программа состоит из слова X, программы получения A из пары $\langle B, X\rangle$ и программы получения B из пары $\langle A, X\rangle$, а также инструкций по различению слов A и B (для чего достаточно указать номер бита, который в этих словах отличается; при $A = B$ утверждение тривиально).

$\boxed{\mathbf{321}}$ Пусть A и B — два независимых случайных слова длины n (то есть $KS(A) \approx n$, $KS(B) \approx n$ и $KS(\langle A, B\rangle) \approx 2n$. Укажите явно слово X, удовлетворяющее условиям теоремы. [Ответ: годится побитовая сумма слов A и B.]

Если сложности $KS(A\,|\,B)$ и $KS(B\,|\,A)$ различны, то теорему 232 можно уточнить следующим образом. Пусть, скажем, $KS(A\,|\,B)$ — бóльшая из них (легче получить B из A, чем наоборот). Оказывается, тогда можно разделить слово X на две части: на информацию, необходимую для преобразования A в B (в лёгкую сторону, длиной $KS(B\,|\,A)$) и остаток (длины $KS(A\,|\,B) - KS(B\,|\,A)$), добавление которого позволяет сделать и обратное преобразование.

Формально говоря, верно такое утверждение:

Теорема 233. *Пусть $KS(A\,|\,B) < k$, $KS(B\,|\,A) < l$ и $k > l$. Тогда можно найти такое слово X длины k, что $KS(X\,|\,A, B) = O(\log k)$, $KS(A\,|\,B, X) = O(\log k)$, а также $KS(B\,|\,A, X') = O(\log k)$, где X' — начало X длины l.*

◀ Рассуждаем аналогично доказательству теоремы 232, относя пары к 2^{l+1} классам и требуя, чтобы на одной вертикали было не более одной точки каждого класса, а на одной горизонтали — не более 2^{k-l} точек каждого класса. Тогда номер класса позволяет восстановить B по A, а для восстановления A по B нужно ещё $k - l$ битов информации (порядковый номер появления среди элементов данного класса). ▶

$\boxed{\mathbf{322}}$ Докажите утверждение теоремы 233 в форме, приведённой в [5]: в предположениях теоремы 233 найдутся слово Y сложности $k - l$ и слово X сложности l, для которых $KS(B, Y\,|\,A, X) = O(\log k)$ и $KS(A\,|\,B, Y, X) = O(\log k)$.

12.7. Кодирование при двух условиях

Рассмотрим теперь задачу передачи информации, которая в некотором смысле обобщает две предыдущие (рис. 12.5).

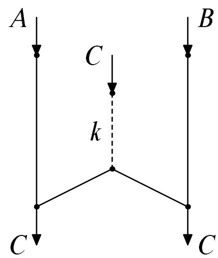


Рис. 12.5. Кодирование C при условиях A, B.

Если в этой схеме положить $A = B$, то получится схема передачи информации раздела 12.3 (симметрично продублированная). Если же положить $C = \langle A, B\rangle$, то мы приходим к задаче раздела 12.6 (в каждом из нижних узлов одно из слов A и B известно, поэтому восстановить пару означает восстановить второе слово).

Необходимыми условиями разрешимости задачи являются неравенства

$$KS(C\,|\,A) \leqslant k, \quad KS(C\,|\,B) \leqslant k.$$

Как установил Ан. А. Мучник [116], эти условия являются и достаточными. Вот точная формулировка:

Теорема 234. *Пусть A, B и C — произвольные слова сложности не более n, а k — натуральное число, причём $KS(C\,|\,A) \leqslant k$ и $KS(C\,|\,B) \leqslant k$. Тогда найдётся слово X длины не более $k + O(\log n)$, при котором $KS(X\,|\,C) = O(\log n)$, $KS(C\,|\,A, X) = O(\log n)$ и $KS(C\,|\,B, X) = O(\log n)$.*

В терминах программ эта теорема звучит так: для любых трёх слов A, B, C длины не более n существует программа сложности $\max(KS(C\,|\,A), KS(C\,|\,B)) + O(\log n)$, имеющая логарифмическую сложность относительно C, переводящая любое из слов A и B в C. (Как и раньше, в эту программу надо включить информацию, позволяющую различить A и B; она имеет логарифмическую сложность.)

Заметим, что это утверждение остаётся содержательным, даже если не требовать простоты программы относительно C: никакого другого доказательства его в этом частном случае не известно. Другими словами, доказательство достаточности (пока?) не удаётся упростить, даже если добавить два новых ребра (рис. 12.6).

Можно также сформулировать следующее обобщение теоремы 234 на случай различных условных сложностей:

Теорема 235. *Пусть A, B и C — произвольные слова сложности не более n, а $k \geqslant l$ — натуральные числа, причём $KS(C\,|\,A) \leqslant k$ и $KS(C\,|\,B) \leqslant l$. Тогда найдётся слово X длины k, для которого $KS(X\,|\,C) = O(\log n)$, $KS(C\,|\,A, X) = O(\log n)$ и $KS(C\,|\,B, X') = O(\log n)$ для слова X' длины l, являющегося началом слова X.*

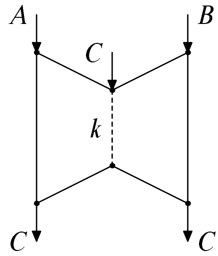


Рис. 12.6. Дополнительные рёбра.

323 Сформулируйте утверждение этой теоремы в терминах передачи информации по некоторой сети. [Указание: X передаётся по ребру пропускной способности k, из конца которого выходит ребро пропускной способности l.]

Все эти утверждения доказаны в той же работе Мучника [116]; приведём доказательство теоремы 235.

◀ Будем использовать тот же метод «отпечатков», что и раньше: слово X будет одним из немногих отпечатков слова C.

Однако рассуждение требует некоторых изменений. Если k и l отличаются, то отпечатки должны быть разных длин. Даже и в случае $k = l$ возникает очевидная проблема: среди отпечатков может найтись X, для которого $KS(C\,|\,A, X)$ мало, а также X', для которого $KS(C\,|\,B, X')$ мало, но нам ведь нужно одно и то же слово и для A, и для B.

Можно пытаться найти среди отпечатков «вдвойне хорошее» слово X, которое порождает мало коллизий и в S_A, и в S_B (через S_A и S_B мы обозначаем множества слов, простых относительно A и B соответственно). Такие слова существуют и их даже большинство (по тем же причинам, что и раньше: в $S_A \cup S_B$ лишь вдвое больше слов, чем в каждом из множеств S_A и S_B в отдельности), так что продолжая рассуждение с экспандером, можно заключить, что для большинства слов в S_A и для большинства слов в S_B найдётся вдвойне хороший правый сосед. Но дальше рассуждение не проходит: мы хотели бы сказать, что остальные слова имеют малую сложность, поскольку их мало и их можно порождать, но для их порождения надо знать сразу и A, и B, а у нас в условии лишь одно из слов A и B.

Что же делать? Будем рассматривать условия A и B независимо, но требовать, чтобы у слова C был не просто один хороший отпечаток, а чтобы больше половины отпечатков были хорошими. Если этого удаётся добиться для A и для B в отдельности, отсюда следует существование отпечатка, одновременно хорошего и для A, и для B.

Соответственно придётся изменить и свойство типа экспандера, которого мы требуем от двудольного графа $E \subset P \times Q$ (с левой долей P и правой долей Q). Теперь мы хотим, чтобы для любого множества $U \subset Q$ не слишком большого размера множество тех вершин $x \in P$, у которых половина или более соседей справа попадает в U, было бы малым. (Раньше мы считали плохими вершины, у которых *все* соседи попадали в U, а теперь достаточно половины.)

Более того, поскольку теперь нас интересуют разные условные сложности, мы будем рассматривать не только сами отпечатки, но и их начала (сразу всех длин, так проще). Утверждение о существовании графа с нужными свойствами теперь выглядит так (через $[u]_m$ обозначается начало слова u, имеющее длину m):

Лемма. *Пусть даны натуральные числа n и N, а также положительное число ε, причём*

$$n2^{N+2n+1}\varepsilon^{N/2} < 1.$$

Тогда существует семейство отображений

$$\chi_1, \ldots, \chi_N \colon \mathbb{B}^n \to \mathbb{B}^n$$

с таким свойством: для любого $m \in \{1, \ldots, n\}$ и для любого непустого подмножества $U \subset \mathbb{B}^m$, число элементов в котором не превосходит $\varepsilon 2^m$, количество тех $x \in \mathbb{B}^n$, для которых

$$[\chi_i(x)]_m \in U \quad \textit{для половины или более значений } i \in \{1, \ldots, N\},$$

меньше $|U|$ (числа элементов в U).

[Пояснение: мы говорим не о графе со степенью вершин N в левой доле, а о семействе N отображений, поскольку допускаем кратные рёбра ($\chi_i(x) = \chi_j(x)$ при $i \neq j$).]

Доказательство. Покажем, что для случайно выбранных функций $\chi_1, \ldots, \chi_N$ (все значения $\chi_i(x)$ при всех i и x независимы и равномерно распределены в $\mathbb{B}^n$) вероятность нарушения указанного свойства меньше единицы. Эту вероятность мы оценим сверху. Для каждого $m \leqslant n$, для каждого $t \leqslant \varepsilon 2^m$ и для любых множеств $T \subset \mathbb{B}^n$ и $U \subset \mathbb{B}^m$, содержащих по t элементов, оценим вероятность того, что для каждого элемента $x \in T$ не менее половины значений $[\chi_i(x)]_m$ (при $i = 1, \ldots, n$) попадает в U. Для фиксированного $x \in T$ вероятность того, что не менее половины его соседей попадает в U, не превосходит $2^N \varepsilon^{N/2}$, поскольку для каждого из не более чем 2^N подмножеств множества $\{1, \ldots, N\}$, содержащих более $N/2$ элементов, вероятность того, что все входящие в него значения i ведут внутрь U, не превосходит $\varepsilon^{N/2}$ (значения $[\chi_i(x)]_m$ при разных i независимы и равномерно распределены в $\mathbb{B}^m$, а доля U среди элементов $\mathbb{B}^m$ не превосходит ε). Такое событие должно произойти независимо для всех $x \in T$, так что полученную оценку надо возвести в степень t.

Таким образом, для интересующей нас вероятности (которая должна быть меньше единицы) мы получаем оценку

$$\sum_{m=1}^{n} \sum_{t=1}^{\varepsilon 2^m} \sum_{T \subset \mathbb{B}^n, |T| = t} \sum_{U \subset \mathbb{B}^m, |U| = t} (2^N \varepsilon^{N/2})^t.$$

Число различных множеств T не превосходит 2^{nt} (столько есть последовательностей длины t, составленных из элементов $\mathbb{B}^n$); число различных множеств U не превосходит 2^{mt}. Учитывая это, получаем верхнюю оценку

$$\sum_{m=1}^{n} \sum_{t=1}^{\varepsilon 2^m} 2^{tn} 2^{tm} 2^{tN} \varepsilon^{Nt/2}$$

или
$$\sum_{m=1}^{n} \sum_{t=1}^{\varepsilon 2^m} (2^n 2^m 2^N \varepsilon^{N/2})^t.$$
Внутренняя сумма представляет собой геометрическую прогрессию. В условиях леммы знаменатель этой прогрессии меньше $1/2$, и потому сумма прогрессии не превосходит удвоенного первого члена, который не зависит от t. Учитывая это, получаем верхнюю оценку
$$2n \cdot (2^{N+n+m} \varepsilon^{N/2}) \leqslant n 2^{N+2n+1} \varepsilon^{N/2},$$
что меньше единицы по условию леммы. Лемма доказана.

Мы будем использовать эту лемму при $\varepsilon = 1/8$. В этом случае можно переписать условие леммы как
$$n 2^{N+2n+1} < 8^{N/2}$$
или
$$\log n + N + 2n + 1 < 3N/2.$$
Видно, что можно положить $N = 6n$, и условие леммы будет выполнено при всех достаточно бóльших n.

Продолжим доказательство теоремы 235. Как и раньше, можно заменить слово C его кратчайшим (безусловным) описанием и считать его словом длины n, то есть элементом $\mathbb{B}^n$. (Сложность слов A и B роли не играет; в частности, она может быть и больше n.) Согласно лемме при $N = 6n$ и $\varepsilon = 1/8$ можно найти N отображений $\chi_1, \ldots, \chi_N \colon \mathbb{B}^n \to \mathbb{B}^n$ с указанными в лемме свойствами. При этом, как уже говорилось, взяв первый в каком-либо естественном порядке набор функций с такими свойствами, мы можем считать, что сложность этого набора есть $O(\log n)$, поскольку для его задания достаточно указать число n.

Пусть $KS(C\,|\,A) = k$ и $KS(C\,|\,B) = l$. (В условии теоремы были неравенства $KS(C\,|\,A) \leqslant k$ и $KS(C\,|\,A) \leqslant l$, но мы имеем право доказывать теорему для уменьшенных значений k и l, от этого утверждение становится только сильнее.) Оставляя от хеш-значений только k или l первых битов, мы получим N отображений $\mathbb{B}^n$ в $\mathbb{B}^k$ (соответственно в $\mathbb{B}^l$). Эти семейства задают двудольные графы на $\mathbb{B}^n \times \mathbb{B}^k$ и $\mathbb{B}^n \times \mathbb{B}^l$, в которых степень каждой вершины слева равна N (считая кратные рёбра). Нас интересуют ограничения этих графов на S_A и S_B, где S_A состоит из слов длины n, имеющих сложность не более k относительно A, а S_B состоит из слов длины n, имеющих сложность не более l относительно B. Мы выделяем в $\mathbb{B}^k$ плохие вершины, имеющие более n^c соседей в S_A; аналогичным образом плохими вершинами в $\mathbb{B}^l$ мы считаем те, у которых имеется более n^c соседей в S_B. (Точное значение достаточно большой константы c мы выберем позже.)

Число плохих вершин в обоих случаях не превосходит соответственно
$$2N \cdot 2^k / n^c \text{ и } 2N \cdot 2^l / n^c,$$
так как степень плохой вершины больше n^c, общее число рёбер в графе не больше $|S_A| \cdot N$ (соответственно $|S_B| \cdot N$), а $|S_A| < 2 \cdot 2^k$ и $|S_B| < 2 \cdot 2^l$.

Назовём плохими в S_A те вершины, у которых половина или более соседей в графе на $S_A \times \mathbb{B}^k$ (с учётом кратности) являются плохими. Лемма гарантирует, что число плохих вершин в S_A меньше

$$2N \cdot 2^k / n^c$$

(мы считаем, что c достаточно велико, поэтому оценка на число плохих вершин меньше $\varepsilon 2^k$, где $\varepsilon = 1/8$; напомним, что $N = 6n$). Поскольку плохие вершины можно перечислять, зная n, k и A, сложность любой из них относительно A не превосходит

$$\log(2N \cdot 2^k / n^c) + O(\log n) \leqslant k - c \log n + O(\log n).$$

При достаточно большом c все плохие вершины имеют сложность (относительно A) меньше k, и слово C не попадает в их число. Это значит, что больше половины значений $[\chi_i(C)]_k$ (среди N вариантов $i = 1, 2, \ldots, N$) являются хорошими в $\mathbb{B}^k$.

Повторяя те же рассуждения для графа в $S_B \times \mathbb{B}^l$, мы получаем, что больше половины значений $[\chi_i(C)]_l$ хороши в S_B. Следовательно, найдётся значение i, которое даёт хороших соседей в обоих случаях. Тогда $X = [\chi_i(C)]_k$ и $X' = [\chi_i(C)]_l$ являются искомыми. ▶

$\boxed{\textbf{324}}$ Сформулируйте и докажите аналогичное утверждение для трёх условий (или для полиномиального их числа).

Может возникнуть желание ещё усилить результаты теорем 229 и 235. Нельзя ли ограничиться одним-единственным отпечатком? Скажем, нельзя ли для слова A сложности n найти такое слово X длины $n/2$, что $KS(A \mid X, B) \approx 0$ для любого слова B, для которого $KS(A \mid B) \leqslant n/2$?

Легко понять, что этого всё-таки добиться нельзя. В самом деле, пусть такое слово X есть. Тогда условная сложность A относительно X не превосходит $n/2$ (это следует из неравенства $KS(A \mid X, B) \approx 0$ для B, равного половине кратчайшего описания A). Если взять X в качестве B, то в условиях X и B информация будет дублироваться, и сложность $KS(A \mid X, B)$ будет примерно равна $n/2$ (а не нулю, как нам хотелось).

$\boxed{\textbf{325}}$ Покажите, что двух (и любого фиксированного числа) отпечатков недостаточно. [Указание. Допустим, можно обойтись k отпечатками. Без ограничения общности можно считать их все несжимаемыми. Для каждого i и каждого отпечатка рассмотрим слово, составленное из начала длины i этого отпечатка, и обозначим через B_i конкатенацию полученных слов. При некотором $i = i_0$ условная сложность A относительно B_i примерно равна $n/2$. В самом деле, условная сложность A относительно конкатенации целых отпечатков не больше $n/2$ (поскольку это верно даже если в условии оставить только отпечаток, обслуживающий слово B, составленное из $n/2$ первых битов минимального описания A) и при увеличении i на 1 условная сложность $KS(A \mid B_i)$ изменяется незначительно. Ясно, что это i_0 не меньше $n/2k$. Ни один из k отпечатков не может обслужить B_{i_0}, поскольку при добавлении его к условию сложность $KS(A \mid B_{i_0})$ понижается максимум на $n/2 - i_0 \leqslant n/2 - n/2k$ (так как $i_0 k$ его битов уже есть в условии), а должна на $n/2$.]

12.8. Поток информации через разрез

Мы рассмотрели несколько схем передачи информации; для каждой из них были указаны необходимые и достаточные условия разрешимости соответствующей задачи. Во всех приведённых примерах эти условия получаются по некоторой единой схеме, которую мы сейчас укажем явно.

Пусть задан граф передачи информации (ориентированный ациклический граф, на некоторых рёбрах которого указаны максимальные пропускные способности; для некоторых вершин заданы также входные и выходные слова). Мы хотим указать необходимые условия, то есть условия на эти слова, которые заведомо будут выполнены, если соответствующая задача разрешима.

Выделим произвольное множество I вершин графа (разрез) и будем изучать поток информации, проходящей через этот разрез (идущей извне множества I внутрь него). Рассмотрим пропускные способности всех рёбер, начало которых лежит вне I, а конец внутри I. Если среди них есть хоть одно ребро с неограниченной пропускной способностью, то никакого необходимого условия для такого I не получится. Пусть все пропускные способности этих рёбер ограничены и равны $u_1, \ldots, u_k$. Пусть $V_1, \ldots, V_l$ — все входные слова для вершин из I, а $W_1, \ldots, W_m$ — все выходные слова для вершин из I. Тогда можно записать такое необходимое условие:

$$KS(W_1, \ldots, W_m \mid V_1, \ldots, V_l) \leqslant u_1 + \ldots + u_k.$$

(Как всегда, неравенства рассматриваются с точностью до логарифма суммарных длин входящих в них слов.) В самом деле, если мы знаем все слова $V_1, \ldots V_l$, а также все слова, написанные на ведущих внутрь I рёбрах, то можем восстановить (с логарифмической дополнительной информацией) все слова, выходящие из вершин множества I, включая $W_1, \ldots, W_m$. Это нужно делать, рассматривая вершины множества I по очереди (начало любого ребра должно предшествовать его концу; такой порядок существует, так как в графе нет циклов).

По существу это рассуждение оценивает «поток информации через разрез».

Покажем на примере, как получаются необходимые условия в рассмотренных нами случаях. Рассмотрим схему передачи информации раздела 12.6 и в качестве множества I возьмём множество из трёх вершин, показанное на рис. 12.7 внутри пунктирной линии.

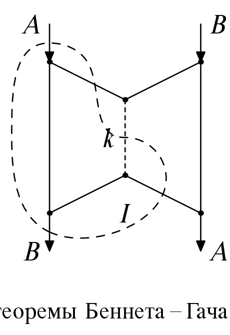


Рис. 12.7. Разрез для теоремы Беннета – Гача – Ли – Витаньи – Цурека.

В это множество входят слово A и слово длины k (по ребру ограниченной пропускной способности); два других ребра графа, пересекающие границу I, ведут изнутри наружу (напомним, что по нашему соглашению все рёбра идут сверху вниз). Получается условие $KS(B\,|\,A) \leqslant k$, которое мы и рассматривали.

326 Покажите, что все остальные необходимые условия, указанные нами для рассмотренных выше задач, также могут быть получены при подходящем выборе множества I на соответствующем графе.

12.9. Сети с одним источником

Возникает естественный вопрос: являются ли необходимые условия, указанные в предыдущем разделе (и взятые для всех множеств I, для которых они имеют смысл), одновременно и достаточными? В наших предыдущих примерах это оказывалось именно так. В общем случае, как мы увидим дальше, это неверно. Однако в случае, когда мы хотим передать какое-то одно слово A из одного источника в несколько мест назначения (другими словами, когда входное слово только одно, а все выходные слова равны входному), наши условия будут и достаточными.

Для классической теории информации аналогичная задача была рассмотрена в работах [1, 86]; наше рассуждение следует использованной там схеме (с некоторыми изменениями, связанными с переходом от шенноновской энтропии к колмогоровской сложности).

Начнём с примера: пусть мы хотим передать слово A длины $2k$ в три пункта назначения, как показано на рис. 12.8; все каналы связи имеют неограниченную пропускную способность, кроме трёх первых, которые имеют пропускную способность k битов. Можно ли это сделать?

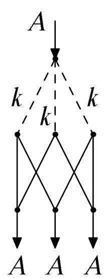


Рис. 12.8. Деление информации на части.

Легко понять, что в каждый из пунктов назначения *по отдельности* можно передать слово A. Например, чтобы передать его в левую из трёх вершин, надо разбить его на две половины, каждая по k битов, и передать эти половины по двум левым каналам связи пропускной способности k (третий канал пропускной способности k бесполезен, так как идущая по нему информация не может попасть в нужную нам вершину).

Аналогичным способом легко передать A в любую из трёх вершин, использовав два канала из трёх. Но передать слово A *одновременно* в три вершины уже не так просто: для этого его надо разрезать на «три половины», причём так, чтобы из любых двух половин его можно было составить обратно.

Стандартный способ («разделение секрета с помощью линейных отображений») состоит в том, чтобы передать по трём каналам слова A_1, A_2 и $A_1 \oplus A_2$, где A_1 и A_2 — две половины слова A (каждая содержит k битов), а $A_1 \oplus A_2$ — побитовая сумма слов A_1 и A_2 по модулю 2. По любым двум из этих трёх k-битовых слов можно восстановить третье (оно равно их побитовой сумме), а потому можно восстановить и слово A.

Оказывается, аналогичный метод можно применить и в общем случае, и справедливо такое утверждение.

Теорема 236. *Рассмотрим схему передачи информации с единственным входным словом A длины n и выходными словами A, в которой заданы целые ограничения на пропускные способности некоторых рёбер.*

Пусть выполнены все необходимые условия описанного типа, то есть для любого множества вершин J, не содержащего вершину, в которую входит A, и содержащего хотя бы одну вершину, из которой выходит A, сумма пропускных способностей рёбер с началами вне J и концами в J не меньше n (или есть хотя бы одно ведущее внутрь ребро с неограниченной пропускной способностью).

Тогда задача передачи информации по этой схеме разрешима с точностью до $O(\log n)$: можно написать на рёбрах такие слова, чтобы соответствующие условные сложности в каждой вершине не превосходили $O(\log n)$.

(Константа, подразумеваемая в $O(\log n)$, зависит от графа, но не зависит от числа n, пропускных способностей и слова A.)

◄ Рассмотрим сначала случай одного выходного слова. В этом случае требуется передать некоторый объём информации (а именно, n битов) из одной вершины (назовём её s) в какую-то другую вершину t (только одну!) по рёбрам.

Запакуем каждый бит в отдельный конверт; получится n конвертов, которые изначально находятся в вершине s. Организуем доставку этих конвертов в вершину t по рёбрам графа. При этом мы следим за пропускной способностью и требуем, чтобы по ребру с ограничением k было перевезено не более k конвертов.

Эта задача разрешима по теореме Форда – Фалкерсона о максимальном потоке и минимальном разрезе (см., например, [33]; поскольку все ограничения целочисленные, то и поток будет целочисленным).

Теперь на каждом ребре можно написать те биты, которые содержатся в перевозимых по нему конвертах. Точнее, после применения алгоритма для задачи Форда – Фалкерсона мы для каждого ребра имеем некоторый перечень битов (список их номеров в возрастающем порядке; общее число номеров не превосходит пропускной способности ребра), и записываем на этом ребре подпоследовательность, состоящую из битов с этими номерами.

Покажем, что выполнено ограничение на условную сложность. Рассмотрим вершину и слова на выходящих и входящих рёбрах этой вершины. Слова на выходящих рёбрах составлены путём перемешивания битов в словах на входящих рёбрах; схема

этого перемешивания не зависит от слова A и алгоритмически вычисляется, если известно число n и пропускные способности рёбер. Без ограничения общности можно предполагать, что пропускные способности рёбер не больше n, поэтому схема перемешивания имеет сложность $O(\log n)$ (константа зависит от графа). А зная эту схему, можно получить выходящие слова по известным входящим.

Случай передачи информации в одну вершину разобран.

Идея доказательства теоремы в общем случае состоит в том, чтобы производить линейное кодирование. До сих пор мы лишь перекоммутировали биты (перекладывали конверты из входящих линий в выходящие). Более общий способ: в каждом узле применяется линейное отображение. Каждый выходящий бит будет линейной функцией от входящих битов. Для начала будем рассматривать биты как элементы поля $\mathbb{F}_2$ (поле из двух элементов: нуля и единицы, при этом $1 + 1 = 0$). Тогда l-битовые слова будут элементами l-мерного векторного пространства над этим полем.

Пусть в некоторую вершину входят рёбра с пропускной способностью $i_1, \ldots, i_p$, а выходящие рёбра имеют пропускные способности $j_1, \ldots, j_q$. (Мы будем считать, что все пропускные способности рёбер конечны, заменив бесконечность на n, число битов в передаваемом слове.) Тогда преобразование информации в этой вершине задаётся матрицей размера $(j_1 + \ldots + j_q) \times (i_1 + \ldots + i_p)$; умножая её на столбец входных битов (со всех входящих рёбер), получаем столбец выходных битов (для всех выходящих рёбер). Заметим, что по ребру пропускной способности k передаётся ровно k битов (независимо от того, насколько это ребро кажется полезным).

В приведённом выше примере (когда слово делилось на две части, после чего они складывались побитово) как раз и выполнялись линейные преобразования.

Пусть для каждой вершины такие линейные преобразования (матрицы) заданы. Тогда для каждого выхода возникает отображение входа в этот выход — некоторое линейное отображение n-мерных пространств над полем $\mathbb{F}_2$. Нам хотелось бы, чтобы все эти отображения были обратимы; в этом случае каждое выходное слово содержало бы полную (с точностью до линейного взаимно однозначного соответствия) информацию о входных битах, и задача передачи информации была бы разрешима для произвольного входного слова A (на каждом ребре графа нужно было бы написать передаваемый по нему набор битов для слова A).

В самом деле, сами линейные отображения можно выбрать небольшой сложности: если мы знаем, что существуют такие матрицы преобразования в вершинах, при которых все отображения входа в выходы одновременно обратимы, то первый (в каком-либо естественном порядке) набор таких матриц имеет логарифмическую сложность, поэтому для каждой вершины условная сложность выходящих слов при известных входящих логарифмическая.

Проблема в том, что не всегда можно добиться одновременной обратимости всех отображений вход $\to$ выход.

327 Рассмотрим граф на рис. 12.9; входное слово имеет длину 2, пропускная способность всех рёбер равна 1. Покажите, что нельзя указать такие линейные преобразования в вершинах, чтобы все шесть отображений входа в выход были бы

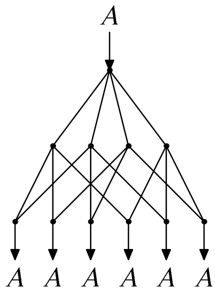


Рис. 12.9. Двух элементов в поле мало.

обратимыми. [Указание. Существует всего три ненулевых линейных функционала на $\mathbb{F}_2^2$, поэтому в двух промежуточных вершинах будет одинаковая информация.]

Подчеркнём, что для каждого выхода по отдельности можно подобрать преобразования в вершинах, при которых соответствующее отображение обратимо, как видно из рассуждения для случая одного выхода. (При этом преобразования в вершинах будут перестановками битов, этого достаточно.) Проблема именно в том, чтобы сделать это одновременно для всех входов.

Временно изменим постановку задачи и вместо битов будем рассматривать элементы произвольного поля F. Входом тогда будет вектор из F^n, через ребро пропускной способности k будет проходить вектор из F^k, а преобразования в вершинах будут линейными над F, то есть будут задаваться матрицами с элементами из F.

Мы докажем, что если поле F достаточно велико, то можно так выбрать преобразования в вершинах, чтобы отображения входа во все выходы были бы одновременно обратимы.

Будем считать элементы матриц преобразования в вершинах переменными (принимающими значения в F). Тогда элементами матрицы преобразования входа в выход будут многочлены от этих переменных, и потому её определитель тоже будет многочленом. Степень этого многочлена, как легко проверить, не превосходит nE, где E — число рёбер графа (при движении от входа к выходу в каждом узле степень каждого матричного элемента увеличивается на единицу, а степень определителя матрицы $n \times n$ в n раз больше степени матричных элементов).

Таким образом, для каждого выхода у нас есть многочлен не очень большой степени (определитель соответствующего преобразования), причём известно, что этот многочлен не обращается тождественно в нуль (ведь на этот выход в отдельности мы передавать информацию умеем). Теперь вспомним простой алгебраический факт:

Лемма. *Многочлен степени d от m переменных над полем F либо равен нулю тождественно, либо принимает нулевое значение с вероятностью не более $d/|F|$ (все элементы F^m считаем равновероятными, $|F|$ — число элементов поля F).*

(Говоря о степени многочлена, мы имеем в виду суммарную степень по всем переменным.)

Доказательство проходит индукцией по m. Для $m = 1$ лемма утверждает, что число корней многочлена с одной переменной не превосходит его степени (и доказывается разложением на множители). Для бóльших m запишем многочлен как многочлен некоторой степени d_1 от одной переменной, коэффициенты которого представляют собой многочлены от $m-1$ переменных. Обозначим через d_2 степень его старшего коэффициента как многочлена от $m-1$ переменной. Этот старший коэффициент может быть нулевым или ненулевым в зависимости от значений остальных переменных; вероятность того, что он окажется нулевым, не больше $d_2/|F|$ по предположению индукции, а если старший коэффициент ненулевой, то вероятность попасть в корень такого многочлена от одной переменной не больше $d_1/|F|$. Так что всего получаем $(d_2 + d_1)/|F| \leqslant d/|F|$. Лемма доказана.

Теперь заметим, что если вероятность обращения в нуль определителя для каждого выхода меньше единицы, делённой на число выходов, то найдутся значения переменных, при которых все определители ненулевые.

Подсчитаем: степень многочлена не больше nE, выходных рёбер тоже не больше E, так что если $nE^2 < |F|$, то наше рассуждение доказывает, что есть линейные преобразования в вершинах, делающие сразу все отображения входа в выходы обратимыми одновременно. (И такое преобразование можно найти перебором, так что среди них есть простые.)

Как это можно применить в ситуации, когда имеется слово из n битов, а вовсе не элементов конечного поля? Как это принято в теории кодирования, будем разбивать слово на блоки некоторой длины k (в количестве n/k) и считать каждый блок элементом поля из 2^k элементов (такое поле, как известно из алгебры, существует).

Если оказалось, что число n, а также все пропускные способности делятся на k, и при этом $2^k > (n/k)E^2$, то всё хорошо.

Как быть в общем случае? Для начала нужно выбрать значение k, при котором $2^k > nE^2$ (мы берём k с запасом, пренебрегая делением на k в правой части). Отметим, что при этом $k = O(\log n)$. Затем надо округлить n и пропускные способности до целых кратных числа k; при этом n надо округлять с уменьшением, а пропускные способности — с увеличением, чтобы не нарушить условия на пропускные способности разрезов. При этом погрешность будет порядка $O(\log n)$, и остаётся только воспользоваться доказанным утверждением. ►

328 Используя описанный метод, постройте вероятностный полиномиальный алгоритм отыскания величины максимального потока в ориентированном графе без циклов с целыми пропускными способностями. [Указание. Достаточно научиться определять, есть ли в графе поток размера n, и находить его, если есть (для любого данного n). Для этого сопоставим каждому ребру линейное отображение, как в доказательстве теоремы. При этом пропускные способности выходного и входного ребра будем считать равными n. Существование потока размера n равносильно невырожденности линейного отображения (вход $\to$ выход) для некоторых линейных отображениях на рёбрах. А существование таких линейных отображений на рёбрах можно проверять вероятностным алгоритмом с помощью алгебраической леммы.]

Теперь мы переходим к примерам, где условия на поток информации через разрез оказываются необходимыми, но не достаточными.

12.10. Выделение общей информации

Один пример такого рода (где необходимые условия на поток информации не являются достаточными) мы по существу уже рассматривали. Это задача о выделении общей информации, рассмотренная в главе 11. В разделе 11.2 для данных слов x, y и чисел α, β и γ мы интересовались, найдётся ли слово z, для которого

$$KS(z) < \alpha, \quad KS(x\,|\,z) < \beta, \quad KS(y\,|\,z) < \gamma.$$

Легко понять, что с точностью до логарифмических слагаемых эту задачу можно сформулировать как задачу передачи информации в графе рис. 12.10. В самом

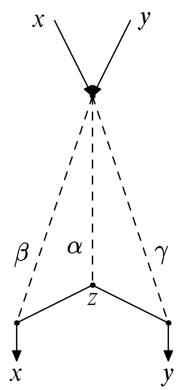


Рис. 12.10. Задача выделения общей информации.

деле, если удаётся найти слово z, для которого выполнены указанные неравенства, то его можно передавать по среднему ребру, а по крайним рёбрам передать условные описания x и y при известном z. (И это слово, и условные описания можно найти перебором при известных x, y, так что в верхней вершине новой информации не возникает.) Напротив, если задача о передаче информации с указанными ограничениями разрешима, то слово z, передаваемое по среднему ребру, удовлетворяет неравенствам (с логарифмической точностью).

Видно, что условия на пропускные способности разрезов соответствуют неравенствам

$$KS(x) \leqslant \alpha + \beta, \quad KS(y) \leqslant \alpha + \gamma, \quad KS(x, y) \leqslant \alpha + \beta + \gamma,$$

и вся глава 11 была посвящена примерам ситуаций, когда эти условия оказываются недостаточными для существования слова z («общей информации»).

12.11. Упрощение программы

В предыдущем разделе приведён пример задачи передачи информации на графе, в которой необходимые условия на потоки не являются достаточными для её

разрешимости. Представляют интерес другие примеры такого рода, по возможности попроще. Оказывается, что утверждение теоремы 229 лежит довольно близко к границе: если рассмотреть чуть более общую постановку задачи, то необходимое условие перестанет быть достаточным.

Рассмотрим задачу рис. 12.11 (её предложил М. Вьюгин). Разница с теоремой Мучника в том, что внизу требуется не восстановить одно из двух входных слов (P и A), а получить некоторое третье слово B. Оказывается, что и в этой задаче

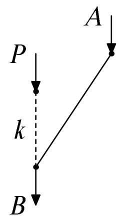


Рис. 12.11. Задача упрощения программы.

необходимые условия на поток информации $KS(B \mid A) \leqslant k$ и $KS(B \mid A, P) = 0$ (как всегда, рассматриваемые с логарифмической точностью) не являются достаточными.

Подробно эта задача рассмотрена в статье [121], где приводятся игровое, вероятностное и комбинаторное объяснения этого факта.

12.12. Минимальная достаточная статистика

Другая задача, в которой необходимые условия также не совпадают с достаточными, показана на рис. 12.12. Здесь на выходе нужно получить одно из входных

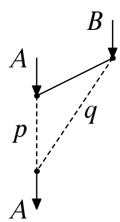


Рис. 12.12. Два канала ограниченной пропускной способности.

слов, но теперь уже пропускные способности обоих каналов ограничены. При этом мы дополнительно разрешаем использования информацию о слове B при формировании сообщения по левому каналу.

Эта задача связана с понятием «минимальной достаточной статистики» в теории вероятностей. Пусть имеется пара совместно распределённых случайных величин θ и X. Первую из них считают «параметром» и для каждого её значения рассматривают условное распределение X_θ второй величины. Например, можно считать, что

θ выбирается равномерно на отрезке $[0, 1]$, после чего в качестве X берётся результат n независимых испытаний с вероятностью успеха θ (условное распределение X_θ представляет собой бернуллиево распределение с параметром θ на множестве $\mathbb{B}^n$ слов длины n).

Предположим, что мы хотим восстановить значение неизвестного параметра θ, зная его априорное распределение и наблюдая значение величины X_θ. Не вся информация, содержащаяся в X_θ, при этом существенна. В нашем примере важно лишь число успехов, а какие именно испытания были успешными, роли не играет. Формально говоря, величина $N(X)$, равная числу единиц в X, содержит всю информацию о θ, которую можно извлечь из X:

$$I(N(X):\theta) = I(X:\theta).$$

Для произвольной функции N левая часть не превосходит правой; те функции, где это неравенство обращается в равенство, называют *достаточными статистиками*. Это же условие можно сформулировать иначе: θ и X независимы при известном $N(X)$. Ещё одна переформулировка: $H(\theta \mid N(X)) = H(\theta \mid X)$.

Сама величина X, очевидно, является достаточной статистикой; из примера видно, что статистика может содержать много «лишней информации». Достаточную статистику называют *минимальной*, если она является функцией любой другой достаточной статистики. Для случайных величин с конечным числом значений минимальная достаточная статистика всегда существует: надо склеить те значения X, которым соответствуют одинаковые условные распределения параметра θ. Минимальная достаточная статистика имеет наименьшую энтропию среди всех достаточных статистик.

$\boxed{\mathbf{329}}$ Предположим, что все возможные значения θ имеют ненулевую вероятность. Докажите, что в определении достаточной статистики важны только условные вероятности $P[X = x \mid \theta = t]$ для всех пар x, t. [Указание. Сопоставим каждому x вектор, состоящий из чисел $P[X = x \mid \theta = t]$ для всех возможных значений t случайной величины θ. Тогда функция N является достаточной статистикой тогда и только тогда, когда любые два склеенных ей слова имеют пропорциональные вектора.]

Задачу поиска минимальной достаточной статистики можно сформулировать так: среди величин X', для которых $H(X' \mid X) = 0$ (т. е. X' есть функция от X) мы выделяем те, для которых $H(\theta \mid X')$ минимально возможная (равна $H(\theta \mid X)$), и среди них минимизируем $H(X')$.

В сложностной постановке мы имеем дело с двумя словами A (соответствует параметру θ) и B (соответствует X). Мы хотим извлечь из B некоторую часть информации B', для которой $KS(A \mid B')$ минимально возможная (приближается к $KS(A \mid B)$), и $KS(B')$ как можно меньше. На рисунке B' передаётся по каналу с ограничением q, а условное описание A при известном B' передаётся по каналу с ограничением p. Минимальная достаточная статистика соответствует минимизации q при минимально возможном $p \approx KS(A \mid B)$.

Мы рассматриваем более общий вопрос: при каких p и q передача информации по указанной схеме возможна. Необходимые условия, связанные с потоками

информации через разрезы, таковы: $KS(A) \leqslant p + q$ и $KS(A \mid B) \leqslant p$. (Как всегда, мы опускаем оговорки о логарифмических поправках.)

330 Нарисуйте соответствующие разрезы.

Для некоторых пар слов эти необходимые условия являются также и достаточными.

331 Проверьте, что если у слов A и B полностью выделяется общая информация (например, слова A и B являются перекрывающими кусками случайного слова), то эти необходимые условия являются также и достаточными. [Указание. Условие $KS(A \mid B) \leqslant p$ позволяет передать по левому каналу часть слова A, не попавшую в B, и ещё сколько-то, а остаток можно передать по правому каналу (условие $KS(A) \leqslant p + q$ гарантирует, что места в канале хватит).]

Однако в общем случае необходимые условия могут не быть достаточными. Для наглядности зафиксируем сложности и условные сложности слов A и B: пусть оба слова A и B имеют сложность $2n$, а пара $\langle A, B\rangle$ имеет сложность $3n$ (и потому условные сложности равны n). Необходимые условия $p + q \geqslant 2n$ и $p \geqslant n$ для этого случая показаны на рис. 12.13.

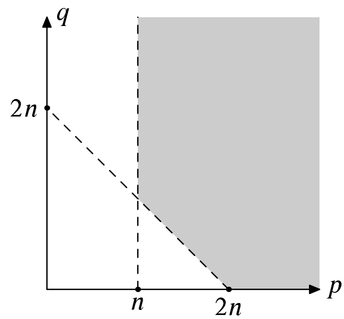


Рис. 12.13. Необходимые условия.

Теперь попытаемся понять, при каких p и q задача заведомо разрешима (для любой пары слов A и B указанной сложности). Можно передавать слово A целиком по левому каналу, поэтому задача разрешима при $p = 2n$, $q = 0$ (а также для всех бóльших p и q). Можно передавать слово B целиком по правому каналу, поэтому задача разрешима при $p = n$, $q = 2n$ (а также для всех бóльших p и q). Это соответствует двум квадрантам с вершинами в $(2n, 0)$ и $(n, 2n)$. Более того, если от B отрезать (скажем, последние) k битов, то условная сложность $KS(A \mid B)$ увеличится не более чем на k, поэтому задача разрешима при $q = 2n - k$, $p = n + k$. Таким образом, вся тёмно-серая область на рисунке 12.14 (обозначенная там как G) соответствует парам $\langle p, q\rangle$, для которых задача разрешима.

Как и в случае с выделением общей информации (глава 11), можно сказать, что «профиль» пары слов A, B (множество пар $\langle p, q\rangle$, при которых задача разрешима) зависит от того, какую именно пару слов (с данной сложностью и условной сложностью) мы возьмём. Как мы видели в задаче 331, для некоторых пар слов

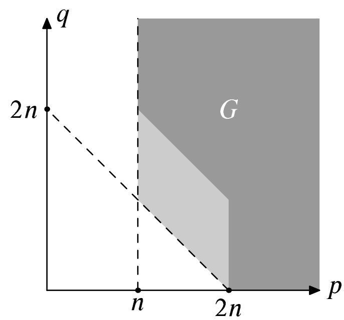


Рис. 12.14. Достаточные условия.

этот профиль совпадает с максимально возможным. Следующая теорема утверждает, что бывает и наоборот — есть пара слов, при которых профиль совпадает с минимально возможным (областью G).

Нужно только понять, как это правильно сформулировать. Хотелось бы сказать, что для всех B', простых относительно B, пара $\langle KS(A\,|\,B'), l(B')\rangle$ находится в $O(\log n)$-окрестности множества G, причём константа в $O(\log n)$ не зависит ни от n, ни от B'. Однако выражение «простые относительно B» требует количественного уточнения: мы должны указать некоторую границу r и рассматривать слова B', для которых $KS(B'\,|\,B) < r$. Естественно ожидать, что чем больше r, тем дальше может отходить пара $\langle KS(A\,|\,B'), l(B')\rangle$ от G. Как мы увидим, связь тут линейная.

Теорема 237. *При любом n существуют слова A, B сложности $2n + O(\log n)$, для которых $KS(A, B) = 3n + O(\log n)$ и для любого B' пара $\langle KS(A\,|\,B'), l(B')\rangle$ находится в $O(\log n + KS(B'\,|\,B))$-окрестности множества G.*

◄ Утверждение этой теоремы, как обычно, соответствует некоторой игре. Мы опишем эту игру, укажем выигрышную стратегию и выведем отсюда утверждение теоремы.

Пусть фиксировано значение n; при каждом n будет своя игра. В этой игре мы имеем право для каждого B длины $2n$ указать до 2^n слов длины $2n$, условно именуя их «n-простыми относительно B».

Противник для каждых p, q и r (из некоторого множества M допустимых троек натуральных чисел; это множество мы опишем позднее) имеет право:

- для каждого слова B длины $2n$ указать до 2^r слов длины q, которые условно именуются «r-простыми относительно B»;
- для каждого слова B' длины q указать до 2^p слов длины $2n$, которые условно именуются «p-простыми относительно B'».

Для каждой тройки $\langle p, q, r\rangle$ (из множества допустимых троек) это происходит независимо. Можно сказать, что мы играем против команды противника, в которой для каждой тройки $\langle p, q, r\rangle$ есть свой игрок, делающий свои объявления и подчиняющийся своим ограничениям, но цель игры у них общая.

Кроме этого, в команде противника есть ещё два дополнительных игрока. Один из них имеет право браковать слова длиной $2n$ (но не более 2^{2n-1} штук, то есть половины всех слов), называя их «плохими». Второй имеет право для каждого слова B длины $2n$ забраковать до 2^{n-2} слов длины $2n$ (свои для каждого B), объявляя их «плохими для данного B». (Рассматриваемый далее противник будет браковать слова длиной $2n$, имеющие сложность менее $2n-1$, а также для каждого слова B браковать слова A, для которых $KS(A\,|\,B) < n-2$, но в определение игры это не входит.) Мы используем границу $n-2$ (а не, скажем, $n-1$), поскольку нам потребуется некоторый запас, см. ниже.

Ходы делаются игроками постепенно: в любой момент каждый из игроков может объявить новое слово плохим (не нарушая количественных ограничений). Поскольку общее число возможных ходов конечно, игра рано или поздно кончится, хотя внешний наблюдатель (не знающий используемых игроками стратегий) не сможет сказать, так ли это — игроки не объявляют о конце игры.

Остаётся определить, кто выигрывает в предельной позиции. Будем считать, что выиграл противник, если для каждой пары слов $\langle A, B\rangle$ длины $2n$ (каждое), в которой слово A объявлено n-простым относительно слова B, причём A и B не забракованы (не объявлены плохими) и A не забраковано для данного B (не объявлено плохим для этого B), найдётся допустимая тройка $\langle p, q, r\rangle$ и слово B' длины q, для которой

- B' объявлено r-простым относительно B;
- A объявлено p-простым относительно B'.

Мы укажем простую стратегию, позволяющую выигрывать в этой игре (при некотором множестве допустимых троек $\langle p, q, r\rangle$), а затем выведем из этого утверждение теоремы.

Выигрывающая стратегия состоит в том, что на каждый ход противника (любого из участников противостоящей команды), состоящий в объявлении нового «простого» или «плохого» слова, мы отвечаем одним своим ходом. Этот ход состоит в добавлении (для некоторого слова B длины $2n$) одного n-простого слова A длины $2n$, которое помешало бы противнику выиграть (если он не сделает нового хода). Для этого нужно, чтобы выполнялись следующие условия:

- выбранные слова A и B (ещё) не забракованы (не объявлены плохими);
- слово A (ещё) не объявлено плохим относительно B;
- ни для одной допустимой тройки $\langle p, q, r\rangle$ не найдётся слова B' длины q, которое было бы объявлено $\langle p, q, r\rangle$-игроком команды противника r-простым относительно B, и для которого A объявлено этим же игроком p-простым относительно B'.

Почему это возможно? На любой момент игры имеется не менее 2^{2n-1} незабракованных слов. Взяв в качестве B одно из них, мы имеем право объявить любое слово n-простым относительно B, если только уже не объявили 2^n слов простыми относительно B. Но если такое случилось для всех незабракованных B, это значит, что мы уже сделали $2^{2n-1}\cdot 2^n = 2^{3n-1}$ ходов, отвечая одним своим ходом на каж-

дый ход противника, а противник так много ходов сделать не сможет (см. далее). Таким образом, слово B мы выбрать сможем.

Выбрав B, мы начинаем выбирать A. Фиксируем допустимую тройку $\langle p, q, r\rangle$ и подсчитаем, сколько слов не годятся из-за неё. Имеется не более 2^r слов B' длины q, объявленных r-простыми относительно B. Для каждого B' есть не более 2^p слов длины $2n$, объявленных p-простыми относительно этого B'. Таким образом, нам не годятся 2^{p+r} слов (для каждой тройки $\langle p, q, r\rangle$ из множества допустимых троек M), всего $2^{p+r} \cdot |M|$. Ещё нам не подходят слова, объявленные плохими (не более 2^{2n-1} штук), а также слова, объявленные плохими относительно выбранного B (их не более 2^{n-2}). Таким образом, искомый ход заведомо возможен, если

$$2^{p+r} \cdot |M| + 2^{2n-1} + 2^{n-2} < 2^{2n}. \qquad (*)$$

Теперь надо подсчитать, сколько ходов (в течение одной партии) может сделать противник. Каждый из $|M|$ игроков, отвечающих за допустимые тройки $\langle p, q, r\rangle$, делает не более 2^{2n+r} ходов, объявляя слова r-простыми, а также не более 2^{q+p} ходов, объявляя слова p-простыми. Поэтому общее число ходов этих игроков не больше

$$|M| \cdot (2^{\max(2n+r)} + 2^{\max(q+p)})$$

(где максимумы в показателе берутся по всем тройкам из M). Сюда ещё надо добавить 2^{2n-1} ходов при объявлении слов плохими и $2^{2n} \cdot 2^{n-2} = 2^{3n-2}$ ходов при объявлении слов плохими относительно других слов. Таким образом, обещанное условие на число ходов противника (меньше 2^{3n-1}) будет выполнено, если

$$|M| \cdot (2^{\max(2n+r)} + 2^{\max(q+p)}) + 2^{2n-1} + 2^{3n-2} < 2^{3n-1}. \qquad (**)$$

Учитывая, что $2^k + 2^l$ близко к $2^{\max\{k,l\}}$, легко заметить, что условия $(*)$ и $(**)$ будут выполнены, если все тройки $\langle p, q, r\rangle \in M$ удовлетворяют неравенствам

$$\begin{aligned} p + r &< 2n - 3\log n - O(1),\\ 2n + r &< 3n - 3\log n - O(1),\\ p + q &< 3n - 3\log n - O(1). \end{aligned}$$

(Таких троек не более $O(n^3)$, и мы как раз вычли $3\log n + O(1)$ как верхнюю оценку для $\log|M|$). Мы не указываем явно константу в $O(1)$ но, скажем, заведомо годится 10. Включим в M все тройки, удовлетворяющие этим неравенствам; в этом случае в описанной игре существует выигрышная стратегия.

Эту стратегию мы будем применять против стратегии противника, которая игнорирует делаемые нами ходы. Она объявляет плохими слова длины $2n$, имеющие сложность менее $2n - 1$, объявляет плохими относительно слова B все слова, имеющие условную сложность менее $n - 2$, а также объявляет u-простыми [относительно простыми] все слова, сложность [соответственно, относительная сложность] которых меньше u. (Точнее говоря, этим правилом руководствуется каждый $\langle p, q, r\rangle$-игрок для каждой тройки $\langle p, q, r\rangle \in M$.)

Запустим две описанные нами стратегии играть друг против друга. Это будет алгоритмический процесс, для задания которого достаточно знать число n. Момент

его окончания алгоритмически найти нельзя, но он существует, и в этот момент найдётся пара слов $\langle A, B\rangle$, которая обеспечивает наш выигрыш. Покажем, что для них выполнено утверждение теоремы.

Длина слов A и B равна $2n$; сложность равна $2n + O(1)$ (она не может быть меньше, так как иначе эти слова были бы объявлены плохими). Поскольку в нашем процессе мы для каждого B указываем не более 2^n различных слов A, то $KS(A \mid B) \leqslant n + O(\log n)$; сложность $KS(A \mid B)$ не может быть меньше $n - O(1)$, иначе слово A было бы объявлено плохим относительно B.

Пусть дано произвольное слово B', для которого $KS(B' \mid B) < r$ и длина слова B' равна q. Можно считать, что r много меньше n, скажем, $r < n/2$ (иначе слагаемое $O(r)$ в правой части утверждения теоремы делает его тривиальным). Тогда второе неравенство из трёх выполнено. Поэтому для любого p, удовлетворяющего первому и третьему неравенству, сложность A относительно B' больше p. Значит для любой пары $\langle p, q\rangle$ из светло-серой области на рис. 12.14, отстоящей от её границы с областью G не менее, чем на $O(r + \log n)$, не существует B' длины q для которого $KS(B' \mid B) < r$ и $KS(A \mid B') < p$. Это и требовалось установить.

Теорема 237 доказана. ▶

Приведём другое доказательство того же утверждения, использующее вероятностный метод. Оно во многом параллельно игровому[1], но есть важные различия. Во-первых, вместо того, чтобы реагировать на ходы противника, мы сделаем все ходы сразу же и будем надеяться, что при любых действиях противника мы выиграем. Во-вторых, мы теперь не указываем выигрышные ходы явно, а доказываем, что случайный набор ходов с положительной вероятностью окажется выигрышным.

Техническое замечание: при таком подходе можно без ограничения общности считать, что оба игрока указывают максимальное разрешённое число слов.

Перейдём к формальному изложению, отмечая параллели с игровым доказательством в квадратных скобках. Пусть имеется отображение

$$U\colon \mathbb{B}^{2n} \times \mathbb{B}^{n} \to \mathbb{B}^{2n}$$

[значения $U(B, X)$ при данном B и всевозможных X соответствуют n-простым относительно B словам в игровом доказательстве]. Пусть фиксировано некоторое конечное множество M троек натуральных чисел и для каждой тройки $\langle p, q, r\rangle \in M$ имеются два отображения

$$V_{p,q,r}\colon \mathbb{B}^{2n} \times \mathbb{B}^{r} \to \mathbb{B}^{q}$$

и

$$W_{p,q,r}\colon \mathbb{B}^{q} \times \mathbb{B}^{p} \to \mathbb{B}^{2n}.$$

Пусть, кроме того, имеется отображение

$$S\colon \mathbb{B}^{2n-2} \to \mathbb{B}^{2n},$$

а также отображение

$$T\colon \mathbb{B}^{2n} \times \mathbb{B}^{n-2\log n} \to \mathbb{B}^{2n}.$$

[1] Читатель мог уже заметить, что многие доказательства в этой книге имеют игровой характер. Более подробно об игровом методе в теории колмогоровской сложности рассказано в обзорах [171, 117].

[Отображения $V_{p,q,r}$ и $W_{p,q,r}$ соответствуют ходам $\langle p, q, r\rangle$-игрока в команде противника. Именно, слова $V_{p,q,r}(B, X)$ при всевозможных X суть r-простые относительно B слова; слова $W_{p,q,r}(B', X)$ при всевозможных X суть p-простые относительно B' слова. Отображения S и T соответствуют ходам двух дополнительных игроков. Именно, $S(X)$ суть плохие слова в количестве не более 2^{2n-2} штук, а $T(B, X)$ суть плохие относительно B слова в количестве не более чем $2^{n-2\log n}$ штук. Границы для числа плохих слов уменьшены по сравнению с игровым доказательством: вместо $2n-1$ взято $2n-2$, а вместо $n-2$ взято $n-2\log n$. Это будет использовано в оценках.]

Будем говорить, что отображение U покрыто четвёркой V, W, S, T (состоящей из двух семейств отображений и двух отображений), если для всякого слова $B \in \mathbb{B}^{2n}$ и для всякого слова A, равного $U(B, X)$ при некотором $X \in \mathbb{B}^n$ [для любого слова B и любого слова A, объявленного нами n-простым относительно B], пара $\langle A, B\rangle$ удовлетворяет одному из четырёх условий:

(1) Слово B входит в область значений отображения S (то есть $B = S(Y)$ при некотором $Y \in \mathbb{B}^{2n-2}$). [Слово B противник объявил плохим.]

(2) Слово A входит в область значений отображения S (то есть $A = S(Y)$ при некотором $Y \in \mathbb{B}^{2n-2}$). [Слово A противник объявил плохим.]

(3) Слово A равно $T(B, Y)$ при некотором $Y \in \mathbb{B}^{n-2\log n}$. [Слово A противник объявил плохим для данного B.]

(4) Найдётся тройка $\langle p, q, r\rangle \in M$ и слово B' длины q, для которых

(а) $B' = V_{p,q,r}(B, Y)$ для некоторого $Y \in \mathbb{B}^r$ [$\langle p, q, r\rangle$-игрок команды противника объявил, что слово B' является r-простым относительно B];

(б) $A = W_{p,q,r}(B', Z)$ для некоторого $Z \in \mathbb{B}^p$ [$\langle p, q, r\rangle$-игрок команды противника объявил, что слово A является p-простым относительно B'].

Мы докажем (при некотором условиях на множество M, указанных далее), что существует отображение U, не покрытое ни одной четвёркой V, W, S, T. Доказательство вероятностное: мы подсчитаем для данной четвёрки, сколько отображений U ими покрывается (т.е. вероятность для случайного отображения U оказаться покрытым), умножим эту вероятность на число четвёрок и установим, что произведение останется меньше 1.

Подсчёт для одной четвёрки делается так. Пусть фиксированы V, W, S, T. Имеется не менее 2^{2n-1} слов B длины $2n$, нарушающих условие (1). Чтобы U было покрыто, необходимо, чтобы при любом таком B каждое из 2^n значений $A = U(B, X)$ для всех X длины n было покрыто одним из условий (2)–(4). Мы убедимся, что для данных B и X вероятность такого события не больше $1/2$. Из независимости следует, что вероятность для случайного U оказаться покрытым не больше

$$(1/2)^{2^{2n-1}\times 2^n} = (1/2)^{2^{3n-1}}.$$

Для данных B и X: не годятся слова A, покрытые условием (2) в количестве 2^{2n-2} штук, покрытые условием (3) в количестве $2^{n-2\log n}$ штук, а также покрытые условием (4) слова $W_{p,q,r}(V_{p,q,r}(B, Y), Z)$ в количестве $2^r \times 2^p$ штук для каждой тройки $\langle p, q, r\rangle \in M$. Всего получается не больше

$$2^{2n-2} + 2^{n-2\log n} + 2^{r+p} \cdot |M|,$$

что не больше 2^{2n-1} (половины всех слов), если

$$r + p + \log|M| < 2n - 3 \qquad (*)$$

при всех $\langle p, q, r\rangle \in M$ (это и будет первое из списка требований к M).

Осталось оценить количество всех четвёрок V, W, S, T. Для $V_{p,q,r}$ (при данных p, q, r) имеется не более

$$(2^q)^{2^{2n}\times 2^r} = 2^{q\cdot 2^{2n+r}}$$

вариантов, для $W_{p,q,r}$ (при данных p, q, r) имеется не более

$$(2^{2n})^{2^q\times 2^p} = 2^{2n\cdot 2^{q+p}}$$

вариантов, для S имеется не более

$$(2^{2n})^{2^{2n-2}} = 2^{2n\cdot 2^{2n-2}}$$

вариантов, для T имеется не более

$$(2^{2n})^{2^{2n}\times 2^{n-2\log n}} = 2^{(2^{3n}/n)+1}$$

вариантов. Первые две оценки возводим в степень $|M|$ и получаем оценку на число вариантов для V и W в целом, после чего полученные оценки перемножаем и заключаем, что общее количество четвёрок V, W, S, T не превосходит

$$2^{q\cdot 2^{2n+r}\times|M|}\cdot 2^{2n\cdot 2^{q+p}\times|M|}\cdot 2^{2n\cdot 2^{2n-2}}\cdot 2^{(2^{3n}/n)+1},$$

а двоичный логарифм этого числа не превосходит

$$q\cdot 2^{2n+r}\times|M| + 2n\cdot 2^{q+p}\times|M| + 2n\cdot 2^{2n-2} + (2^{3n}/n) + 1,$$

что будет меньше 2^{3n} (как требуется для завершения доказательства), если

$$2n + r + \log q + \log|M| < 3n - O(1) \qquad (**)$$

и

$$q + p + \log n + \log|M| < 3n - O(1). \qquad (***)$$

(Мы пользуемся тем, что $2^a + 2^b$ отличается не более чем в константу раз от $2^{\max(a,b)}$; два других условия в логарифмической шкале имеют вид $2n - 2 + \log 2n < 3n - O(1)$ и $\log(2^{3n}/n + 1) < 3n - O(1)$, и выполнены автоматически.)

Все три условия $(*)-(***)$ будут заведомо выполнены, если

$$\begin{aligned} p + r &< 2n - 3\log n - O(1),\\ 2n + r &< 3n - 4\log n - O(1),\\ p + q &< 3n - 4\log n - O(1) \end{aligned}$$

для всех $\langle p, q, r\rangle$ из M, поскольку в таком случае $|M| = O(n^3)$.

Поэтому, взяв в качестве M множество троек, удовлетворяющих трём указанным только что неравенствам, мы заключаем, что существует отображение U, не покрытое никакой четвёркой V, W, S, T (при этом M).

При известном n такое отображение U можно найти перебором, поэтому (важный момент!) первое отображение U с такими свойствами будет иметь сложность не больше $\log n$.

Возьмём это U и конкретную четвёрку V, W, S, T, которой оно (по доказанному) не покрыто. А именно, пусть $\{S(\cdot)\}$ (множество значений отображения S) — все слова длины $2n$, имеющие сложность менее $2n-2$. Пусть при любом $B \in \mathbb{B}^{2n}$ множество $\{T(B,\cdot)\}$ составляют все слова, имеющие условную сложность (относительно B) менее $n - 2\log n$. Пусть при данных p, q, r среди $V_{p,q,r}(B,\cdot)$ встречаются все слова условной сложности (относительно B) менее r, а среди $W_{p,q,r}(B',\cdot)$ встречаются все слова условной сложности (относительно B') менее p.

Раз U не покрыто, найдётся пара слов A и B, не обладающая ни одним из свойств (1)–(4). Отсюда следует, что $KS(A) = 2n + O(1)$ (слово имеет длину $2n$ и не может иметь сильно меньшую сложность, иначе его покрывает S). Аналогично $KS(B) = 2n + O(1)$. Условная сложность $KS(A\,|\,B)$ равна $n + O(\log n)$: она не может быть больше, так как $A = U(B, X)$ для некоторого слова X длины n, а сложность U есть $O(\log n)$; она не может быть меньше, так как иначе было бы выполнено свойство (3). Наконец, ни при каких p, q и r не найдётся слова B' длины q, при котором $KS(B'\,|\,B) < r$ и $KS(A\,|\,B') < p$, иначе было бы выполнено свойство (4).

Далее рассуждаем как раньше (в игровом доказательстве); на этом вероятностное доказательство теоремы 237 заканчивается.

Наконец, можно привести и «геометрическую» конструкцию, доказывающую то же самое утверждение. (В отличие от задачи выделения общей информации, здесь геометрическое рассуждение даёт примерно те же оценки сложности.)

А именно, возьмём поле из 2^n элементов (или примерно такого количества, если мы хотим ограничиться вычетами по простому модулю) и рассмотрим двумерную плоскость над этим полем. Пусть $\langle A, B\rangle$ — случайная пара, состоящая из точки этой плоскости и проходящей через неё прямой. Тогда сложности как раз такие, как требуется в теореме 237.

Далее, пусть имеется слово B', для которого

$$KS(B'\,|\,B) \leqslant r, \quad KS(B') \leqslant q, \quad KS(A\,|\,B') \leqslant p. \qquad (*)$$

Мы хотим показать, что пара $\langle p, q\rangle$ находится в $O(r) + O(\log n)$ окрестности множества G, установив, что в противном случае пара $\langle A, B\rangle$ имела бы меньшую сложность. В самом деле, оценим количество пар $\langle A, B\rangle$, для которых выполняются условия $(*)$ при некотором B'. Каждое из 2^q слов B' задаёт два множества:

- те слова A (длины $2n$), при которых $KS(A\,|\,B') \leqslant p$ (обозначим это множество $U_{B'}$);
- те слова B (длины $2n$), при которых $KS(B'\,|\,B) \leqslant r$ (обозначим его $V_{B'}$).

Множество $U_{B'}$ содержит 2^p элементов (точнее, $O(2^p)$, но для простоты записи мы ограниченные множители опускаем). Множество $V_{B'}$ может иметь разный размер (это зависит от B'), но известно, что эти множества покрывают множество из 2^{2n} слов длины $2n$ не более чем в 2^r слоёв (для каждого B есть не более 2^r слов B', простых относительно него).

Нам надо доказать, что объединение всех $U_{B'} \times V_{B'}$ покрывает лишь небольшую часть всех пар (так что случайная пара в него не попадёт). Для оценки количества покрытых пар применим уже известное свойство графа инцидентности (две прямые не могут проходить через две точки) и вытекающую из него оценку (лемма о четырёхугольниках, с. 398).

Нам будет удобно использовать утверждение этой леммы в такой форме (равносильной, как легко проверить): если в прямоугольной таблице $l \times L$ расставлены звёздочки, и никакие четыре не стоят на пересечении двух строк и двух столбцов, то число звёздочек не превосходит

- $O(L)$ при $l \leqslant \sqrt{L}$;
- $O(l\sqrt{L})$ при $l \geqslant \sqrt{L}$.

При этом следует отдельно рассматривать случай «больших» и «малых» $V_{B'}$. Начнём с первого. Если $V_{B'}$ содержит больше $\sqrt{|U_{B'}|}$, то есть больше $2^{p/2}$ элементов, то число покрытых пар не больше $2^{p/2}|V_{B'}|$ элементов для данного B'. Сумма по всех B' не больше $O(2^{p/2}2^{2n}2^r)$ (множество размера 2^{2n} покрыто не более чем в 2^r слоёв).

Теперь перейдём к малым $V_{B'}$. Для них $U_{B'} \times V_{B'}$ покрывает не более $O(2^p)$ элементов, и всего для 2^q различных B' получается $O(2^{p+q})$.

Таким образом, если $p + q < 3n - O(\log n)$ и $(p/2) + 2n + r < 3n - O(\log n)$, то случайная пара $\langle A, B\rangle$ не будет обслужена ни одним из слов B'. (Тут ещё надо заметить, что множество обслуженных пар можно перечислять, зная n, p, q, r, то есть $O(\log n)$ битов информации.) Второе неравенство можно переписать как $p + 2r < 2n$; хотя это и немного хуже, чем неравенство $p + r < 2n$, которое было в первом доказательстве, но всё равно мы остаёмся в пределах оценки $O(r)$, упомянутой в теореме.

Третье доказательство теоремы 237 закончено.

Замечание. Это третье доказательство позволяет получить простое множество пар, большинство пар в котором удовлетворяют теореме (другими словами, позволяют получить стохастическую в смысле раздела 14.2 пару, удовлетворяющую условиям теоремы).

Того же самого, хотя и не в столь наглядной форме, можно добиться и модификацией второго (вероятностного) доказательства. Мы считали отображение U покрытым, если для всех пар определённого вида нечто верно; ослабим это условие и будем говорить, что U покрыто, если для половины пар нечто верно.

Для доказательства существования непокрытого U можно воспользоваться следующей (тривиальной) оценкой. Если вероятность каждого из независимых 2^k событий не больше $1/16$, то вероятность того, что произойдёт не менее 2^{k-1} событий, не больше $2^{2^k} \cdot (1/16)^{2^k/2} = 2^{-k}$. Поэтому нужно заменить вероятность $1/2$ на $1/16$, а всё остальное как раньше.

(Можно и иначе: не рассматривать множеств S и T и параллельно все допустимые тройки, а доказывать, что с близкой к единице вероятностью для случайно выбранного U доля пар, для которых выполнено условие (4), мала. Эти малые доли и малые отклонения от единицы затем складываются для всех троек из M.)

13. Информация и логика

13.1. Задачи, логические операции, сложность

Под *задачей* будем понимать произвольное множество двоичных слов, а под *решением* задачи — любой его элемент. Смысл этого определения можно пояснить следующим образом. Решение любой точно поставленной задачи можно записать в подходящем формальном языке. Можно считать, что решение задачи закодировано при помощи некоторого двоичного слова. Мы будем интересоваться лишь количеством информации в решениях задачи, а не содержательной её стороной, поэтому задачу естественно отождествлять с множеством её решений.

Определим сложность задачи X как наименьшую сложность её решений:

$$KS(X) = \min\{KS(x) \mid x \in X\}.$$

При этом минимум пустого множества (задачи, не имеющей решений) полагаем равным $+\infty$.

Например, сложность синглетона $\{x\}$ совпадает со сложностью самого x. Другой пример: в разделе 1.2 мы сталкивались с задачей описания натурального числа, не меньшего заданного числа n. Её сложность обозначалась через $KS_{\geqslant}(n)$. Используя новую терминологию, можно сказать, что $KS_{\geqslant}(n)$ есть сложность задачи, множество решений которой состоит из всех чисел, больших или равных n.

Если есть две задачи X и Y, то можно рассмотреть задачу «решить *обе* задачи X и Y», а также задачу «решить *хотя бы одну* из задач X и Y (указав при этом, какая именно из двух задач решена)». Другими словами, решениями задачи «X и Y» можно считать пары, первая компонента которых есть решение задачи X, а вторая — задачи Y. А решениями задачи «X или Y» можно считать решения любой из задач с дополнительным указанием, какая именно из двух задач решена. Таким образом, мы приходим к таким определениям логических операций над задачами:

$$X \wedge Y = \{[x, y] \mid x \in X, y \in Y\},$$
$$X \vee Y = \{[0, x] \mid x \in X\} \cup \{[1, y] \mid y \in Y\}.$$

Здесь $[x, y]$ обозначает код пары слов $\langle x, y\rangle$ при некотором вычислимом однозначном кодировании пар слов.

Сложность задачи $\{x\} \wedge \{y\}$ равна сложности пары $\langle x, y\rangle$, а сложность задачи $\{x\} \vee \{y\}$ равна минимуму из сложностей x и y (с точностью до константы). И вообще, для любых задач X, Y сложность задачи $X \vee Y$ равна наименьшей из сложностей задач X и Y (с $O(1)$-точностью).

Задача $X \wedge Y$ называется *конъюнкцией* задач X и Y, а задача $X \vee Y$ — *дизъюнкцией* задач X и Y. Есть альтернативное естественное определение дизъюнкции задач, отражающее следующую идею. Можно предложить решающему написать два слова, из которых первое должно быть решением первой задачи или второе — решением второй задачи, при этом решающий может и не указывать, какой из двух случаев имеет место (что-то похожее происходит на письменных экзаменах по математике, когда требуется решить хотя бы одну задачу из двух, студент записывает два решения двух задач, и одно из них оказывается правильным, но студент не обязан указывать, какое). Этой идее соответствует такое формальное определение «псевдодизъюнкции» задач:

$$X \tilde{\vee} Y = \{[x, y] \mid x \in X \text{ или } y \in Y\}.$$

Сложность псевдодизъюнкции отличается от сложности дизъюнкции всего лишь на $O(1)$, однако это принципиально разные задачи, о чём мы поговорим позже.

$\boxed{\mathbf{332}}$ Докажите, что $KS(X \tilde{\vee} Y) = KS(X \vee Y) + O(1)$.

Другой (промежуточный) вариант «псевдодизъюнкции» получится, если просто объединить множества решений задач X и Y. (Студент пишет только одно решение, но не указывает номер задачи, к которой оно относится.)

Условную сложность $KS(y\,|\,x)$ тоже можно понимать как сложность некоторой алгоритмической задачи, а именно задачи преобразования x в y. Эта задача получается из задач $\{x\}$ и $\{y\}$ помощью операции над задачами, называемой *импликацией.* А именно, решениями задачи $X \to Y$ будут программы, которые дают некоторое решение задачи Y в применении к любому решению задачи X. Точнее говоря, зафиксируем некоторый язык программирования, в котором программами являются двоичные слова, и будем обозначать $[p](x)$ результат применения программы p к слову x. Если вычисление не заканчивается, то $[p](x)$ не определено. Язык программирования должен быть универсальным, то есть иметь программы для всех вычислимых функций. Более того, мы предполагаем, что он допускает эффективную трансляцию с любого другого языка программирования (задаваемая им универсальная функция должна быть гёделевой, или главной, см. [175]).

Итак, положим

$$X \to Y = \{p \mid \forall x\ (x \in X \Rightarrow [p](x) \text{ определено и } [p](x) \in Y)\}.$$

Например, в разделе 6.4 мы изучали величину $KS(x\,|\!\geqslant n)$, которая определялась как наименьшая сложность программы, которая выдает x, получив на вход любое натуральное число, не меньшее n. Используя новые обозначения, можно написать $KS(x\,|\!\geqslant n) = KS(\{m \in \mathbb{N} \mid m \geqslant n\} \to \{x\})$. Другой пример, с которого мы начали: сложность задачи $\{x\} \to \{y\}$ равна условной сложности y при известном x (с точностью до константы).

Некоторые теоремы из главы 12 естественным образом могут быть сформулированы как утверждения о сложности тех или иных задач.

А именно, решениями задачи $(x \to y) \wedge (y \to x)$ (мы опускаем скобки при записи синглетонов) являются пары программ, из которых первая преобразует x в y, а

вторая — y в x. Мы рассматривали задачу о минимальной возможной сложности такой пары и установили, что она равна максимальной из условных сложностей $KS(y\,|\,x)$ и $KS(x\,|\,y)$ (с точностью до $O(\log KS(x,y))$).

Другой пример. Решениями задачи $(x \to z) \wedge (y \to z)$ являются пары программ, из которых первая преобразует x в z, а вторая — y в z. Мы доказали, что минимально возможная сложность такой пары примерно равна максимуму условных сложностей $KS(z\,|\,x)$ и $KS(z\,|\,y)$ (с точностью до $O(\log KS(x,y,z))$).

$\boxed{\textbf{333}}$ Найдите сложность задачи $a \to (b \to c)$. [Указание. Эта задача эквивалентна $(a \wedge b) \to c$.]

$\boxed{\textbf{334}}$ Найдите сложность задачи $a \wedge (b \to c)$ (с логарифмической точностью). [Ответ: $KS(a) + KS(c\,|\,a,b)$. Указание: имея программу для a и программу получения c из a и b, можно указать и программу получения c из b. Неравенство в другую сторону: прибавив к предлагаемому ответу $KS(b\,|\,a)$, мы получим $KS(a,b,c)$, поэтому достаточно доказать, что по a, программе переделки a в b и программе переделки b в c можно восстановить тройку (a,b,c).]

$\boxed{\textbf{335}}$ Докажите, что сложность задачи $(x \vee y) \to (x \mathbin{\tilde{\vee}} y)$ ограничена константой, однако сложность обратной задачи $(x \mathbin{\tilde{\vee}} y) \to (x \vee y)$ такая же, как и у задачи $x \vee y$ с точностью до $O(\log n)$, где n — наибольшая из длин x, y. [Указание. Пусть p — любое решение задачи $(x \mathbin{\tilde{\vee}} y) \to (x \vee y)$. Будем говорить, что пара слов (u, v) длины не больше n согласована с p, если для всех слов w длины не больше n программа p в применении к обоим словам $[u, w]$, $[w, v]$ дает некоторое решение задачи $u \vee v$. Тогда для любой согласованной с p пары $u = x$ или $v = y$.]

$\boxed{\textbf{336}}$ Докажите, что сложность задачи

$$((x \mathbin{\tilde{\vee}} y) \to (x \vee y)) \to (x \mathbin{\tilde{\vee}} y)$$

есть $O(\log n)$, где n — наибольшая из длин x, y.

Две последних задачи объясняют разницу между обычной дизъюнкцией и псевдодизъюнкцией. В частности, из них следует, что задачи $x \mathbin{\tilde{\vee}} x$ и x существенно различны, хотя они и имеют почти одинаковые сложности. А сложность задачи $(x \mathbin{\tilde{\vee}} x) \to x$ оказывается примерно равной сложности самого x.

13.2. Сложность задач и интуиционистская логика

Импликация обладает следующим свойством: если задачи X и $X \to Y$ просты, то и задача Y проста:

$$KS(Y) \leqslant KS(X) + KS(X \to Y).$$

Это неравенство выполняется с точностью $O(\log KS(X))$, а также и с точностью $O(\log KS(X \to Y))$. Оно является обобщением неравенства

$$KS(y) \leqslant KS(x) + KS(y\,|\,x),$$

верного для любых слов x, y (также с логарифмической точностью). Как и неравенство для слов, оно может быть усилено:

$$KS(X \wedge Y) \leqslant KS(X) + KS(X \to Y).$$

Однако обратное неравенство уже неверно (в отличие от аналогичного неравенства для слов).

337 Приведите пример множеств X, Y, для которых $KS(X \wedge Y)$ значительно меньше, чем $KS(X) + KS(X \to Y)$. [Указание. В качестве X можно взять множество всех случайных слов длины n, а в качестве Y — множество всех случайных слов длины $2n$.]

Операции $\wedge, \vee, \to$ как операции над задачами восходят к определениям Колмогорова [61] и Клини [60], введённым для интерпретации интуиционистского исчисления высказываний (IPC); сложность задачи в описанном выше смысле рассматривалась в [157]. Читателя, желающего больше узнать о логических исчислениях, в том числе об IPC, мы отсылаем к книге [174]. А здесь мы обсудим связь между выводимостью формулы в интуиционистском исчислении высказываний и максимальной возможной сложностью порождённой ею задачи.

Пусть $\Phi(p, q, \ldots)$ — пропозициональная формула со связками $\wedge, \vee, \to$. Для любых множеств слов X, $Y, \ldots$ рассмотрим задачу $\Phi(X, Y \ldots)$, которая получится, если подставить в формулу Φ вместо переменных эти множества.

Сейчас будет вполне уместным следующее замечание. В нашем определении операций над задачами имеется произвол: в выборе спаривающей функции $x, y \to [x, y]$, в выборе языка программирования и, наконец, в выборе слов $0, 1$ при определении дизъюнкции. Однако сложности любой задачи, посчитанные при двух различных выборах, отличаются не более чем на константу. Точнее надо сказать так. Пусть $\Phi(p, q, \ldots)$ — пропозициональная формула, а X, $Y, \ldots$ — произвольные множества. Пусть $\Phi'(X, Y, \ldots)$ и $\Phi''(X, Y, \ldots)$ — две задачи, которые получились, если использовать разные спаривающие функции $[x, y]'$, $[x, y]''$, разные языки программирования $U'(p, x) = [p]'(x)$ и $U''(p, x) = [p]''(x)$, и, наконец, разные пары слов a', b' и a'', b'' в определении дизъюнкции. Тогда разность сложностей задач $\Phi'(X, Y, \ldots)$ и $\Phi''(X, Y, \ldots)$ ограничена константой по абсолютной величине. Доказать это можно по индукции. Точнее, по индукции для любой формулы можно построить две вычислимых функции: f_{12}^{Φ} преобразует любое решение задачи $\Phi'(X, Y, \ldots)$ в некоторое решение задачи $\Phi''(X, Y, \ldots)$, а f_{21}^{Φ} — любое решение задачи $\Phi''(X, Y, \ldots)$ в некоторое решение задачи $\Phi'(X, Y, \ldots)$.

Для пропозициональных переменных обе функции — это тождественная функция. Если формула Φ является конъюнкцией формул Ψ и Θ, для которых обе функции уже определены, то f_{12}^{Φ} действует так: данное слово s представляем в виде $[u, v]'$, затем применяем к u и к v функции f_{12}^{Ψ} и f_{12}^{Θ}, соответственно; полученные результаты спариваем с помощью второй спаривающей функции. Функция f_{21}^{Φ} действует аналогично. Аналогично рассматривается случай дизъюнкции.

Наконец, если формула Φ есть $\Psi \to \Theta$, то функция f_{12}^{Φ} определяется следующим образом. Рассмотрим вычислимую функцию $V(s, b) = f_{12}^{\Theta}([s]'(f_{21}^{\Psi}(b)))$. Если s — решение задачи $\Psi'(X, Y, \ldots) \to \Theta'(X, Y, \ldots)$, а b — решение задачи $\Psi''(X, Y, \ldots)$,

то $V(s,b)$ будет решением задачи $\Theta''(X, Y, \ldots)$. Будем рассматривать V как язык программирования; мы предполагали, что возможна эффективная трансляция этого языка в язык U'' (формально говоря, мы используем, что U'' — главная универсальная функция). Поэтому существует всюду определённая вычислимая функция $t\colon \Xi \to \Xi$, для которой $[t(s)]''(b) = V(s,b)$. Функция t и будет искомой функцией f_{12}^{Φ} для формулы Φ.

338 Проведите это рассуждение подробно.

Связь введённых понятий с интуиционистским исчислением высказываний обнаруживается в следующем факте: если формула $\Phi(p, q \ldots)$ выводима в интуиционистском исчислении высказываний, то сложность задачи $\Phi(X, Y, \ldots)$ ограничена константой (не зависящей от $X, Y, \ldots$). Более того, в этом случае существует строка s, которая является решением задачи $\Phi(X, Y, \ldots)$ для всех множеств $X, Y, \ldots$. Это доказывается индукцией по длине вывода в IPC. Например, для формулы $p \to (q \to p)$ (одна из аксиом IPC) эта строка есть следующая программа: «преобразовать данное нам x в программу, которая печатает x на любом входе».

339 Завершите это рассуждение и докажите, что для каждой выводимой в IPC формулы $\Phi(p, q \ldots)$ существует строка s, которая является решением задачи $\Phi(X, Y, \ldots)$ для всех множеств $X, Y, \ldots$.

Верно и обратное: если формула $\Phi(p, q, \ldots)$ *без отрицаний* не выводится в IPC, то сложность задачи $\Phi(X, Y, \ldots)$ не ограничена константой. Более того, справедлива следующая теорема о полноте для формул без отрицаний:

Теорема 238. *Пусть пропозициональная формула $\Phi(t_1, \ldots, t_k)$ со связками $\wedge, \vee, \to$ (без отрицаний) невыводима в IPC. Тогда найдётся положительное ε и последовательность непустых конечных множеств $X_1^n, \ldots, X_k^n$ (для $n = 1, 2, \ldots$), состоящих из слов длины не более n, для которой сложность задачи $\Phi(X_1^n, \ldots, X_k^n)$ не меньше εn при всех достаточно больших n.*

Это — главное утверждение настоящей главы. Мы его докажем, приняв на веру некоторое чисто логическое утверждение о невыводимых в IPC формулах. В этом доказательстве важную роль играют оценки сложности задач, которые возникают при подстановке одноэлементных множеств в (невыводимые в IPC) пропозициональные формулы. Некоторые из таких задач интересны сами по себе (см. уже приведённые примеры). Сейчас мы займемся анализом сложности некоторых других формул, представляющих интерес с логической точки зрения.

13.3. Примеры

Начнём с формулы[1]

$$((p \to q) \to p) \to p.$$

Эта формула, называемая законом Пирса (Peirce's Law), является тавтологией (истинна при любых булевых значениях переменных), но невыводима в интуиционистском вычислении высказываний. При этом сложность задачи $((x \to y) \to x) \to x$

[1] Примеры этого раздела взяты из работы [157]

(как обычно, мы опускаем скобки при записи синглетонов) есть $O(\log n)$ для любых слов x, y длины не более n (теорема 239 ниже). С другой стороны, как мы уже говорили, взяв интуиционистски невыводимую формулу и подставляя вместо переменных подходящие конечные непустые множества слов длины не более n, можно получить задачу сложности не меньше εn. В частности, это верно и для закона Пирса. Тут нет противоречия: сложность задачи $((X \to Y) \to X) \to X$ мала по сравнению со сложностями X, Y, если X, Y — синглетоны, но может быть сравнимой со сложностями X, Y, если X, Y — непустые конечные множества.

Теорема 239. *Сложность задачи $((x \to y) \to x) \to x$ есть $O(\log n)$ для любых слов x, y длины не более n.*

◄ Нам достаточно указать алгоритм, который по n и любому решению задачи $(x \to y) \to x$ находит x. Этот алгоритм работает так. Пусть p — решение задачи $(x \to y) \to x$. Пусть S есть обозначает множество всех слов длины не больше n. Для любой всюду определённой функции $\tau\colon S \to S$ (их конечное число) зафиксируем некоторую программу l_τ, вычисляющую эту функцию. Назовем пару $(u, v) \in S \times S$ *согласованной с* p, если $[p](l_\tau) = u$ для всех $\tau\colon S \to S$ с $\tau(u) = v$. Исходная пара (x, y) согласована с p. Имея p и n, мы начинаем перечислять согласованные с p пары и выдаем первую компоненту u первой попавшейся пары (u, v). Докажем, что в самом деле $u = x$. Действительно, в противном случае существует функция τ, для которой $\tau(x) = y$ и $\tau(u) = v$, и программа p на входе l_τ выдаёт одновременно два различных слова u и x. ►

С формальной точки зрения в предыдущем доказательстве есть небольшой пробел. А именно, мы неявно использовали, что имеется алгоритм который находит программу l_τ, имея функцию τ (заданную списком своих значений). Это следует из предположений о нашем языке программирования (список значений тоже можно рассматривать как программу для функции в некотором другом языке программирования, и этот второй язык можно эффективно транслировать в наш по предположению). Заметим ещё, что в наших рассуждениях можно было рассматривать не все функции, а только функции некоторого класса: важно лишь, что для любых двух пар (аргумент, значение) с различными аргументами найдётся функция из класса, через них проходящая. Например, годятся линейные функции $ax + b$ над конечным полем.

$\boxed{340}$ Докажите, что в условии предыдущей теоремы можно заменить $O(\log n)$ на $O(\log k)$, где k — максимальная из сложностей x, y. [Указание. Надо перейти от слов x, y к их программам длины не более k относительно оптимального декомпрессора. В доказательстве следует заменить S на множество всех слов длины не больше k, а программа l_τ должна работать так: на входе u она находит первое описание слова u длины не более k, применяет к нему τ и применяет оптимальный декомпрессор к получившемуся слову.]

Из доказанного утверждения и неравенства

$$KS(Y) \leqslant KS(X) + KS(X \to Y) + O(\log KS(X))$$

следует, что $KS(x) \leqslant KS((x \to y) \to x) + O(\log n)$ (где n — наибольшая из длин x, y). Но это не единственный способ доказывать неравенства подобного вида: суще-

ствуют такие формулы $A(p,q)$ и $B(p,q)$, что для любых слов x, y сложность задачи $B(x,y)$ не превосходит сложности задачи $A(x,y)$, однако сложность импликации $A(x,y) \to B(x,y)$ велика по сравнению со сложностями x, y (при некоторых x, y). Вот пример таких формул: $(x \to y) \to y$ и $x \vee y$. (Эти формулы, кстати, эквивалентны в классической логике.) Их сложности отличаются всего лишь на $O(\log n)$, что мы сейчас докажем, однако сложность импликации $((x \to y) \to y) \to (x \vee y)$, как мы увидим, может быть примерно равна n.

Начнем с вычисления сложности задачи $(x \to y) \to y$.

Теорема 240. *Сложность задачи $(x \to y) \to y$ равна сложности задачи $x \vee y$ (с точностью $O(\log n)$, где n — наибольшая из длин x, y).*

Эта сложность, как мы знаем, равна минимуму сложностей слов x и y с точностью $O(1)$.

◀ Нам достаточно предъявить алгоритм, который по любому решению задачи $x \vee y$ даёт некоторое решение задачи $(x \to y) \to y$, и другой алгоритм, который по любому решению задачи $(x \to y) \to y$ и логарифмическому от n количеству дополнительной информации даёт некоторое решение задачи $x \vee y$.

Первый алгоритм, получив на вход $[0, x]$ или $[1, y]$, должен найти некоторую программу, которая преобразует любое решение задачи $(x \to y)$ в y. Если нами получено слово $[1, y]$, то мы выдаём такую программу: не читая входа, печатать y. Если же нам дано слово $[0, x]$, то мы выдаём следующую программу: применяем данное нам решение задачи $(x \to y)$ к x, получая тем самым y, и печатаем его.

Теперь предъявим алгоритм, который по любому решению задачи $(x \to y) \to y$, числу n и ещё одному дополнительному биту информации находит некоторое решение задачи $x \vee y$.

Пусть p — любое решение задачи $(x \to y) \to y$. Обозначим через S множество всех слов длины не более n. Для любой функции $\tau \colon S \to S$ можно алгоритмически найти программу l_τ, вычисляющую эту функцию. На этот раз назовём пару $(u, v) \in S \times S$ *согласованной с p*, если $[p](l_\tau) = v$ для всех τ таких, что $\tau(u) = v$.

По определению пара (x, y) согласована с p. Могут существовать и другие пары, согласованные с p. Однако у любых двух согласованных с p пар (u', v') и (u'', v'') совпадают первые или вторые компоненты. Действительно, если первые компоненты различны, то существует функция τ, отображающая u' в v', а u'' в v''. Применение p к l_τ должно давать одновременно и v', и v''. Поэтому вторые компоненты должны совпадать, если не совпадают первые.

Множество согласованных с p пар можно перечислять, зная p и n. Запустим процесс его перечисления. Пусть (u, v) — первая появившаяся пара. Мы знаем, что либо $u = x$, либо $v = y$, но не знаем, какой из двух случаем имеет место. Именно здесь нам нужен дополнительный бит информации: получив его, мы выдаём x или y (в зависимости от значения этого бита). ▶

Это рассуждение можно обобщить и получить такой результат:

Теорема 241. *Сложность задачи $(x \to y) \to z$ (с точностью $O(\log n)$, где n — наибольшая из длин x, y) совпадает со сложностью задачи $z \vee (x \wedge (y \to z))$.*

Из утверждения задачи 334 следует, что сложность последней задачи равна минимуму из $KS(z)$ и $KS(x) + KS(z \,|\, x, y)$.

◄ Нам достаточно предъявить алгоритм, который по любому решению второй задачи даёт решение первой задачи, и другой алгоритм, который по любому решению первой задачи и логарифмическому от n количеству дополнительной информации даёт решение второй задачи.

Сначала предъявим первый алгоритм. По определению решением исходной задачи является любая программа, которая преобразует любое решение задачи $(x \to y)$ в z. Решения же второй задачи — это слово z, а также пары, состоящие из x и программы, переводящей y в z. Если данное нам решение второй задачи есть слово z, то мы выдаём следующее решение исходной задачи: не читая данной нам программы, печатать z. Если же данное нам решение второй задачи есть слово x и программа, преобразующая y в z, то мы выдаём следующее решение исходной задачи: применяем данное нам решение задачи $(x \to y)$ к x, получая тем самым y. Затем применяем программу, переводящую y в z, к найденному y.

Теперь предъявим алгоритм, который по любому решению исходной задачи, числу n и ещё одному дополнительному биту информации находит некоторое решение второй задачи.

Пусть p — любое решение задачи $(x \to y) \to z$. Обозначим через S множество всех слов длины не более n. Для любой функции $\tau\colon S \to S$ зафиксируем некоторую программу l_τ, вычисляющую эту функцию. Назовем тройку $(u, v, w) \in S \times S \times S$ *согласованной с* p, если $[p](l_\tau) = w$ для всех τ с $\tau(u) = v$.

Покажем, что либо у всех согласованных троек одинаковы первые компоненты, либо одинаковы третьи компоненты. Предположим, что есть две согласованные тройки с различными третьими компонентами: (u, v, w) и (u', v', w'), причём $w \neq w'$. Заметим, что тогда $u = u'$, иначе функция, которая переводит u в v и u' в v', должна одновременно переводиться в w и w', что невозможно. Теперь можно заметить, что и у всех остальных согласованных троек (если они есть) первая компонента та же, иначе эту тройку можно было бы совместить с любой из двух имеющихся и получить два разных значения третьей компоненты — противоречие.

По определению тройка (x, y, z) согласована с p. Множество согласованных с p троек можно перечислять, зная p и n. Запустим процесс его перечисления. Пусть (u, v, w) — первая появившаяся тройка. Мы уже знаем, что $u = x$ или $w = z$.

Первый случай: $w = z$. В этом случае мы (если нам скажут дополнительно, что этот случай имеет место) знаем z.

Второй случай: $u = x$. Тогда мы можем найти x, и остаётся решить задачу $y \to z$. Но эта задача проще, чем задача $(x \to y) \to z$, поскольку из y можно получить решение задачи $x \to y$.

Итак, имея решение задачи $(x \to y) \to z$, число n и один дополнительный бит, можно найти либо z, либо решение задачи $x \wedge (y \to z)$. (Дополнительный бит нам нужен, чтобы понять, какой из двух случаев имеет место.) ►

13.4. Ещё несколько примеров и доказательство теоремы 238

Теперь мы докажем что сложность задачи $((x \to y) \to y) \to (x \vee y)$ может быть близка к длине слов x, y, о чём уже говорилось ранее.

Теорема 242. *Сложность задачи* $((x \to y) \to y) \to (x \vee y)$ *для слов* x *и* y *длины не более* n *равна* $\min(KS(x\,|\,y), KS(y\,|\,x)) + O(\log n)$. *В частности, если* x, y *— случайные и независимые слова длины* n, *сложность этой задачи примерно равна* n.

◀ Легко убедиться, что сложность задачи $((x \to y) \to y) \to (x \vee y)$ не превосходит $KS(y\,|\,x)$ (с точностью $O(1)$): имея программу p, перерабатывающую x в y, а также любое решение q задачи $(x \to y) \to y$, можно найти y как $[p](q)$.

Докажем теперь, что сложность задачи $((x \to y) \to y) \to (x \vee y)$ не превосходит также $KS(x\,|\,y)$ (с точностью $O(\log n)$). Для этого достаточно установить, что имея тройку, состоящую из n, любой программы p, преобразующей y в x, и любого решения q задачи $(x \to y) \to y$, можно найти x или y. Пусть S обозначает множество всех слов длины не больше n. Как и в предыдущем доказательстве, назовём пару $(u, v) \in S \times S$ *согласованной с* q, если $[q](l_\tau) = v$ для всех $\tau \colon S \to S$ таких, что $\tau(u) = v$.

Напомним, что у любых двух согласованных с q пар совпадают первые или вторые компоненты. Из этого следует, что либо первые компоненты всех согласованных пар равны x, либо вторые компоненты всех согласованных пар равны y (либо и то, и другое). Действительно, пусть есть пара, у которой первая компонента отличается от x (и потому вторая компонента равна y). Обозначим её (x', y). Пусть есть и пара, у которой вторая компонента не равна y (и потому первая равна x); обозначим её (x, y'). Тогда у пар (x, y') и (x', y) различаются обе компоненты, в противоречии со сказанным выше.

Имея n, p и q, найдём первую пару (u, v), согласованную с q. Затем продолжим поиск других согласованных с q пар и одновременно запустим проверку того, что $[p](v) = u$. Либо то, либо другое обязательно произойдёт: если $[p](v)$ не определено или $[p](v) \neq u$, пара (u, v) отличается от (x, y) и потому есть более одной согласованной пары.

Если мы нашли другую пару (u', v'), согласованную с p, то при $v \neq v'$ мы знаем x (поскольку тогда $u = u' = x$), а при $u \neq u'$ знаем y (тогда $v = v' = y$). А если мы обнаружили, что $[p](v) = u$, то можем быть уверены, что $u = x$ (если $u \neq x$, то $v = y$ и $x = p(y) = p(v) = u$, противоречие).

Половина утверждения теоремы (неравенство в одну сторону) доказана. Осталось доказать, что сложность задачи $((x \to y) \to y) \to (x \vee y)$ не может быть существенно меньше $\min\{KS(y\,|\,x), KS(x\,|\,y)\}$. Пусть нам дана программа p, решающая задачу $((x \to y) \to y) \to (x \vee y)$. Имея p, мы построим пару программ (r_1, r_2) с таким свойством: либо r_1 на входе x выдаёт y, либо r_2 на входе y выдаёт x (но при этом мы не будем знать, какой из случаев имеет место). Другими словами, мы предъявляем решение задачи

$$(((x \to y) \to y) \to (x \vee y)) \to ((x \to y) \mathbin{\tilde{\vee}} (y \to x)),$$

и этого достаточно для оценки сложности.

Идею построения можно объяснить так. Пусть дана программа p, решающая задачу $((x \to y) \to y) \to (x \vee y)$. Нам надо либо преобразовать x в y, либо наоборот. Любое из слов x и y может быть использовано для построения решения

задачи $(x \to y) \to y$. В самом деле, зная y, мы можем отображать всё что угодно в y. А зная x, мы можем из решения задачи $x \to y$ получить y.

Получив (исходя из x или исходя из y) решение задачи $(x \to y) \to y$, мы можем применить к нему программу p и получить решение задачи $x \vee y$, то есть найти x или y. Было бы здорово, если бы p выдавала y, если мы начинали с x, или, наоборот, p выдавала x, если мы начинали с y. Но какие основания у нас на это рассчитывать?

Применим технику, известную из теории вычислимости, и рассмотрим пару A, B непересекающихся перечислимых неотделимых множеств. (Неотделимость означает, что не существует разрешимого множества, содержащего одно из множеств A, B и не пересекающегося с другим.) Например, можно взять в качестве A программы без входа, останавливающиеся с результатом 0, а в качестве B — останавливающиеся с результатом 1. (См., например, [175].) Удобно считать, что элементами A и B являются натуральные числа.

Для каждого числа i и для любых двух слов u, v рассмотрим программу $q_i(u,v)$, которая действует на входе s следующим образом. (Вход s мы рассматриваем как программу.)

$q_i(u,v)$ на входе s:
$i \in A$: выдать v;
$i \in B$: выдать $[s](u)$;
$[s](u) = v$: выдать v (равное $[s](u)$)

Имеется в виду, что программа $q_i(u,v)$ ждёт выполнения одного из трёх указанных (перед двоеточиями) условий, перечисляя A и B и вычисляя $[s](u)$. Заметим, что первые два условия несовместны друг с другом, а третье может пересекаться с первыми двумя — но в этом случае и указанный ответ один и тот же, так что всё согласовано.

Наше построение гарантирует, что:

(1) если $i \in A$, то для любого u программа $q_i(u,y)$ является решением задачи $((x \to y) \to y)$;

(2) если $i \in B$, то для любого v программа $q_i(x,v)$ является решением задачи $((x \to y) \to y)$;

(3) для любого i программа $q_i(x,y)$ является решением задачи $(x \to y) \to y$.

Таким образом, во всех трёх случаях применение программы p к $q_i(u,v)$ позволит получить решение задачи $x \vee y$.

Теперь мы можем описать обещанные программы r_1 и r_2 и доказать, что либо r_1 отображает x в y, либо r_2 отображает y в x, либо и то, и другое. Программа r_1 применяет p ко всем $q_i(x,v)$ при всех $i \in B$ и при всех v, ожидая появления ответа вида $[1,z]$ для некоторого z; как только такой ответ появится, программа останавливается и выдаёт z. Аналогичным образом программа r_2 применяет p ко всем $a_i(u,y)$ при всех $i \in A$ и при всех u, ожидая появления ответа вида $[0,z]$ при некотором z, и выдаёт это z.

Указанные выше свойства q_i гарантируют, что ответ программ r_1 и r_2, если только он появится, будет правильным. Остаётся доказать, что хотя бы в одном из случаем ответ-таки появится.

Пусть это не так. Это значит, что $[p](q_i(x,v))$ всегда начинается на 0 (при всех $i \in B$ и при всех v), а $[p](q_i(u,y))$ всегда начинается на 1 (при всех $i \in A$ и при всех u). В частности, такое происходит при $v = y$ и $u = x$. Вспомним, однако, что $[p](q_i(x,y))$ определено при всех i. Таким образом, первая компонента пары $[p](q_i(x,y))$ позволяет тогда отделить A от B разрешимым множеством, что противоречит предположению. ►

Формула, о которой идёт речь в только что доказанной теореме, и так довольно сложная. Но для доказательства теоремы 238 нам понадобиться ещё немного усложнить её, рассмотрев обобщения формул вида $((x \to y) \to y) \to (x \vee y)$. Пусть даны $k \geqslant 2$ слов, которые мы обозначим $u_1, \ldots, u_k$. Пусть также даны два непустых непересекающихся подмножества I, J множества индексов $\{1, \ldots, k\}$. Рассмотрим сначала задачу

$$((X \to Y) \to Y) \to Z,$$

где X есть конъюнкция синглетонов u_i по всем $i \in I$ (то есть, попросту говоря, кортеж, состоящий из этих слов), Y — дизъюнкция синглетонов u_j при $j \in J$, а Z — дизъюнкция всех синглетонов $u_1, \ldots, u_k$. Например, при $k = 2$, $I = \{1\}$, $J = \{2\}$ мы получаем уже рассмотренную задачу $((u_1 \to u_2) \to u_2) \to (u_1 \vee u_2)$. Другой пример (где для наглядности переменные обозначены разными буквами):

$$((x_1 \wedge x_2 \to y_1 \vee y_2) \to y_1 \vee y_2) \to x_1 \vee x_2 \vee y_1 \vee y_2 \vee z.$$

Теорема 243. *Сложность задачи $((X \to Y) \to Y) \to Z$ не меньше минимальной из условных сложностей слов u_i при известных остальных словах из $u_1, \ldots, u_k$ (с точностью до $O(1)$).*

◄ Доказательство будет аналогично предыдущему: мы построим k алгоритмов; при этом для любой программы p, решающей задачу $((X \to Y) \to Y) \to Z$, найдётся такое m, что m-й алгоритм по p и кортежу слов $u_1, \ldots, u_k$ с пропущенным словом u_m находит u_m.

Алгоритмы эти будут действовать по той же схеме: зная остальные (кроме u_m) слова, мы строим решения задачи $(X \to Y) \to Y$ и применяем к ним p. Заметим, что если мы знаем все слова за одним исключением, то знаем либо все переменные в X (и тогда знаем решение задачи X), либо все переменные в Y (и тогда знаем решение задачи Y — точнее, даже несколько решений, поскольку это дизъюнкция; нам достаточно было бы одного). Может даже случиться, что мы знаем и то, и другое (например, если пропущенная переменная не входит ни в X, ни в Y). Во всех случаях мы можем (как и раньше) построить решение задачи $(X \to Y) \to Y$ и применить к нему программу p. Результатом такого применения будет какое-то слово u_t (в паре с самим номером t), и если вдруг окажется, что $t = m$, то мы тем самым восстановили u_m. (Здесь мы считаем, что решениями задач Y и Z являются пары вида $[t, u_t]$; формально говоря, дизъюнкцию более чем двух задач мы не определяли явно, и надо было бы говорить о $(u_1 \vee (u_2 \vee (\ldots (u_{k-1} \vee u_k) \ldots)))$, но это по существу то же самое.)

Вопрос только в том, как гарантировать, что (для некоторого m) такое событие ($t = m$) произойдёт, и здесь нам снова понадобится техника из общей теории

вычислимости. Вместо перечислимых неотделимых множеств мы рассмотрим диагональную функцию: это такая вычислимая функция d, с которой любая вычислимая функция u совпадает хотя бы в одной точке: для всякой вычислимой функции u найдётся i, при котором $d(i) \simeq u(i)$, то есть либо $d(i) = u(i)$, либо оба значения $d(i)$ и $u(i)$ не определены. Строится такая функция d диагональной конструкцией: достаточно положить $d(i)$ равным значению i-й функции на i. (Мы рассматриваем натуральные числа как аргументы диагональной функции.)

Это построение можно выполнить для функций со значениями 0 и 1; тогда функция d задаётся двумя непересекающимися перечислимыми множествами (прообразами нуля и единицы), и невозможность построить всюду определённую функцию со значениями 0 и 1, всюду отличающуюся от d, означает, что эти прообразы неотделимы. Нам понадобится диагональная функция для случая k значений $1, 2, \ldots, k$.

Более точно, для любых слов $v_1, \ldots, v_k$ и для любого натурального i мы построим программу $q_i(v_1, \ldots, v_k)$ с такими свойствами:

(1) для исходных слов $u_1, \ldots, u_k$ программа $q_i(u_1, \ldots, u_k)$ является решением задачи $(X \to Y) \to Y$ при любом i;

(2) если $d(i) = m$, то $q_i(u_1, \ldots, u_k)$ останется решением задачи $(X \to Y) \to Y$ и после замены u_m на любое другое слово v_m (то есть значение $q_i(u_1, \ldots, u_{m-1}, v_m, u_{m+1}, \ldots, u_k)$ является решением задачи $(X \to Y) \to Y$ при любом v_m).

Теперь можно описать m-й алгоритм (пытающийся восстановить u_m по остальным u_t): мы запускаем вычисление d на всех входах; как только обнаруживается i с $d(i) = m$, мы применяем данную нам программу p ко всем решениям задачи $(X \to Y) \to Y$, имеющим вид $q_i(u_1, \ldots, u_{m-1}, *, u_{m+1}, \ldots, u_k)$ (вместо звёздочки мы подставляем всевозможные слова — в силу свойства (2) это будут действительно решения). Как только программа p на одном из таких входов выдаст пару вида $[m, *]$, мы останавливаем процесс: второй член этой пары и будет u_m.

Корректность ответа гарантируется свойством (2), но надо убедиться, что хотя бы при одном m такой ответ будет получен (ожидание не будет бесконечным). Тут и понадобится свойство диагональной функции. Пусть ни при одном m описанный процесс не даёт ответа. Тогда, в частности, искомого ответа не получится для $q_i(u_1, \ldots, u_k)$ при всех тех i, при которых $d(i)$ определено. Но в этом случае при всех i мы имеем дело с решением задачи $(X \to Y) \to Y$ в силу свойства (1). Следовательно, $[p](q_i(u_1, \ldots, u_k))$ определено при всех i и по предположению является решением задачи Z, то есть имеет вид $[t, u_t]$. Безуспешность поиска означает, что первый член пары t всегда отличается от $d(i)$, если $d(i)$ определено. Поскольку значение t является всюду определённой вычислимой функцией от i, то мы получаем противоречие с диагональным свойством.

Осталось описать действие программы $q_i(v_1, \ldots, v_k)$ (для произвольного числа i и произвольных слов $v_1, \ldots, v_k$). Получив на вход s, эта программа параллельно делает две вещи:

(1) пытается вычислить $d(i)$;

(2) строит из слов v_t с $t \in I$ предположительное решение задачи X, применяет к нему s и проверяет, что выход s имеет вид $[t, v_t]$ с $t \in J$ (вторая координата совпадает с элементом набора $v_1, \ldots, v_k$ с тем же индексом).

Как только одно из этих событий (включая положительный результат проверки во втором случае) произошло, программа предпринимает такие действия:

- в случае (1), если $d(i)$ определено и равно некоторому m:
  - если $m \notin I$, то строим из v_t с $t \in I$ предположительное решение задачи X, и выдаём результат применения программы s к этому решению в качестве ответа программы $q_i(v_1, \ldots, v_k)$ на входе s;
  - если $m \in I$ (и потому $m \notin J$), выдаём в качестве ответа пару $[t, v_t]$ для некоторого $t \in J$ — для определённости, скажем, для минимального. (При этом ответ не зависит от входа s.)
- в случае (2) программа выдаёт пару $[t, v_t]$, являющуюся результатом работы программы s.

Остаётся проверить требуемые свойства. Если слова $v_1, \ldots, v_k$ совпадают с $u_1, \ldots, u_k$, а программа s является решением задачи $(X \to Y) \to Y$, то второе событие произойдёт обязательно (если раньше не произойдёт первое); в обоих случаях выход программы $q_i(u_1, \ldots, u_k)$ будет решением задачи Z. Значит, в этом случае программа $q_i(u_1, \ldots, u_k)$ будет решением задачи $((X \to Y) \to Y) \to Z$.

Пусть теперь $d(i) = m$, все v_t, кроме (возможно) m-го, совпадают с соответствующими u_t, а s является решением задачи $(X \to Y) \to Y$. При этом может раньше случиться и то, и другое событие. Но в обоих случаях выход программы $q_i(v_1, \ldots, v_k)$ будет решением задачи Z. В самом деле, в случае (1) мы не используем значение v_m вообще. В случае (2) мы его можем использовать, но у нас два резона считать ответ правильным (в силу совпадения), и один из них сработает, поскольку лишь одно v_t отличается от u_t. ►

При доказательстве теоремы 238 нам понадобится небольшое обобщение доказанного утверждения. Пусть опять имеется кортеж слов $u = u_1, \ldots, u_k$. Пусть также имеется некоторое число (N) непересекающихся пар подмножеств множества индексов $\{1, \ldots, k\}$:

$$I_l \cap J_l = \varnothing, \quad I_l, J_l \subset \{1, \ldots, k\}, \; l = 1, \ldots, N.$$

Для каждого l определим задачи X_l и Y_l как раньше, то есть X_l есть конъюнкция синглетонов u_t для $t \in I_l$, а Y_l — дизъюнкция синглетонов u_t для $t \in J_l$. Наконец, пусть Z — дизъюнкция всех синглетонов $u_1, \ldots, u_k$. Рассмотрим теперь задачу

$$\big(((X_1 \to Y_1) \to Y_1) \wedge \ldots \wedge ((X_N \to Y_N) \to Y_N)\big) \to Z.$$

Теорема 244. *Сложность этой задачи не меньше минимума условных сложностей слов u_i при известных остальных словах из набора $u_1, \ldots, u_k$ (с точностью до $O(1)$).*

◄ Доказательство этой теоремы в основном повторяет предыдущее доказательство. Для каждого l, каждого натурального i и каждого кортежа $v = v_1, \ldots, v_k$,

отличающегося от кортежа u не более чем в одной координате, мы, как и раньше, строим программу $q_{li}(v)$, которая будет решением задачи

$$(X_l \to Y_l) \to Y_l,$$

если $d(i)$ определено и равно той координате, где кортежи u и v различаются, или если $v = u$ (в последнем случае неважно, определено ли и чему равно $d(i)$).

При любом m можно рассмотреть следующий алгоритм поиска u_m, если дана любая программа p решения задачи

$$\big(((X_1 \to Y_1) \to Y_1) \wedge \ldots \wedge ((X_N \to Y_N) \to Y_N)\big) \to Z,$$

а также даны все члены кортежа u, кроме m-го. Применяем программу p к кортежу, состоящему из программ $q_{1i}(v), \ldots, q_{Ni}(v)$. В качестве v пробуем все кортежи, совпадающие с u во всех координатах, отличных от m-й, а в качестве i — все такие натуральные числа, что $d(i) = m$. Если нам повезёт и для хотя бы для одной такой пары v, i программа p остановится и выдаст пару с первой компонентой m, то мы найдём u_m. Как и раньше, диагональное свойство функции d гарантирует, что существует m, для которого такие v, i обнаружатся (чтобы убедиться в этом, в качестве v мы берём сам кортеж u). ►

Пропозициональные формулы, соответствующие задачам, рассмотренным в этой теореме, называются *критическими импликациями.* А именно, так называются формулы вида

$$\big(((P_1 \to Q_1) \to Q_1) \wedge \ldots \wedge ((P_N \to Q_N) \to Q_N)\big) \to R,$$

где R есть дизъюнкция переменных $s_1, \ldots, s_k$, формулы P_l представляют собой конъюнкции, а Q_l — дизъюнкции каких-то из этих переменных, причём P_l и Q_l не имеют общих переменных при каждом l. Критические импликации невыводимы в IPC (что следует из только что доказанной теоремы, а также может быть доказано с помощью моделей Крипке). Более того, они являются универсальными невыводимыми формулами интуиционистского исчисления высказываний (IPC) в следующем смысле:

Теорема 245. *Для любой невыводимой в интуиционистском исчислении высказываний формулы $\Phi(t_1, \ldots, t_m)$ со связками $\wedge, \vee, \to$ существуют число k, формулы $T_1, \ldots, T_m$ с переменными $s_1, \ldots, s_k$ и некоторая критическая импликация $J(s_1, \ldots, s_k)$, для которых формула*

$$\Phi(T_1, \ldots, T_m) \to J$$

выводима в IPC. При этом формулы $T_1, \ldots, T_m$ содержат только связки $\wedge, \vee$.

Доказательство этой чисто логической теоремы Ю. Т. Медведева выходит за рамки нашей книги. Интересующегося читателя мы отсылаем к работе [176], в которой приведено полное доказательство. Используя это утверждение, мы можем наконец доказать теорему 238.

◀ Пусть формула $\Phi(t_1,\ldots,t_m)$ невыводима в IPC. По теореме 245 существуют число k, формулы $T_1,\ldots,T_m$ с переменными $s_1,\ldots,s_k$ и связками $\wedge,\vee$ и некоторая критическая импликация $J(s_1,\ldots,s_k)$ такие, что формула

$$\Phi(T_1,\ldots,T_m)\to J$$

выводима в IPC.

Возьмем случайные независимые слова $u_1,\ldots,u_k$ длины n/c (константу c подберём позже) и подставим одноэлементные множества, соответствующие этим словам, вместо переменных в формулы $T_1,\ldots,T_m$. Поскольку эти формулы не содержат импликации, мы получим конечные непустые множества слов $X_1,\ldots,X_m$. Формула $\Phi(T_1,\ldots,T_m)\to J$ выводима в IPC, поэтому сложность задачи

$$\Phi(X_1,\ldots,X_m)\to J(u_1,\ldots,u_k)$$

ограничена сверху некоторой константой. Теорема 244 гарантирует, что сложность задачи $J(u_1,\ldots,u_k)$ не меньше $n/c-O(1)$. Поэтому и сложность задачи $\Phi(X_1,\ldots,X_m)$ также не меньше $n/c-O(1)$, что больше $n/(2c)$ при всех достаточно больших n.

Осталось понять, почему длины всех слов в $X_1,\ldots,X_m$ не превосходят n (при подходящем выборе константы c и при всех достаточно больших n). Любое слово из множеств $X_1,\ldots,X_m$ может быть получено из слов $u_1,\ldots,u_k$ с помощью фиксированного числа применений операции спаривания $v,w\mapsto[v,w]$ и операции спаривания с нулем или единицей. Если длина слова $[v,w]$ ограничена сверху линейной функцией от длин v,w, то утверждение очевидно. Ясно, что спаривающие функции с таким свойством существуют. Осталось заметить, что утверждение теоремы инвариантно относительно замены спаривающей функции на другую, поэтом мы можем предполагать, что спаривающая функция обладает этим свойством. ▶

13.5. Доказательство варианта теоремы 238 с помощью моделей Крипке

Доказывая теорему 238, мы приняли на веру утверждение теоремы 245. Хотелось бы обойтись без этого; Ан. А. Мучник предложил доказательство близкого утверждения, которое использует лишь существование контрмодели Крипке у невыводимых в IPC формул. (Было бы интересно найти аналогичное прямое доказательство полного варианта теоремы 238; авторам такое доказательство не известно.)

Теорема 246. *Пусть формула $\Phi(t_1,\ldots,t_k)$ со связками $\wedge,\vee,\to$ невыводима в IPC. Тогда найдется последовательность задач $X_1^n,\ldots,X_k^n$ сложности не более $O(n)$, для которой сложность задачи $\Phi(X_1^n,\ldots,X_k^n)$ не меньше n при всех достаточно больших n.*

Разница с теоремой 238 в том, что мы теперь ограничиваем лишь сложность подставляемых множеств (а не длину слов в них). Кроме того, мы не требуем их конечности (напротив, все множества, строящиеся в доказательстве ниже, будут бесконечными).

◀ Пусть $\langle K, \leqslant\rangle$ — конечная модель Крипке, в корне которой ложна данная нам формула $\Phi(t_1, \ldots, t_k)$. Объясним, как по этой модели и произвольному числу n построить множества $X_1, \ldots, X_k$. (Для краткости мы опускаем верхний индекс n, так как в дальнейшем рассуждении n фиксировано.)

Отберём некоторое бесконечное множество длин, включающее нулевую длину, так чтобы любые две различные ненулевые длины сильно различались (скажем, не менее чем вдесятеро) и чтобы все ненулевые отобранные длины были не меньше $2n$. Будем называть такие длины *правильными*. Распределим правильные длины между мирами модели K так, чтобы каждому миру досталось бесконечное множество длин, причём нулевая длина досталась корню, и чтобы множества длин, сопоставленных каждому миру, были разрешимыми. Будем называть длины, отнесённые к миру u, *длинами из мира* u, а слова этих длин — *словами из мира* u.

Пусть всё это сделано. Теперь мы можем объяснить, как построить множество X_i, подставляемое вместо переменной t_i. Оно равно объединению следующих двух множеств слов. Первое состоит из множества всех случайных слов из миров, в которых истинна переменная t_i. Второе, обозначаемое C, не зависит от i и равно множеству всех пар вида $\langle x, y\rangle$, где x и y — случайные слова из несравнимых миров. В этом доказательстве слово называется случайным, если его сложность не меньше, чем некоторая линейная функция его длины, скажем, не меньше десятой части длины. Случайным словом нулевой длины будем считать пустое.

Если соблюдать формальности, то в определении C вместо пары $\langle x, y\rangle$ следует брать её код $[x, y]$, а при объединении множеств дописывать нуль или единицу в начало, чтобы элементы разного происхождения не смешивались:

$$
\begin{aligned}
C &= \{[x, y] \mid x, y \text{ случайные слова из несравнимых миров}\},\\
X_i &= \{0x \mid t_i \text{ истинна в мире } v,\ \text{а } x \text{ — случайное слово из мира } v\} \cup\\
&\qquad \cup \{1y \mid y \in C\}.
\end{aligned}
$$

Поскольку слова из разных миров имеют разные длины, по любому слову из X_i можно понять, из каких миров взялись составляющие его слова.

Мы докажем индукцией по построению формулы Ψ от переменных $t_1, \ldots, t_k$, что множество решений задачи $\Psi(X_1, \ldots, X_k)$ по существу совпадает с множеством X_Ψ, которое определяется по тому же правилу, что и значения переменных. По определению X_Ψ есть множество случайных слов из миров, в которых истинна формула Ψ, к которому ещё добавлено C (соблюдая те же формальности, что и при определении X_i). Точнее, мы докажем, что $\Psi(X_1, \ldots, X_k)$ отличается от X_Ψ алгоритмическим преобразованием: для каждой формулы Ψ существуют две вычислимых функции f, g: первая преобразует любое решение задачи $\Psi(X_1, \ldots, X_k)$ в некоторый элемент X_Ψ, а вторая преобразует любой элемент множества X_Ψ в некоторое решение задачи $\Psi(X_1, \ldots, X_k)$.

Если Ψ есть переменная, то множества $\Psi(X_1, \ldots, X_k)$ и X_Ψ совпадают и обе функции тождественны. Поэтому достаточно доказать, что для множеств вида X_Ψ операции над задачами коммутируют с одноименными операциями над формулами:

(А) множество $X_\Psi \vee X_\Theta$ алгоритмически эквивалентно множеству $X_{\Psi\vee\Theta}$;

(Б) множество $X_\Psi \wedge X_\Theta$ алгоритмически эквивалентно множеству $X_{\Psi\wedge\Theta}$;

(В) множество $X_\Psi \to X_\Theta$ алгоритмически эквивалентно множеству $X_{\Psi\to\Theta}$.
Кроме этого, нам нужна ещё устойчивость логических операций над задачами относительно алгоритмической эквивалентности, то есть, такое свойство: если множества U, V алгоритмически эквивалентны соответственно U', V', то множество $U \vee V$ алгоритмически эквивалентно множеству $U' \vee V'$ (и аналогичное свойство для других двух операций). Эти свойства очевидны, и мы переходим к доказательству (А), (Б) и (В).

(А) По определению $X_{\Psi\vee\Theta}$ есть объединение множеств X_Ψ и X_Θ, поэтому по любому элементу $X_\Psi \vee X_\Theta$ легко указать некоторый элемент из $X_{\Psi\vee\Theta}$. В обратную сторону: если нам дан элемент объединения X_Ψ и X_Θ, то можно разобраться, какому из двух множеств он принадлежит и преобразовать его в соответствующий элемент множества $X_\Psi \vee X_\Theta$.

(Б) По определению $X_{\Psi\wedge\Theta}$ есть пересечение множеств X_Ψ и X_Θ, а $X_\Psi \wedge X_\Theta$ есть их декартово произведение. По любому элементу x из пересечения легко найти некоторый элемент декартова произведения, а именно $[x, x]$. В обратную сторону: пусть нам дан элемент $[x, y]$ из декартова произведения X_Ψ и X_Θ. Если хотя бы одно из слов x, y принадлежит C, то оно принадлежит и пересечению этих множеств. Пусть ни одно из них не принадлежит C, то есть x, y — это случайные слова из миров u, v, в которых истинны соответственно Ψ и Θ. Тогда эти миры мы можем найти.

Далее рассмотрим два случая.

(1) Эти миры несравнимы. Тогда мы просто выдаем пару $[x, y]$, которая по определению принадлежит C, а значит, и пересечению X_Ψ и X_Θ.

(2) Миры u, v сравнимы, скажем, u предшествует v. Тогда в силу монотонности истинности, формула Ψ истинна также в мире v, а значит слово y принадлежит пересечению X_Ψ и X_Θ.

(В) Это наиболее интересный случай. Нам надо доказать, что по любому элементу из $X_{\Psi\to\Theta}$ можно найти некоторый элемент $X_\Psi \to X_\Theta$ и, обратно, по любому элементу из $X_\Psi \to X_\Theta$ можно найти некоторый элемент $X_{\Psi\to\Theta}$. Сначала докажем первое утверждение.

Пусть нам дано слово x из $X_{\Psi\to\Theta}$. Нам надо найти некоторое решение задачи $X_\Psi \to X_\Theta$. По-другому можно сказать так: нам надо, имея x и любое слово y из X_Ψ, найти некоторое слово из X_Θ. Если хотя бы одно из слов x, y принадлежит C, мы выдаём это слово. Пусть ни одно из них не принадлежит C. Тогда x — случайное слово из некоторого мира u, в котором истинна формула $\Psi \to \Theta$, а слово y — случайное слово из некоторого мира v, в котором истинна формула Ψ (и эти миры мы можем найти, изучив длины x, y). Рассмотрим три случая.

(1) Если мир v предшествует миру u, то формула Ψ истинна и в мире u, следовательно, и формула Θ истинна в мире u (напомним, что импликация $\Psi \to \Theta$ истинна в мире u). Значит, слово x принадлежит множеству X_Θ.

(2) Если мир u предшествует миру v, то формула Θ истинна в мире v, поскольку в нём истинна формула Ψ и импликация $\Psi \to \Theta$ истинна в мире u. Значит, слово y принадлежит множеству X_Θ.

(3) Если миры u и v несравнимы, то пара $[x, y]$ принадлежит C, а следовательно, и X_Θ.

Осталось доказать, что по любому элементу из $X_\Psi \to X_\Theta$ можно найти некоторый элемент $X_{\Psi\to\Theta}$. Пусть нам дана программа r, преобразующая любое слово из X_Ψ в некоторое слово из X_Θ.

Если формула $\Psi \to \Theta$ истинна в корне, то пустое слово принадлежит $X_{\Psi\to\Theta}$ и мы выдаем его.

Пусть формула $\Psi \to \Theta$ ложна в корне. Тогда существует мир u, в котором истинна формула Ψ, но ложна Θ. Выберем правильную длину l из мира u, значительно большую длины программы r. Мы знаем, что в применении ко всем случайным словам длины l программа r остановится и выдаст слово из X_Θ. К сожалению, мы не знаем, какие слова случайны, а какие нет. Поэтому будем применять параллельно программу r ко всем словам длины l. Если какое-то слово s получится более чем $2^{l/2}$ раз в качестве выхода r, то мы нашли нужное слово. Действительно, неслучайных слов длины l значительно меньше, чем $2^{l/2}$, а значит, хотя бы один раз слово s получено как результат работы на случайном слове. Следовательно, слово s принадлежит $X_\Theta \subset X_{\Psi\to\Theta}$.

Попробуем доказать, что найдётся слово, имеющее так много прообразов. Пусть x — случайное слово длины l, а s — результат работы программы r на входе x. Сложность s не превосходит суммы длин x и r (с логарифмической точностью), и поэтому значительно меньше $2l$. Кроме того, s принадлежит X_Θ, а, следовательно, имеет вид вид $0t$ или $1[y,z]$, где t,y,z — случайные слова разрешённых длин. Поэтому длины t,y,z не могут быть больше l. Более того, длина t должна быть строго меньше l, так как слово t принадлежит миру, в котором истинна Θ, и этот мир не равен u (напомним, что Θ ложна в u). Значит длина t не больше $l/10$. Кроме того, слова y и z имеют разные длины, поскольку находятся в несравнимых мирах, следовательно одна из них должна быть не больше $l/10$. Если бы и другая длина тоже была всегда не больше $l/10$, то множество результатов работы программы r на случайных входах имело бы мощность значительно меньше $2^{l/2}$. А следовательно, какое-то слово имело бы более $2^{l/2}$ прообразов.

Поскольку всё-таки длина одного из слов y,z может быть равна l, мы не можем гарантировать наличия слова с таким количеством прообразов. Тем не менее, наши рассуждения доказывают следующее. Либо некоторое слово s имеет более $2^{l/2}$ прообразов, либо верно следующее: найдётся такое слово y, что для более чем $2^{l/2}$ слов x длины l программа r на входе x выдаёт слово вида $1[y,z]$ (или $1[z,y]$), где длина z равна l. (Для разных входов x слово z может быть разным.) В этом случае y обязано быть случайным словом из некоторого мира, несравнимого с u. Действительно, хотя бы одно из слов x, на которых r выдаёт результат $1[y,z]$ (или $1[z,y]$) случайно, а значит, результат работы r должен принадлежать X_Θ. А это может быть только когда y и z случайны и принадлежат несравнимым мирам. Поскольку слово z принадлежит миру u (его длина равна l), слово y принадлежит миру, несравнимому с u.

Применяя параллельно программу r ко всем словам длины l, найдём такое слово s или такое y. Если реализовался первый случай и найдено слово s с большим числом прообразов, то оно принадлежит $X_\Theta \subset X_{\Psi\to\Theta}$.

Во втором случае поступим следующим образом. У нас есть случайное слово y из некоторого мира v, несравнимого с u. Если в мире v истинна формула $\Psi \to \Theta$,

то слово y и есть искомое. Иначе существует мир u_1 выше мира v, в котором истинна формула Ψ, но ложна Θ, и можно повторить все рассуждения для этого мира. В результате либо мы найдём некоторое слово из X_Θ, либо найдём случайное слово из некоторого мира v_1, несравнимого с миром u_1. При этом мир v_1 не может быть ниже v (в этом случае он был бы и ниже u_1). Значит, мир v_1 либо несравним с миром v (в этом случае мы нашли пару случайных слов в несравнимых мирах и можем остановиться), либо строго выше него. Во втором случае (мир v_1 строго выше мира v), если формула $\Psi \to \Theta$ истинна в мире v_1, то можно остановиться. А иначе можно повторить все рассуждения для мира v_1. Рано или поздно мы остановимся, поскольку модель конечна и, следовательно, не содержит бесконечных возрастающих последовательностей миров.

Итак, мы доказали, что множество решений задачи $\Phi(X_1, \ldots, X_k)$ алгоритмически эквивалентно множеству X_Φ. По условию формула Φ ложна в корне, поэтому длины всех слов в X_Φ не меньше $2n$, значит, сложность задачи $\Phi(X_1, \ldots, X_k)$ больше n. Осталось понять, почему сложности всех множеств X_i можно сделать равными $O(n)$. Можно считать без ограничения общности, что модель K содержит мир, в котором все формулы истинны (добавим новый мир, который больше всех остальных и в котором все переменные истинны; на истинность формул в корне этот мир не повлияет). Также можно считать, что длина $2n$ правильная и принадлежит этому миру. Тогда множество случайных слов длины $2n$ принадлежит всем X_i. ►

13.6. Пример задачи, сложность которой не выражается через условные сложности подставленных слов

Мы оценили сложности многих задач, полученных из одноэлементных множеств логическими операциями $\wedge, \vee, \to$. Могло создаться впечатление, что сложность любой задачи, составленной из синглетонов $\{x\}, \{y\}, \ldots$ с помощью этих операций, можно выразить через сложности слов $x, y, \ldots$, их пар, троек и т.д. Однако это не так. Задача $(x \to z) \wedge (y \to z)$, сложность которой, напомним, примерно равна наибольшей из условных сложностей $KS(z \mid x)$, $KS(z \mid y)$, уже находится на грани той области, где это возможно. А именно, для задачи $(a \to c) \wedge (b \to d)$ это уже невозможно, как мы увидим в теореме 247 (см. [173]). Но сначала несколько простых утверждений:

341 Докажите, что с логарифмической точностью для любых a, b, c, d выполнены неравенства

$$
\begin{aligned}
&KS((a \to c) \wedge (b \to d)) \leqslant KS(c \mid a) + KS(d \mid b),\\
&KS((a \to c) \wedge (b \to d)) \leqslant KS(d \mid b, c) + KS(c),\\
&KS((a \to c) \wedge (b \to d)) \leqslant KS(c \mid a, d) + KS(d),\\
&KS((a \to c) \wedge (b \to d)) \geqslant KS(b, c, d \mid a) - KS(b \mid a, c),\\
&KS((a \to c) \wedge (b \to d)) \geqslant KS(a, c, d \mid b) - KS(a \mid b, d).
\end{aligned}
$$

$\boxed{\mathbf{342}}$ Найдите последовательность четвёрок слов линейной от n длины, для которых наименьшая из трёх верхних оценок сложности задачи $(a \to c) \wedge (b \to d)$ из задачи 341 превышает наибольшую из двух нижних оценок более, чем на $o(n)$. [Указание. Можно положить $a = d$ и $b = c$, где a, b — независимые случайные слова длины n.]

Теорема 247. *Существует положительное δ и последовательности четвёрок слов $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ и $\bar{a}_n, \bar{b}_n, \bar{c}_n, \bar{d}_n$ линейной (от n) длины, для которых сложность задачи $(\tilde{a}_n \to \tilde{c}_n) \wedge (\tilde{b}_n \to \tilde{d}_n)$ более чем на δn превосходит сложность задачи $(\bar{a}_n \to \bar{c}_n) \wedge (\bar{b}_n \to \bar{d}_n)$. При этом сложности слов $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ отличаются от сложностей $\bar{a}_n, \bar{b}_n, \bar{c}_n, \bar{d}_n$ не более чем на $O(\log n)$, и то же самое верно для всех пар, троек и самих четвёрок.*

◄ **Геометрическое доказательство**. Начнём с простого замечания. Для четвёрки $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ из условия теоремы оба неравенства из задачи 341, дающие нижнюю оценку сложности задачи $(a \to c) \wedge (b \to d)$, должны быть строгими — разница между левой и правой частями должна быть не меньше εn. Поэтому начнем с поиска такой четвёрки. А затем мы подберем другую четвёрку с теми же сложностями, для которой эта разница есть $o(n)$.

Будем использовать геометрическую конструкцию, которая нам помогла найти примеры слов, у которых не выделяется общая информация. Возьмем поле из 2^n элементов и будем рассматривать точки, прямые и плоскости в трёхмерном аффинном пространстве над этим полем. Всего имеется 2^{3n} точек. Количество прямых примерно равно 2^{4n} (прямая задаётся парой различных точек — 2^{6n} вариантов, причем каждая прямая таким образом может быть задана примерно 2^{2n} способами, так как содержит 2^n точек). Количество плоскостей примерно равно 2^{3n} (плоскость задаётся тремя точками — 2^{9n} вариантов, а поскольку каждая плоскость содержит 2^{2n} точек, она содержит примерно 2^{6n} троек точек). В качестве $\langle \tilde{a}, \tilde{b} \rangle$ мы возьмём случайную пару различных пересекающихся прямых, $\tilde{c}$ будет их точкой пересечения, а $\tilde{d}$ — плоскостью, в которой обе они лежат. Нетрудно посчитать, что для этой четвёрки обе нижние оценки из задачи 341 равны n (с точностью до $O(\log n)$).

Докажем, что
$$KS((\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})) \geqslant 1{,}5n$$
(с точностью до $O(\log n)$).

Пусть γ есть решение задачи $(\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})$, то есть γ — это пара программ $\langle \alpha, \beta \rangle$, первая из которых преобразует $\tilde{a}$ в $\tilde{c}$, а вторая — $\tilde{b}$ в $\tilde{d}$. Рассмотрим множество S, состоящее из всех пар различных пересекающихся прямых a, b, для которых программа α преобразует a в общую точку a и b, а программа β преобразует b в плоскость, содержащую a и b. При известных α, β и n мы можем перечислять все элементы S. Так как пара $\langle \tilde{a}, \tilde{b} \rangle$ принадлежит S, мы можем заключить, что
$$7n \leqslant KS(\tilde{a}, \tilde{b}) \leqslant KS(\gamma) + \log |S|$$
(с точностью $O(\log n)$). Таким образом, нам достаточно доказать, что количество пар в S не превосходит $O(2^{5,5n})$. Это вытекает из следующей комбинаторной леммы.

Лемма. *Пусть f — некоторая функция, сопоставляющая каждой прямой некоторую точку на этой прямой, а g — некоторая функция, сопоставляющая каждой прямой некоторую плоскость, содержащую эту прямую. Пусть множество S состоит из всех пар прямых $\langle a, b\rangle$ таких, что точка $f(a)$ принадлежит b, а плоскость $g(b)$ содержит a. Тогда S содержит не более $O(2^{5,5n})$ пар.*

Доказательство. Сначала посмотрим, какая оценка получится, если не прибегать к хитростям. Для каждой прямой b существует примерно 2^{2n} прямых a в плоскости $g(b)$, поэтому мощность S не более чем в 2^{2n} раз превосходит количество прямых 2^{4n}, что даёт верхнюю оценку $|S| \leqslant 2^{6n}$. Такую же оценку мы получим, если для каждой прямой a подсчитаем количество прямых b, проходящих через $f(a)$. Заметим, что в первом подсчёте мы не использовали того, что прямая b должна содержать точку $f(a)$, а во втором — того, что прямая a должна лежать в плоскости $g(b)$. Наш план таков: мы модифицируем первое из рассуждений, показав, что в S в среднем на каждую прямую b приходится не более $2^{1,5n}$ прямых a. При этом мы уже будем учитывать условие $f(a) \in b$. (Можно действовать и симметричным образом — показать, что в S в среднем на каждую прямую a приходится не более $2^{1,5n}$ прямых b.)

Разобьём S на слои, поместив в один слой пары $\langle a, b\rangle$ с одинаковым значением $g(b)$. Все пары данного слоя лежат в одной плоскости, а сам слой однозначно задаётся этой плоскостью. Мы ограничим сверху количество пар в каждом слое, а затем просуммируем полученные оценки. Итак, фиксируем плоскость d и оценим сверху количество пар в S, лежащих в слое, соответствующем d.

Для этого рассмотрим произвольную точку c на плоскости d и обозначим через A_c множество прямых a в плоскости d с $f(a) = c$, а через B_c — множество прямых b, содержащих точку c, для которых $g(b) = d$ (из условия следует, что прямая a содержит точку c, а прямая b содержится в плоскости d). Ясно, что размер слоя не превосходит

$$\sum_c |A_c||B_c| \leqslant \sqrt{\sum_c |A_c|^2 \sum_c |B_c|^2}$$

(мы применили неравенство Коши). Обе суммы под корнем легко ограничить сверху, поскольку они имеют ясный смысл. А именно, $\sum_c |A_c|^2$ пропорционально вероятности того, что для двух равномерно и независимо выбранных прямых a', a'' в плоскости d будет выполнено $f(a') = f(a'')$. Действительно, вероятность этого события есть сумма по всем c вероятности пересечения независимых событий $f(a') = c$ и $f(a'') = c$. Вероятности этих событий равны между собой и равны отношению мощности A_c к общему количеству прямых в плоскости d, примерно равному 2^{2n}. С другой стороны, вероятность события $f(a') = f(a'')$ не превосходит примерно 2^{-n} (при любом фиксированном a' вероятность события $f(a') = f(a'')$ не превосходит вероятности того, что прямая a'' проходит через точку $f(a')$, что примерно равно 2^{-n}). Поэтому $\sum_c |A_c|^2$ не превосходит примерно

$$(2^{2n})^2 \cdot 2^{-n} = 2^{3n}.$$

Все эти подсчёты верны с точностью до мультипликативной константы.

Вторая сумма $\sum_c |B_c|^2$ связана со средним количеством общих точек у двух равномерно и независимо выбранных прямых b', b'' в множестве M_d, состоящем из всех прямых b, для которых $g(b) = d$ (все они лежат в плоскости d). Действительно, среднее количество общих точек b' и b'' равно сумме по всем c вероятности того, что точка c принадлежит одновременно b' и b''. События $c \in b'$ и $c \in b''$ независимы и вероятности их равны между собой и равны отношению мощности B_c к общему количеству прямых в M_d. Поэтому среднее количество общих точек у b' и b'' равно отношению суммы $\sum_c |B_c|^2$ к квадрату мощности M_d. С другой стороны, поскольку любые две прямые имеют не более одной общей точки или совпадают (что бывает с вероятностью $1/|M_d|$), среднее количество общих точек не превосходит $1 + 2^n/|M_d|$. Поэтому

$$\sum_c |B_c|^2 \leqslant |M_d|^2(1 + 2^n/|M_d|) = |M_d|^2 + |M_d|2^n \leqslant (|M_d| + 2^n)^2.$$

Напомним, что количество пар в слое, задаваемом плоскостью d, не превосходит $\sqrt{\sum_c |A_c|^2 \sum_c |B_c|^2}$, поэтому оно не больше

$$\sqrt{2^{3n}(|M_d| + 2^n)^2} = 2^{1.5n}(|M_d| + 2^n)$$

(с точностью до умножения на постоянный множитель).

Осталось просуммировать полученные оценки по всем d:

$$|S| \leqslant 2^{1,5n} \sum_d (|M_d| + 2^n) = 2^{1,5n} \left(\sum_d |M_d| + \sum_d 2^n \right).$$

Семейства прямых M_d образуют разбиение множества всех прямых, следовательно, сумма мощностей M_d равна общему количеству прямых 2^{4n}. Кроме того, количество различных плоскостей d примерно равно 2^{3n}, поэтому вторая сумма также примерно равна 2^{4n}. Лемма, а вместе с ней и нижняя оценка сложности задачи $(\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})$, доказана.

Для завершения доказательства теоремы нам достаточно построить другую четвёрку $\langle \bar{a}, \bar{b}, \bar{c}, \bar{d} \rangle$ с теми же сложностями, что и у $\langle \tilde{a}, \tilde{b}, \tilde{c}, \tilde{d} \rangle$, для которой сложность задачи $(\bar{a} \to \bar{c}) \wedge (\bar{b} \to \bar{d})$ примерно равна n.

Для этого возьмём случайное слово длины $7n$ и разрежем его на 7 блоков u, v, w, p, q, r, s длины n. Положим $\bar{a} = uvws$, $\bar{b} = pqrs$, $\bar{c} = ups$, $\bar{d} = vqs$. Нетрудно проверить, что сложности слов обеих четвёрок, их пар и т. д. одинаковы и равны:

$$\begin{gathered}
KS(a) = KS(b) = 4n, \quad KS(c) = KS(d) = 3n,\\
KS(a, b) = 7n,\\
KS(a, c) = KS(a, d) = KS(b, c) = KS(b, d) = KS(c, d) = 5n,\\
KS(a, c, d) = KS(b, c, d) = 6n,\\
KS(a, b, c) = KS(a, b, d) = KS(a, b, c, d) = 7n.
\end{gathered}$$

При этом сложность задачи $(\bar{a} \to \bar{c}) \wedge (\bar{b} \to \bar{d})$ примерно равна n, поскольку, зная побитовую сумму слов p и v (по модулю 2), можно $\bar{a}$ преобразовать в $\bar{c}$, а $\bar{b}$ в $\bar{d}$.

Вероятностное доказательство. Мы опять начнём с предъявления четверки $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ для которой разница между сложностью задачи $(\tilde{a}_n \to \tilde{c}_n) \wedge (\tilde{b}_n \to \tilde{d}_n)$ и нижними оценками из задачи 341 линейно растет как функция от n.

Фиксируем n. Мы определим четвёрку слов $\langle \tilde{a}, \tilde{b}, \tilde{c}, \tilde{d} \rangle$ длины n, сложности которых примерно равны n, сложности всех пар, составленных из них, — примерно $2n$, сложности всех троек, а также сложность самой четвёрки — примерно $3n$. Из этого следует, что обе нижнии оценки задачи 341 близки к n. Но для определенной нами четвёрки сложность задачи $(\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})$ будет близка к $2n$.

Для построения такой четвёрки рассмотрим всевозможные функции Q, отображающие каждую тройку $\langle a, b, c \rangle$ слов длины n в некоторое слово d той же длины.

Лемма. *Для всех достаточно больших n найдется такая функция Q сложности $\log n + O(1)$, что для не менее чем половины троек $\langle a, b, c \rangle$ слов длины n сложность задачи $(a \to c) \wedge (b \to Q(a, b, c))$ не меньше $2n - O(\log n)$.*

Прежде чем доказывать эту лемму, объясним, как из неё будет следовать утверждение теоремы. Выберем функцию Q, удовлетворяющую утверждению леммы. Тогда для не менее чем половины троек $\langle a, b, c \rangle$ выполнено неравенство

$$KS((a \to c) \wedge (b \to Q(a, b, c))) \geqslant 2n - O(\log n).$$

Поэтому найдётся такая тройка сложности не меньше $3n - 1$. Возьмем в качестве $\langle \tilde{a}, \tilde{b}, \tilde{c} \rangle$ любую такую тройку и положим $\tilde{d} = Q(\tilde{a}, \tilde{b}, \tilde{c})$. Сложность четвёрки и сложность тройки $\langle \tilde{a}, \tilde{b}, \tilde{c} \rangle$ будут равны примерно $3n$, как мы и хотели. Поэтому случайны и все пары, составленные из $\tilde{a}, \tilde{b}, \tilde{c}$, и сами $\tilde{a}, \tilde{b}, \tilde{c}$.

Кроме того, из сказанного следует, что и тройка $\langle \tilde{b}, \tilde{c}, \tilde{d} \rangle$ случайна. Действительно, сложность задачи $(\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})$ не меньше $2n$ и ограничена сверху суммой $KS(\tilde{c}) + KS(\tilde{d} \,|\, \tilde{b}, \tilde{c}) \leqslant 2n$. Следовательно, оба слагаемых в этой сумме примерно равны n, то есть слово $\tilde{d}$ независимо от пары $\langle \tilde{b}, \tilde{c} \rangle$. А поскольку сама пара $\langle \tilde{b}, \tilde{c} \rangle$ случайна, то случайна и тройка $\langle \tilde{b}, \tilde{c}, \tilde{d} \rangle$. Единственное, что нам нужно, но не следует из сказанного, это случайность двух оставшихся троек $\langle \tilde{a}, \tilde{b}, \tilde{d} \rangle$ и $\langle \tilde{a}, \tilde{c}, \tilde{d} \rangle$. Как добиться этого, мы обсудим позднее.

Доказательство. Переходя к доказательству леммы, определим некоторое алгоритмически проверяемое свойство функции Q, которое гарантирует, что для большинства троек $\langle a, b, c \rangle$ сложность задачи $(a \to c) \wedge (b \to Q(a, b, c))$ не меньше $2n - O(\log n)$. Затем мы докажем, что вероятность того, что случайная функция Q обладает этим свойством, положительна. Зная n, перебором можно найти функцию Q с таким свойством, поэтому первая найденная функция будет иметь сложность порядка $\log n + O(1)$.

Пусть S обозначает множество всех слов длины n, и пусть M — некоторое множество пар всюду определённых функций из S в S. Скажем, что множество M *обслуживает* четвёрку $\langle a, b, c, d \rangle \in S^4$, если $f(a) = c$ и $g(b) = d$ для некоторой пары $\langle f, g \rangle \in M$. Упомянутое свойство функции Q состоит в следующем:

> *любое множество M пар функций из S в S мощности менее 2^k обслуживает менее $1/8$ четвёрок вида $\langle a, b, c, Q(a, b, c) \rangle$.*

Число k мы определим позднее; оно будет чуть меньше $2n$. Это свойство гарантирует, что для $7/8$ троек $\langle a, b, c \rangle$ сложность задачи $(a \to c) \wedge (b \to Q(a, b, c))$

больше k. Действительно, решениями задачи $(a \to c) \wedge (b \to d)$ являются пары программ $\langle p, q\rangle$, для которых $[p](a) = c$, $[q](b) = d$. Для каждой пары программ $\langle p, q\rangle$ сложности меньше k доопределим любым способом функции $a \mapsto [p](a)$ и $b \mapsto [q](b)$ на всё множество S. Рассмотрим получившееся множество M пар функций. Его мощность меньше 2^k. Следовательно, M обслуживает не более одной восьмой части троек $\langle a, b, c, Q(a, b, c)\rangle$. А по построению M обслуживает все четвёрки $\langle a, b, c, d\rangle$, для которых сложность задачи $(a \to c) \wedge (b \to d)$ меньше k. (Как видно, в формулировке свойства мы могли бы взять $1/2$ вместо $1/8$. Зачем нам нужен этот запас, станет ясно позднее.)

Рассмотрим равномерное распределение на всех функциях Q рассматриваемого вида. Другими словами, значения $Q(a, b, c)$ независимы друг от друга и равномерно распределены в множестве S. Оценим сверху вероятность того, что случайная функция Q не обладает описанным свойством (нам нужно так подобрать $k = n - O(\log n)$, чтобы эта вероятность была меньше 1).

Сначала зафиксируем множество M из 2^k функций и оценим сверху вероятность того, что оно обслуживает более чем $1/8$ четвёрок. При подсчёте этой вероятности удобно будет разделить тройки $\langle a, b, c\rangle$ на «плохие» и «хорошие». Плохих троек будет меньше 2^{3n-4}. А для хороших троек $\langle a, b, c\rangle$ мала будет доля тех d, для которых четверка $\langle a, b, c, d\rangle$ обслуживается M. По закону больших чисел отсюда будет следовать, что доля обслуживаемых четверок $\langle a, b, c, Q(a, b, c)\rangle$, у которых проекция на первые три координаты хорошая, также меньше $1/16$.

Точнее, назовём тройку $\langle a, b, c\rangle$ (а также четвёрку $\langle a, b, c, d\rangle$) *плохой*, если количество пар $\langle f, g\rangle$ из M, для которых $f(a) = c$, больше $|M|2^{-n+4}$ (это свойство не зависит от b и d). Поскольку для каждой пары $\langle f, g\rangle$ из M существует лишь 2^{2n} троек вида $\langle a, b, f(a)\rangle$, количество плохих троек меньше $|M| \cdot 2^{2n}/(|M|2^{-n+4}) = 2^{3n-4}$; остальные будут хорошими.

Для любой хорошей тройки $\langle a, b, c\rangle$ вероятность того, что множество M обслуживает четвёрку $\langle a, b, c, Q(a, b, c)\rangle$, не превосходит $1/32$. Действительно, если M обслуживает четвёрку $\langle a, b, c, d\rangle$, то d принадлежит множеству, состоящему из всех таких $g(b)$, что для некоторой функции f пара $\langle f, g\rangle$ принадлежит M и $f(a) = c$. Это множество содержит не более $2^{-n+4} \cdot |M| = 2^{k-n+4}$ элементов. Следовательно, если выбрать k меньшим $2n - 9$, это множество будет содержать менее чем $(1/32)$-ю часть всех слов длины n.

Теперь воспользуемся неравенством Чернова для больших уклонений: если проводить N независимых испытаний, вероятность успеха в каждом из которых равна p, то для любого $\varepsilon > 0$ доля успешных испытаний будет меньше $p + \varepsilon$ с вероятностью не меньше $1 - e^{-2\varepsilon^2 N/2}$. Мы будем применять следующее его следствие, получаемое при $\varepsilon = p$: если вероятность успеха в отдельном испытании не превосходит p, то вероятность того, что доля успешных испытаний больше $2p$, меньше $e^{-2p^2 N}$. В нашем случае испытания нумеруются хорошими тройками $\langle a, b, c\rangle$, так что $(15/16)2^{3n} \leqslant N \leqslant 2^{3n}$. Испытание с номером $\langle a, b, c\rangle$ заключается в случайном выборе d, а успехом является то, что четвёрка $\langle a, b, c, d\rangle$ обслуживается множеством M. Вероятность успеха не больше $p = 1/32$. Поэтому при случайном выборе функции Q с вероятностью не менее $1 - e^{-\Omega(2^{3n})}$ для не более чем $1/16$-й части хороших троек $\langle a, b, c\rangle$ соответствующая четвёрка $\langle a, b, c, Q(a, b, c)\rangle$ обслужена M.

Если доля обслуженных среди хороших четвёрок не больше $1/16$, то даже если все плохие четвёрки обслужены, общая доля обслуженных четвёрок не больше $1/16 + 1/16 = 1/8$. Поэтому вероятность того, что количество обслуженных четвёрок $\langle a, b, c, Q(a, b, c)\rangle$ больше 2^{3n-3}, меньше $2^{-\Omega(2^{3n})}$.

Осталось убедиться, что эта вероятность даже при умножении на количество различных множеств M мощности 2^k остаётся малой. Количество множеств M не превосходит квадрата количества функций, отображающих слова длины n в слова длины n (оно равно 2^{n2^n}), возведённого в степень 2^k. Получается формула

$$2^{2n2^n2^k} = 2^{2^{n+k+\log n+1}}.$$

Сравним эту формулу с формулой для вероятности того, что доля обслуженных четвёрок больше $1/8$. При перемножении этих формул показатели экспонент первого этажа складываются и нам нужно чтобы отрицательное слагаемое $-\Omega(2^{3n})$ было больше положительного $2^{n+k+\log n+1}$ по абсолютной величине. Для этого достаточно положить $k = 2n - \log n - O(1)$. (Напомним, что для того, чтобы было мало плохих троек, нам достаточно, чтобы k было меньше $2n - 9$, так что наши подсчёты остаются верными.) Лемма доказана.

Теперь надо объяснить, какими свойствами функции Q будет обеспечена случайность двух оставшихся троек $\langle\tilde{a}, \tilde{b}, \tilde{d}\rangle$, $\langle\tilde{a}, \tilde{c}, \tilde{d}\rangle$. Самое простое — это действовать так же, как в доказательстве случайности тройки $\langle\tilde{b}, \tilde{c}, \tilde{d}\rangle$. Для этого нам нужно, чтобы для половины троек $\langle a, b, c\rangle$ сложной была не только задача $(a \to c) \wedge (b \to d)$, но также и симметричные ей задачи $(c \to b) \wedge (a \to d)$ и $(b \to a) \wedge (c \to d)$. А для этого надо ужесточить требования к функции Q, добавив два новых условия, симметричных предыдущим:

Лемма. *Для всех достаточно больших n найдётся такая функция Q сложности $\log n + O(1)$, что для половины троек $\langle a, b, c\rangle$ слов длины n сложность каждой из трёх задач $(a \to c) \wedge (b \to Q(a, b, c))$, $(c \to b) \wedge (a \to Q(a, b, c))$, $(b \to a) \wedge (c \to Q(a, b, c))$ больше $2n - O(\log n)$.*

Доказательство. Предыдущая лемма доказывалась вероятностным методом: мы рассмотрели некоторое свойство функции Q и доказали, что для случайной функции оно не выполнено с экспоненциально малой вероятностью. Теперь вместо одного свойства надо рассмотреть три симметричных свойства. Каждое из них также не выполнено с экспоненциально малой вероятностью. Поэтому для всех достаточно больших n найдётся функция, обладающая всеми тремя свойствами. Для такой функции количество троек, не обслуживаемых каким-то из трех симметричных способов, меньше $1/8 + 1/8 + 1/8 < 1/2$. Лемма доказана.

Для завершения второго доказательства теоремы осталось предъявить другую четвёрку с теми же сложностями, что и у построенной, но для которой сложность задачи $KS((a \to c) \wedge (b \to d))$ существенно меньше, чем $2n$. Для этого возьмём случайное слово длины $3n$, разрежем его на три части длины n. Это будут a, b и c; затем положим $d = a \oplus b \oplus c$. Зная $a \oplus c$, мы можем по a найти c, а по b найти d, следовательно $KS((a \to c) \wedge (b \to d)) = n$ (с точностью до константы). ▶

$\boxed{\textbf{343}}$ Докажите, что сложность задачи $(p \vee q) \to (r \vee s)$ также не выражается через сложности p, q, r, s, их пар, троек, и четвёрки. [Можно положить $p = a$,

$q = b$, $r = ac$, $s = bd$, где $\langle a, b, c, d\rangle$ — одна из двух четвёрок, использованных в доказательстве теоремы (скажем, в первом). Сложность полученной задачи зависит от того, какую из двух четвёрок взять, а сложности p, q, r, s, их пар, троек, и четвёрки не зависят от этого.]

Полезно сравнить вероятностное доказательство с геометрическим. Геометрическое доказательство более конструктивно — в нем первая четвёрка задана более явно, чем в вероятностном доказательстве. Зато в вероятностном доказательстве для первой четвёрки сложность задачи $(a \to c) \wedge (b \to d)$ равна верхним оценкам из задачи 341, которые все равны $2n$. Для четвёрок из геометрического доказательства эти верхние оценки также все равны $2n$, однако для первой четвёрки мы получили лишь нижнюю оценку $1{,}5n$ для сложности задачи $(a \to c) \wedge (b \to d)$. Можно ли получить лучшую оценку, неизвестно.

14. Алгоритмическая статистика

14.1. Постановка задачи. Дефект случайности.

Общими словами задачу математической статистики можно описать так: имеются результаты наблюдений, нужно предложить правдоподобную вероятностную гипотезу, их объясняющую (согласованную с ними).

Пусть, скажем, имеется некоторое устройство («чёрный ящик»). Мы включили его, и оно выдало последовательность из миллиона битов (= число от 0 до $2^{1\,000\,000}-1$). Что можно сказать о внутреннем строении «чёрного ящика», зная эту последовательность?

Классическая математическая статистика, увы, про это ничего не говорит. Вот если бы мы имели данные от большого числа одинаковых независимых устройств, или могли включать наше устройство много раз подряд (и предположить, что результаты последовательных включений независимы), тогда другое дело. (Заметим в скобках, что ситуации однократного и невоспроизводимого эксперимента не так уж и редки на практике.) Или, скажем, если бы мы заранее имели некоторое параметрическое семейство гипотез, — тогда можно было бы выбирать максимально правдоподобное значение параметра. А так ничего не скажешь — ведь все $2^{1\,000\,000}$ возможных исходов с точки зрения математической статистики совершенно равноправны (на множестве возможных исходов нет никакой структуры).

Тем не менее здравый смысл иногда позволяет выдвинуть кажущиеся разумными предположения. Например, если наше устройство выдало миллион нулей, скорее всего оно только и умеет что выдавать нули. А если оно выдало последовательность нулей и единиц безо всяких закономерностей, то похоже, что это результат миллиона независимых бросаний монеты. И хорошо бы как-нибудь уточнить механизм таких догадок.

В первом примере (слово из одних нулей) мы предлагали в качестве объяснения гипотезу, согласно которой только это слово и могло появиться. Формально говоря, наша гипотеза указывала в качестве множества возможных результатов эксперимента одноэлементное множество, состоящее из слова $00\ldots0$. Примерно то же естественно сказать и для любого двоичного слова x малой сложности: можно предположить, что чёрный ящик как раз и предназначен для порождения такого слова (имеет множество возможностей $\{x\}$).

Второй пример (слово без закономерностей, у которого сложность примерно равна длине) указывает другой крайний случай, когда мы говорим, что чёрный ящик мог выдать любое слово, то есть указываем в качестве множества возможностей множество всех слов длины 1 000 000.

Бывают и промежуточные примеры. Пусть оказалось, что первые 500 000 битов нашего слова равны нулю, а дальше идут 500 000 «случайных битов» (то есть некоторое слово длины и сложности 500 000). Видимо, разумная гипотеза в этом случае такова: чёрный ящик устроен так, что он сначала выдаёт 500 000 нулей, а потом уже 500 000 случайно выбранных символов. Соответствующее множество возможностей состоит из всех двоичных слов длины 1 000 000, у которых первая половина состоит только из нулей (таких слов, очевидно, $2^{500\,000}$).

Общая схема всех наших примеров такова. Нам дают некоторое слово x. Мы предлагаем в качестве «объяснения» этого слова некоторое множество A, содержащее x. При этом мы хотим, чтобы:

- множество A было простым (колмогоровская сложность A была бы мала);
- слово x было бы «типичным представителем» множества A.

Как уточнить эти пожелания? Чтобы говорить о колмогоровской сложности конечного множества, запишем элементы этого множества в список (заранее договорившись о порядке — скажем, в лексикографическом порядке), этот список закодируем двоичным словом и возьмём сложность этого слова. (Разные естественные способы кодирования дадут варианты сложности, отличающиеся не более чем на константу, поскольку от одного из них можно алгоритмически переходить к другому.)

Понятие «типичного представителя» также можно уточнить в терминах колмогоровской сложности. Вспомним, что если множество A содержит N элементов, то условная сложность $KS(x \mid A)$ любого его элемента при известном A не превосходит $\log_2 N + O(1)$, поскольку каждый элемент может быть задан своим порядковым номером в множестве A. Большинство элементов множества A имеют условную сложность, близкую к $\log_2 N$ (поскольку элементов меньшей сложности ма́ло по сравнению с N), и типичными представителями множества A можно назвать как раз входящих в это большинство.

Другими словами это можно объяснить так. Назовём (неотрицательную с точностью до константы) величину

$$d(x \mid A) = \log_2 |A| - KS(x \mid A)$$

дефектом случайности слова x в множестве A (мы будем использовать это равенство лишь в случае, когда $x \in A$, хотя правая часть имеет смысл всегда; при $x \notin A$ естественно считать, что $d(x \mid A) = +\infty$, поскольку в этом случае гипотеза A «максимально непригодна» для объяснения слова x). «Типичными представителями» множества A мы считаем те $x \in A$, для которых величина $d(x \mid A)$ мала.

344 Докажите, что вероятность того, что случайно взятый элемент данного множества A имеет дефект больше k, не превосходит 2^{-k}.

(Вероятность понимается здесь просто как доля элементов с большим дефектом в множестве A. Чтобы это утверждение было формально верным, надо брать целую часть снизу от $\log_2 |A|$ в определении дефекта, но поскольку сложность по существу определена с точностью до константы, этим можно пренебречь.)

Заметим ещё, что функция d (как функция двух аргументов) перечислима снизу (мы можем постепенно получать всё более точные нижние оценки для неё, хотя и

не можем сказать, когда мы дошли до предела). Это непосредственно следует из того, что функция KS перечислима сверху.

345 Пусть $\delta(x \,|\, A)$ определено для любого элемента x любого конечного множества A, причём: (а) функция δ перечислима снизу; (б) для любого множества A и любого числа k доля элементов A, для которых $\delta(x \,|\, A) > k$, не превосходит 2^{-k}. Тогда $\delta(x \,|\, A) \leqslant d(x \,|\, A) + O(1)$.

Смысл этой задачи (которая является простым следствием аналогичного утверждения для условной сложности, см. теорему 19 на с. 46) можно объяснить так. Могут быть разные мнения о том, какие элементы в каких множествах «нетипичны» (разные способы измерения дефекта). Если мы все их отнормируем по доле нетипичных элементов, потребовав, чтобы доля k-нетипичных элементов не превосходила 2^{-k}, а также будем требовать, чтобы нетипичность можно было рано или поздно алгоритмически обнаруживать, то среди всех способов измерения дефекта существует наилучший (с точностью до константы он обнаруживает не меньше дефектов, чем любой другой).

Другими словами, дефект случайности в конечном множестве можно рассматривать как конечный аналог дефекта случайности для бесконечных последовательностей (раздел 3.5). Точнее, приведённое определение соответствует ограниченному по вероятности дефекту; можно рассмотреть и аналог ограниченного в среднем дефекта, как показывает следующая задача.

346 Покажите, что *префиксный дефект случайности* элемента x относительно конечного множества A, определяемый как $d_P(x \,|\, A) = \log_2 |A| - KP(x \,|\, A)$, является максимальной перечислимой снизу функцией δ двух аргументов (x и A), для которой среднее значение $(1/|A|)\sum_{x \in A} 2^{\delta(x \,|\, A)}$ не превосходит единицы для любого конечного множества A. [Указание: воспользуйтесь описанием условной префиксной сложности как логарифма априорной вероятности.]

Таким образом, задачу поиска хорошего объяснения для данного слова x можно сформулировать так: надо найти *простое множество A, для которого дефект $d(x \,|\, A)$ мал.* Всегда ли (для любого ли x) это возможно? Этот вопрос мы рассмотрим в следующем разделе.

Отметим ещё, что знатоки теории вероятностей могут законно упрекнуть нас в том, что мы рассматриваем лишь равномерные распределения в качестве вероятностных гипотез. Вместо этого можно было бы говорить о произвольных распределениях вероятностей (для простоты — с конечным носителем и рациональными значениями вероятностей) и определять дефект слова x относительно распределения P как $-\log_2 P(x) - KS(x \,|\, P)$ (считая его бесконечным для тех x, для которых $P(x) = 0$: для таких x объяснение P совсем непригодно).

Для равномерных распределений (все элементы конечного множества имеют вероятность $1/|A|$) это определение совпадает с прежним. Можно заметить, что и в общем случае отличие не слишком велико, как показывает следующая задача:

347 Пусть для слова x сложности n найдено вероятностное распределение сложности не больше k, при котором дефект x не больше l. Тогда существует множество A сложности не больше $k + O(\log(l + n))$, содержащее x, относитель-

но которого дефект x не превосходит $l + O(\log(l + n))$. [Указание. Пусть p — вероятность слова x относительно распределения P, округлённая вниз до ближайшей степени двойки. Положим $A = \{y \mid P(y) \geqslant p\}$.]

Поэтому в дальнейшем мы в основном ограничиваемся конечными множествами (и равномерными распределениями вероятностей на них) в качестве описаний. При этом важно подчеркнуть, что говоря о сложности конечного множества, мы имеем в виду его сложность как конструктивного объекта (списка элементов), а не, скажем, сложность его перечисления (сложность программы, порождающей все его элементы). Если допустить последнее, то определение стохастичности станет неинтересным: любой объект x сложности n содержится в множестве S_n всех объектов сложности не больше n, которое имеет $O(2^n)$ элементов и которое можно порождать программой сложности $O(\log n)$, но вряд ли стоит считать S_n хорошим «объяснением» для x.

Аналогичным образом (в определении с мерами) рассматривать сложность меры как конечного объекта (если её значения рациональны) или даже сложность мер с вычислимыми значениями, имея в виду сложность соответствующей программы, но нельзя рассматривать перечислимые снизу меры и сложность программ, приближающих их снизу — ибо в этом случае всякое слово имеет простое объяснение — максимальную перечислимую снизу полумеру.

14.2. Стохастические объекты

Будем говорить, что слово x является (α, β)-стохастическим, если для него существует «хорошее объяснение»: существует конечное множество A, содержащее слово x, для которого $KS(A) \leqslant \alpha$ и $d(x \mid A) \leqslant \beta$.

Возникает такой естественный вопрос: будем рассматривать все слова длины n и положим α и β равными $O(\log n)$ или $o(n)$ (тем самым сложность гипотезы, как и полагается, мала по сравнению с длиной объясняемых слов). Будут ли при таких α и β существовать нестохастические («необъяснимые») объекты? Да, как показывает следующая теорема.

Теорема 248. *При $2\alpha + \beta < n - O(\log n)$ существуют слова длины n, не являющиеся (α, β)-стохастическими.*

(Точнее: существует такая константа c, что для всех достаточно бóльших n и для всех α и β с $2\alpha + \beta < n - c \log n$ существуют слова длины n, не являющиеся (α, β)-стохастическими.)

◄ Рассмотрим список всех конечных множеств сложности не больше α. Сложность такого списка не больше $\alpha + O(\log \alpha) = \alpha + O(\log n)$ (см. с. 36). Для краткости мы будем игнорировать члены порядка $O(\log n)$, надеясь, что читатель к этому уже привык и легко исправит рассуждение; в частности, мы будем считать, что список имеет сложность не больше α.

Отберём в этом списке множества, имеющие не более $2^{\alpha+\beta}$ элементов. Результирующий список также имеет сложность α и содержит не более 2^{α} множеств размера $2^{\alpha+\beta}$. Всего в них $2^{2\alpha+\beta} < 2^n$ элементов, и поэтому есть слова длины n,

не попадающие в эти множества. Рассмотрим первое такое слово (скажем, в алфавитном порядке). Оно имеет сложность не более α, поскольку для его задания (помимо числа n) достаточно знать этот список.

Покажем, что это слово (обозначим его t) не может быть (α, β)-стохастическим. Если t содержится в некотором множестве A сложности не больше α, то размер этого множества больше $2^{\alpha+\beta}$, поскольку все меньшие множества мы исключили по построению. Тогда $d(t\,|\,A) = \log|A| - KS(t\,|\,A) \geqslant (\alpha+\beta) - KS(t) \geqslant$ $\geqslant (\alpha+\beta) - \alpha \geqslant \beta$.

(Во всех этих построениях нужно, конечно, оставить запас размера $c\log n$ при достаточном c, чтобы компенсировать добавки порядка $O(\log n)$.) ►

С другой стороны имеется следующая очевидная оценка:

Теорема 249. *Если $\alpha + \beta > n + O(\log n)$, то все слова длины n являются (α, β)-стохастическими.*

◄ Достаточно разбить все слова длины n на 2^α множеств размера 2^β. ►

Как мы увидим, реальная ситуация (с какого момента все слова становятся стохастическими) ближе к этой нижней оценке, чем к предыдущей верхней (см. задачу 362 на с. 496).

Естественно поинтересоваться, насколько часто встречаются нестохастические объекты. Например, можно спросить, какую долю они составляют среди слов длины n. Заранее ясно, что эта доля не превосходит $2^{-\beta}$, поскольку в множестве A, состоящем из всех слов длины n, слова с дефектом β составляют долю не более $2^{-\beta}$.

С другой стороны, если $2\alpha + \beta$ много меньше n, можно продолжить доказательство теоремы 248, используя имеющийся в нём запас. А именно, для некоторого h можно взять все множества сложности не больше α с числом элементов не больше $2^{\alpha+\beta+h}$ и затем взять первые 2^h элементов, не покрытых этими множествами; такие элементы найдутся, если $2\alpha + \beta + h < n$. Сложность этих элементов будет не больше $\alpha + h$ и потому их дефект в любом множестве из более чем $2^{\alpha+\beta+h}$ элементов будет больше β. Проведя эти рассуждения аккуратно (не забывая о членах порядка $O(\log n)$), получаем такой результат:

Теорема 250. *Среди слов длины n доля не-(α, β)-стохастических слов не меньше $2^{-2\alpha-\beta-O(\log n)}$.*

Можно также интересоваться не просто долей нестохастических слов (другими словами, вероятностью получения такого слова при бросании монеты), а их суммарной априорной вероятностью (вероятностью их получения на выходе оптимальной вероятностной машины). Формально говоря, пусть $m(x)$ — априорная вероятность слова x (в смысле главы 4, равная $2^{-KP(x)+O(1)}$). Рассмотрим сумму $m(x)$ по всем словам длины n, не являющимся (α, β)-стохастическими.

Теорема 251. *Если $2\alpha + \beta < n - O(\log n)$ и $\alpha < \beta - O(\log n)$, то указанная сумма есть $2^{-\alpha+O(\log n)}$.*

◄ Нам нужно указать верхнюю и нижнюю оценку для этой суммы. Нижняя оценка получается из доказательства теоремы 248; в самом деле, построенное там

нестохастическое слово имеет сложность α и потому уже для него одного априорная вероятность равна $2^{-\alpha}$ (как всегда, мы опускаем слагаемые $O(\log n)$).

Для получения верхней оценки рассмотрим сумму $m(x)$ по *всем* словам длины n. Получится некоторое действительное число ω, не превосходящее единицы. Пусть $\bar{\omega}$ — его приближение снизу с точностью $2^{-\alpha}$ (первые α битов оставляем, остальное выбрасываем). Все слагаемые в сумме $\omega = \sum\{m(x) \mid l(x) = n\}$ перечислимы снизу; будем строить их рациональные приближения и остановимся, как только суммы этих приближений превысят $\bar{\omega}$.

В результате мы получим некоторую меру P с рациональными значениями на множестве всех слов длины n, имеющую сложность α (поскольку она вычислимо строится по слову $\bar{\omega}$ длины α). Мера P отличается от априорной вероятности не более чем на $2^{-\alpha}$ (суммарно по всем словам).

Задача 347 позволяет использовать в определении стохастичности произвольные меры (а не только равномерные) с погрешностью $O(\log n)$ в параметрах. Остаётся показать, что суммарная априорная вероятность всех слов, для которых дефект относительно P больше β, не превосходит $2^{-\alpha}$. В самом деле, для таких слов x мы имеем

$$-\log P(x) - KS(x\,|\,P) > \beta.$$

Сложность P не больше α и потому $KS(x)$ превосходит $KS(x\,|\,P)$ не более чем на α, откуда

$$-\log P(x) - KS(x) > \beta - \alpha.$$

Поскольку члены порядка $O(\log n)$ мы игнорируем, можно заменить обычную сложность на префиксную:

$$-\log P(x) - KP(x) > \beta - \alpha.$$

Вспоминая, что префиксная сложность связана с априорной вероятностью, получаем, что

$$\log(m(x)/P(x)) > \beta - \alpha$$

для любого слова x с дефектом (относительно P) больше β. По условию $\alpha < \beta$ с запасом, достаточным для компенсации всех допущенных нами ошибок, так что можно считать, что для всех таких слов выполнено неравенство $P(x) < m(x)/2$, или $(m(x) - P(x)) > m(x)/2$. Вспоминая, что сумма $m(x) - P(x)$ по всем вообще x не превосходит $2^{-\alpha}$ по построению меры P, заключаем, что и сумма всех $m(x)$ по словам дефекта больше β относительно P не превосходит $2^{-\alpha+O(\log n)}$, что и требовалось. ►

Понятие стохастического объекта можно рассматривать как конечный аналог понятия последовательности, случайной по некоторой вычислимой мере. Тут есть не только аналогия, но и прямая связь, как показывает следующая задача.

348 Покажите, что если последовательность ω случайна в смысле Мартин-Лёфа относительно некоторой вычислимой меры, то её начальные отрезки длины n являются $(O(\log n), O(\log n))$-стохастическими словами. [Указание: можно воспользоваться определением стохастичности с мерами.]

Выведите отсюда, что существуют последовательности, не случайные в смысле Мартин-Лёфа ни по какой вычислимой мере. [Указание: добавление начального отрезка малой длины не сильно нарушает нестохастичность.]

14.3. Двухчастные описания

Можно оценивать качество статистических гипотез и в немного другой системе координат. Начнём с такого замечания. Если слово x является элементом конечного множества A, то можно задать это слово, указав:

- множество A;
- порядковый номер слова x в этом множестве (относительно какого-либо естественного порядка на словах, скажем, алфавитного).

Из этого следует, что $KS(x) \leqslant KS(A) + \log|A|$ для любого элемента x любого конечного множества A (опять же с логарифмической точностью).

Одно и то же слово x может иметь много таких «двухчастных» описаний. Какие из них лучше, какие хуже? Ясно, что вообще-то мы хотели бы сделать обе части описания как можно меньше: если удалось найти более простое множество того же размера или найти меньшее множество той же сложности (среди множеств, содержащих x), то это хорошо. Но как сравнивать простое, но большое множество с меньшим, но более сложным? Одна из возможностей — сравнить длины соответствующих двухчастных описаний и предпочесть более короткое. (По-английски этот принцип сравнения называют Minimal Description Length principle, или MDL.)

Следующее простое наблюдение показывает, что мы можем перераспределять информацию между частями двухчастного описания, перемещая её из второй части в первую (уменьшая размер множества за счёт увеличения сложности).

Теорема 252. *Пусть конечное множество A содержит слово x и $i < \log|A|$. Тогда можно найти конечное множество A', содержащее x, для которого $|A'| \leqslant |A|/2^i$ и $KS(A') \leqslant KS(A) + i + O(\log i)$.*

◄ Запишем элементы A в каком-то фиксированном порядке (например, алфавитном) и разобьём их на 2^i частей (первые $|A|/2^i$ элементов, следующие и т.д.; мы опускаем очевидные оговорки для случая, когда нацело не делится). В качестве A' возьмём часть, в которую попало слово x. Чтобы задать её, достаточно указать A, число i и номер части, на который нужно не более i битов. ►

Это утверждение удобно иллюстрировать картинкой. Для данного слова x рассмотрим множество P_x всех пар $\langle k, l\rangle$, при которых существует конечное множество $A \ni x$ с $KS(A) \leqslant k$ и $\log|A| \leqslant l$. Из определения ясно, что вместе с каждой точкой в это множество входят все точки справа от неё (с бóльшими k) и все точки сверху от неё (с бóльшими l). Только что доказанная теорема показывает, что можно сдвигаться вправо-вниз на вектор $\langle k, -k\rangle$ (с логарифмической точностью).

Как мы увидим, обратное движение возможно далеко не всегда, так что из двухчастных описаний одной и той же (суммарной) длины более ценным является описание с более простым множеством большего размера (из него можно получить другие, но не наоборот).

Схематически строение множества P_x можно описать так (все утверждения с логарифмической точностью). В него входит точка $\langle 0, l(x)\rangle$, соответствующая множеству A всех слов той же длины, что и x. Кроме того, в него входит точка $\langle KS(x), 0\rangle$, соответствующая $A = \{x\}$. Множество P_x ограничено кривой, идущей от первой точки до второй, причём эта кривая не заходит в треугольник $k + l \leqslant KS(x)$ (пунктирная линия) и всюду идёт вниз под углом 45° или более, согласно теореме 252 (рис. 14.1).

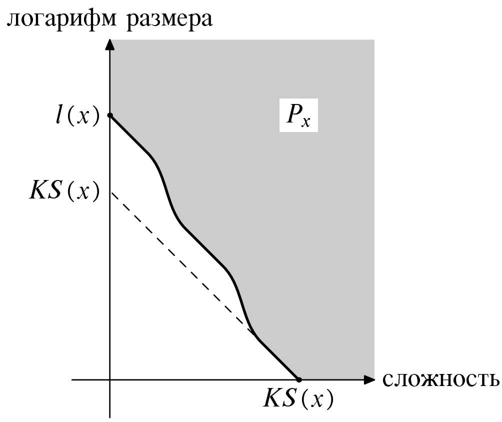


Рис. 14.1. Множество P_x.

При взгляде на рисунок возникает естественный вопрос: какие граничные кривые возможны? Может ли, например, граница области P_x пройти по пунктирной линии? Может, для этого достаточно в качестве x взять случайное слово, к которому приписать недостающие (до желательной длины) нули. Тогда точка $(0, KS(x))$ в начале пунктирной кривой реализуется множеством A, состоящим из всех слов, на конце которых стоят эти самые нули. По теореме 252 угол спуска границы P_x не может быть меньше 45°. А поскольку граница P_x не может опуститься ниже пунктирной кривой, она совпадает с ней (с логарифмической точностью).

Не столь очевиден ответ на другой вопрос: может ли кривая идти под углом 45° до самого конца, где круто спускаться из точки $\langle KS(x), l(x) - KS(x)\rangle$ к оси абсцисс (в точке $\langle KS(x), 0\rangle$), как сплошная линия на рис. 14.2. Неформально говоря, это означает, что слово x допускает по существу лишь два вида статистических гипотез: множество всех слов длины $l(x)$ (и его части, получаемые по теореме 252), и одноэлементное множество $\{x\}$.

349 Покажите, что в этом случае слово x не будет (α, β)-стохастическим при α, β, малых по сравнению с $KS(x)$ и $n - 2\,KS(x)$.

Оказывается, что возможны не только два крайних случая (они изображены на рис. 14.2), но и все промежуточные: любая (достаточно простая) кривая с указан-

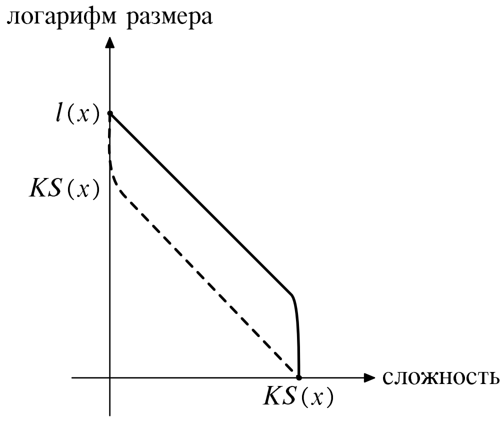


Рис. 14.2. Две граничные кривые.

ными выше свойствами может быть границей множества P_x при некотором x (с логарифмической точностью). Более точно, имеет место следующая

Теорема 253. *Пусть даны $k \leqslant n$ и последовательность*

$$n = t_0 > t_1 > \ldots > t_k = 0$$

из натуральных чисел, имеющая сложность m. Тогда существует слово x сложности $k + O(\log n) + O(m)$ и длины $n + O(\log n)$, для которого множество P_x отстоит от множества $T = \{\langle i,j\rangle \mid (i \leqslant k) \Rightarrow (j \geqslant t_i)\}$ не более чем на $O(\log n) + O(m)$.

(Говорят, что множество P отстоит от множества Q не более чем на ε, если P содержится в ε-окрестности Q и наоборот.)

◄ Мы уже говорили, что из этой теоремы следует существование нестохастических слов. Поэтому не удивительно, что её доказательство использует тот же приём, что и доказательство теоремы 248.

Для каждого i от 0 до k составим список всех множеств сложности не больше i и размером не больше 2^{t_i}; объединив все такие множества (при данном i), получим некоторое множество S_i, состоящее не более чем из 2^{i+t_i} элементов. (Здесь и далее при оценке числа элементов мы опускаем постоянные и полиномиальные по n множители, поскольку они соответствуют слагаемым $O(\log n)$ в оценках длин и сложностей.) Поскольку t_i убывают (что соответствует наклону 45° на рисунке), то $i + t_i$ не возрастают с ростом i, и потому любое из S_i содержит не более $2^{t_0} = 2^n$ элементов. Объединение всех S_i также содержит не более чем 2^n элементов (с точностью до полиномиального множителя), и потому среди слов длины n (точнее, $n + O(\log n)$) есть слово, не входящее ни в одно из множеств S_i. Первое (в каком-либо, например, алфавитном порядке) такое слово и будет искомым словом x.

По построению P_x лежит выше кривой t_i (содержится в множестве T). Остаётся оценить сложность слова x и показать, что множество P_x прилегает к кривой (то есть что T содержится в окрестности P_x).

Верхняя оценка сложности слова x следует из того, что список всех объектов сложности не больше k с указанием их сложности имеет сложность $k + O(\log k)$ (достаточно знать k и самую долгоиграющую программу длины не больше k, см. с. 36); помимо этого списка нам надо знать последовательность $t_0, t_1, \ldots, t_k$, имеющую сложность m.

Нижняя оценка: меньше k сложность слова x быть не может, так как все одноэлементные множества такой сложности выброшены.

Остаётся показать, что при любом $i \leqslant k$ слово x можно поместить в множество A сложности i (или чуть больше) и размера 2^{t_i} (или чуть больше). Искомое множество A строится «с нескольких попыток».

Вначале возьмём первые 2^{t_i} слов нужной длины в качестве A и параллельно развернём процесс порождения множеств сложности не больше j и размера не больше 2^{t_j} при всех $j = 0, \ldots, k$. Для каждого j получится список множеств, который мы будем называть j-списком. В ходе этого процесса будут обнаруживаться подлежащие выбрасыванию слова — элементы всех множеств во всех списках; появление нового множества приводит к одномоментному выбрасыванию всех входящих в него слов. Пока не все слова из A оказались выброшенными (пока объединение порождённых множеств не покроет целиком множество A), мы можем не беспокоиться, поскольку первое по порядку невыброшенное слово принадлежит A. Как только все слова из A будут выброшены, придётся заменить A, взяв в качестве нового кандидата первые 2^{t_i} невыброшенных (на данный момент) слов нужной длины. После этого мы спокойно ждём до тех пор, пока все слова нового множества A не будут выброшены (не попадут в объединение множеств из всех j-списков для всех j); когда и если это случится, мы заменим A, взяв первые 2^{t_i} пока ещё не выброшенных слов, и так далее.

Рано или поздно этот процесс выбрасывания остановится, и текущее значение множества A на этот момент будет содержать наше слово x. Таким образом, мы нашли множество правильного размера, содержащее x, но какова сложность этого множества?

Чтобы описать процесс порождения выбрасываемых множеств, достаточно знать последовательность $t_0, \ldots, t_k$ (и длину слов, равную примерно n), то есть всего $m + O(\log n)$ битов информации. Поэтому с интересующей нас точностью сложность каждого из кандидатов на роль множества A оценивается логарифмом его порядкового номера (сколько кандидатов перед ним пришлось отбросить). Таким образом, нам осталось доказать, что число отброшенных кандидатов ненамного больше 2^i; тогда сложность любого из кандидатов (в том числе и последнего, правильного) будет ненамного больше i.

Итак, сколько раз мы могли заменять множество A в связи с появлением новых множеств в списках? Отдельно посмотрим на появление множеств небольшой сложности и большого размера, а именно, входящих в j-списки с $j \leqslant i$. Такое событие могло случиться не более $O(2^i)$ раз, так как в j-списке не более $O(2^j)$ множеств. Поэтому мы можем считать лишь интервалы между двумя соседними изменения-

ми множества A, в течение которых таких событий не случалось. Это значит, что все элементы предыдущего A были покрыты множествами из j-списков при $j > i$. Во всех этих множествах не так много элементов: общее количество слов во всех множествах j-списка не больше $2^j \cdot 2^{t_j}$; поскольку $t_j + j \leqslant t_i + i$ и количество разных значений j невелико, получается всего примерно $O(2^{t_i+i})$ элементов, и после деления на $|A| = 2^{t_i}$ остаётся 2^i интервалов. ►

Эта теорема показывает, что, помимо сложности $KS(x)$, слово x может иметь другие «сложностные характеристики»; такой характеристикой можно считать границу множества P_x (или, если угодно, само множество). Эта характеристика, если можно так выразиться, «бесконечномерна» (представляет собой не одно число и не набор из двух, трёх или скольких-то чисел, а целую кривую).

350 Докажите, что в теореме 253 нельзя потребовать, чтобы множество P_x отстояло от T на расстояние $O(\log n)$ (а не на $O(\log n) + O(m)$). [Указание. Слов длины $n + O(\log n)$ существенно меньше, чем множеств T, удовлетворяющих условию теоремы.]

351 Докажите, что не существует алгоритма, который по всякому данному слову x находил границу множества P_x с точностью $O(\log l(x))$.

Более сильные результаты о невычислимости функции границы множества P_x по x имеются в работе [177].

Классификацию слов по сложности можно представлять себе как вложенную цепочку множеств $S_0 \subset S_1 \subset S_2 \subset \ldots$, где S_i — множество слов сложности меньше i. В этой цепочке множества S_i перечислимы равномерно по i и содержат $O(2^i)$ элементов.

Теперь вместо этой классификации у нас есть двумерная классификация множеств $S_{i,j}$, где множество $S_{i,j}$ состоит из всех слов x, входящих в конечные множества A с $KS(A) < i$ и $\log|A| < j$. (Такое множество A мы будем называть $(i * j)$-*описанием* любого его элемента x.) Мы получаем двумерную таблицу из множеств $S_{i,j}$, монотонную по обоим направлениям ($S_{i,j}$ растёт с ростом i и с ростом j). Кроме того, теорема 252 показывает, что эта таблица монотонна по диагонали: $S_{i,j} \subset S_{i+k,j-k}$. (Как всегда, мы пренебрегаем логарифмическими поправками: точнее следовало бы написать $S_{i,j} \subset S_{i+k+O(\log k),j-k}$.)

Чтобы лучше понять смысл этой двумерной «стратификации», полезно посмотреть на эквивалентные определения множеств $S_{i,j}$. При этом мы будем, как обычно, пренебрегать логарифмическими слагаемыми и считать два варианта определения (обозначаемые S и S') эквивалентными, если $S_{i,j} \subset S'_{i+O(\log n),j+O(\log n)}$ и наоборот (здесь и далее $n = i + j$).

Говоря о «перечне» в формулировке следующей теоремы, мы подразумеваем алгоритм, который работает и время от времени выдаёт на выход некоторые двоичные слова (возможно, с повторениями); длиной перечня мы будем называть общее число выданных слов (с учётом кратностей). В условии (в) мы считаем, что алгоритм может выдавать на выход слова не поодиночке, а «порциями» произвольного размера (который может меняться от порции к порции).

Теорема 254. *Следующие свойства слова x эквивалентны в описанном смысле* (*из одного следует другое с логарифмической добавкой к параметрам*):

(а) *слово x принадлежит $S_{i,j}$* (*имеет* ($i * j$)*-описание*);

(б) *существует простой* (*сложности $O(\log n)$*) *перечень длины не более 2^{i+j}, в котором слово x в первый раз появляется за 2^j или более шагов от конца*;

(в) *существует простой* (*сложности $O(\log n)$*) *перечень длины не более 2^{i+j}, включающий в себя x, в котором слова выдаются не более чем в 2^i порций*;

(г) *в любом простом* (*сложности $O(\log n)$*) *перечне, включающем все слова сложности не более $i + j$, слово x встречается за 2^j или более шагов от конца.*

◀ Пусть выполнено (а). Будем перечислять все множества сложности не более i и размера не более 2^j. Когда очередное множество появляется, добавляем в перечень все его элементы. При этом получится не более чем 2^i порций с 2^j (или меньше) элементов в каждой, так что всего будет не более чем 2^{i+j} элементов, как и требуется в условии (в). При этом сложность перечня логарифмическая, так как нужно знать лишь i и j. Итак, (а) $\Rightarrow$ (в).

Чтобы получить (б), надо немного модифицировать конструкцию и добавлять после каждой порции 2^j произвольных ещё не порождённых элементов; общее число элементов увеличится на 2^{i+j}, что в пределах допустимого. Таким образом, (а) $\Rightarrow$ (б).

Напротив, из (б) легко следует (а), достаточно лишь нарезать элементы перечня на части размера 2^j; при этом получится не более чем 2^i частей и останутся ненарезанными не более чем 2^j элементов. Следовательно, x войдёт в одну из частей. Каждая часть задаётся своим порядковым номером и потому имеет сложность i (плюс логарифмическая добавка, включающая в себя сложность перечня).

Чтобы получить (а) из (в), при появлении каждой новой порции элементов перечня мы разбиваем её на части из 2^j элементов или меньше (для последней, неполной, части в каждой порции). Количество полных частей не превосходит 2^i, поскольку длина перечня не превосходит 2^{i+j}. Поэтому сложность каждой полной части не больше i (описываем каждую полную часть её номером в порядке, задаваемым перечнем). Количество неполных частей также не больше 2^i, поэтому сложность каждой из них также не больше i.

Итак, свойства (а) – (в) равносильны (с логарифмической точностью), и осталось показать, что они равносильны (г). Очевидно, что (г) является усилением (б), так что достаточно проверить, что из (а) следует (г).

Итак, пусть слово x содержится в некотором конечном множестве A сложности не более i, причём $\log|A| \leqslant j$. Все элементы множества A, в том числе и x, имеют сложность не более $i + j + O(\log(i + j))$; пренебрегая логарифмическими поправками, мы будем опускать последнее слагаемое (надеясь, что привыкший к этому читатель легко внесёт необходимые исправления).

Пусть также дан некоторый простой перечень, включающий в себя все слова сложности не более $i + j$. Мы хотим доказать, что слово x появится в этом перечне не слишком поздно (после него будут ещё как минимум 2^j слов). Зная множество A, мы можем порождать перечень, пока в нём не обнаружатся все элементы A. Порождённая к этому моменту часть перечня представляет собой конечное множество B

сложности не более i (поскольку для задания B достаточно знать A и сам перечень, который прост). Рассмотрим первые (по порядку) 2^j слов вне B: они определяются множеством B (i битов) и своим порядковым номером (j битов) и потому имеют сложность не более $i + j$; это и будут искомые 2^j элементов перечня, стоящие в нём после x. ▶

Можно сказать, что мы ввели дополнительную классификацию слов сложности не больше n, измеряя логарифм расстояния до конца списка таких слов. С точки зрения двумерной таблицы $S_{i,j}$ это соответствует возрастающей последовательности множеств $S_{i,j}$ на диагонали $i + j = n$ (строго говоря, возрастание множеств вдоль этой диагонали — при увеличении i и соответствующем уменьшении j — имеет место лишь с точностью до логарифмических поправок). Случайные слова длины $l \leqslant n - O(\log n)$ (то есть слова сложности l и длины l) стоят в начале этой классификации, имея $(n * 0)$-описания; в конце её стоят (немногочисленные) слова, имеющие лишь $(0 * n)$-описания.

$\boxed{\mathbf{352}}$ Покажите, что все слова сложности n, стоящие в самом конце списка (с логарифмической точностью, то есть слова, за которыми стоит $\mathrm{poly}(n)$ слов), «почти одинаковы»: они имеют условную сложность $O(\log n)$ друг относительно друга.

Покажите, что среди них находятся $B(n - c \log n)$ и $BB(n - c \log n)$, рассмотренные в разделе 1.2 (с. 32 и 34), для любых положительных c и n.

Если угодно, можно считать n минус логарифм расстояния до конца списка слов сложности не выше n мерой «извращённости» слова: случайные слова длины не более $n - O(\log n)$ при этом имеют нулевую извращённость, а слова у конца списка — максимальную (близкую к n). Но надо иметь в виду два важных обстоятельства.

- «извращённость» слов x и y может сильно отличаться, если даже $KS(x \mid y) \approx 0$ и $KS(y \mid x) \approx 0$. (В самом деле, кратчайшее описание y любого слова x является случайным и находится в самом начале соответствующего списка.) Однако если слова x и y переходят друг в друга при простой вычислимой биекции, это уже не так (см. следующую задачу).

- «извращённость» данного слова x сложности n (определяемая его местом в перечне всех слов сложности не выше n) может сильно упасть, если мы рассмотрим то же слово x в перечне слов сложности не больше n' при $n' > n$. Так что фактически характеристикой слова x является функция $n' \mapsto$ расстояние от x до конца перечня слов сложности не выше n'. По существу это та же самая граничная кривая множества P_x, но только в других координатах (теперь мы смотрим на точки, в которых диагонали $i + j = n'$ заходят в множество P_x).

$\boxed{\mathbf{353}}$ Пусть слова x и y соответствуют друг другу при вычислимой биекции сложности t. Докажите, что если $x \in S_{i,j}$, то $y \in S_{i+O(t),j}$.

Напомним, что простая вычислимая биекция, переводящая одно слово в другое, существует, когда тотальная сложность любого из слов относительно второго мала (задача 31 на с. 45).

Возвращаясь к теореме 15 (с. 36), можно сказать, что эта теорема выделяла среди слов сложности n некоторые особые слова (все имеющие малую сложность относительно друг друга). Теперь мы можем объяснить смысл этой «особой точки» ещё одним способом: это слова, стоящие у конца списка и эквивалентные им слова. Среди эквивалентных, отметим, есть и стохастические, длина которых близка к сложности. Такова, например, двоичная запись числа k_n слов сложности не больше n (пункт (б) в теореме 15) или числа m_n слов длины не более n, на которых определён оптимальный декомпрессор.

Кстати, использованное при доказательстве теоремы 254 рассуждение позволяет установить связь между k_n (или m_n) при разных n, как показывает следующая задача:

$\boxed{\textbf{354}}$ Докажите, что при $n' < n$ слово $k_{n'}$ (точнее следовало бы сказать «слово, являющееся двоичной записью числа $k_{n'}$) эквивалентно (с точностью до условной сложности $O(\log n)$) первым n' битам слова k_n. Докажите аналогичное утверждение для слов m_n и $m_{n'}$. [Указание. Для k_n по существу надо доказать, что указание списка слов сложности не больше n с не более чем 2^s пропусками равносильно указанию числа, большего $B(n-s)$. В одну сторону: зная такой список, мы дождёмся T шагов, пока в перечислении слов сложности не больше n появятся все элементы этого списка. Любое число $t > T$ должно иметь сложность не меньше $n-s$, иначе по этому числу можно было построить список сложности меньше $n-s$, содержащий все слова сложности n, кроме 2^s, и первые 2^s слов вне этого списка приводили бы к противоречию. В другую сторону: будем перечислять слова сложности не больше n, выдавая их порциями по 2^s штук; зная $B(n-s)$, можно найти число полных порций (шаг, на котором появляется последняя полная порция, не превосходит $B(n-s)$, поскольку определяется числом полных порций и имеет сложность не больше $n-s$) и тем самым получить искомый список с точностью до последней неполной порции. Аналогичное рассуждение годится и для m_n.]

Следующий результат обобщает утверждение задачи 39 (с. 50). Там по существу утверждалось, что если слово x имеет много описаний данной длины, то оно имеет и описания меньшей длины. Оказывается, что аналогичное утверждение верно и для $(i*j)$-описаний.

Теорема 255. *Пусть для слова x существует не менее 2^k множеств, являющихся его $(i*j)$-описаниями. Тогда оно имеет $(i*(j-k))$-описание и даже $((i-k)*j)$-описание.*

В формулировке этой теоремы мы, как и раньше, опускаем слагаемые вида $O(\log(i+j+k))$, которые разрешается прибавлять к параметрам описаний. Слово «даже» напоминает о теореме 252, которая позволяет от $(i-k)*j$ перейти к $i*(j-k)$.

◄ Первое (более простое) утверждение теоремы легко следует из рассуждений, использованных в доказательстве теоремы 254. Будем перечислять все множества A сложности не более i и размера не более 2^j и смотреть, какие элементы x покрыты ими с кратностью 2^k. Таких элементов будет не более $2^{i+j}/2^k = 2^{i+j-k}$, и они выдаются не более чем в 2^i порций (каждое новое множество A соответствует одной порции). Остаётся вспомнить свойство (в) теоремы 254.

Для доказательства второго утверждения нам понадобится уменьшить число порций до 2^{i-k}. Вот как это делается. По-прежнему перечисляя множества сложности не более i и размера не более 2^j, мы теперь обращаем внимание не только на «полноценные» элементы, покрытые с кратностью 2^k, но и на «кандидатов» — элементы, покрытые с вдвое меньшей кратностью (2^{k-1}). Как только появляется полноценный элемент, не вошедший в перечень, мы включаем в этот перечень не только его, но заодно уж и всех обнаруженных к этому моменту кандидатов (кроме уже включённых в перечень). От этого общее число включённых в перечень элементов увеличится не более чем вдвое (что несущественно при нашей логарифмической точности). Зато число порций сильно уменьшится: ведь каждая из них включает в перечень всех кандидатов, и чтобы появился непокрытый полноценный элемент, нужно не менее 2^{k-1} новых множеств (кратность должна возрасти по меньшей мере от 2^{k-1} до 2^k). Это и даёт необходимое уменьшение числа порций. ▶

Только что доказанную теорему можно переформулировать следующим образом:

Теорема 256. *Если слово x имеет $(i * j)$-описание A, для которого выполнено неравенство $KS(A|x) \geqslant k$, то оно имеет и $(i * (j - k))$-описание и даже $((i - k) * j)$-описание.*

(Как и раньше, мы опускаем логарифмические слагаемые, необходимые в точной формулировке.)

◀ В самом деле, зная слово x, а также значения параметров i и j (последнее требует логарифмического числа битов), мы можем перечислять все $(i * j)$-описания слова x. Поэтому сложность таких описаний (с точностью до логарифмических слагаемых) не превосходит их числа. Если есть описание A с большим $KS(A|x)$, то описаний много и можно применить предыдущую теорему. ▶

Эта теорема гарантирует, что если мы берём описания с предельно возможными параметрами (на границе множества P_x для данного x), то все эти описания будут просты относительно x. Что, с точки зрения здравого смысла, довольно естественно: если в описании есть «лишняя» (не входящая в x) информация, то вряд ли оно будет оптимальным.

14.4. Ограниченные классы гипотез

В этом разделе мы будем предполагать, что в качестве статистических гипотез разрешены только множества из некоторого семейства множеств $\mathscr{A}$. Ограничение класса гипотез означает, что у нас заранее имеется некоторая информация об источнике происхождения данного слова x. А именно, нам известно, что слово x было получено случайным выбором в одном из множеств из $\mathscr{A}$ (но мы не знаем, в каком). Поэтому, подыскивая статистическую гипотезу для x, мы ограничиваемся множествами, принадлежащими $\mathscr{A}$.

Оказывается, основные результаты предыдущих параграфов обобщаются (с некоторым ухудшением оценок) на случай любых семейств $\mathscr{A}$, удовлетворяет следующим трем условиям.

(1) Семейство $\mathscr{A}$ перечислимо. Это означает, что имеется алгоритм, печатающий в некотором порядке списки элементов всех множеств из $\mathscr{A}$; список очередного множества можно начинать только тогда, когда уже закончен список предыдущего.

(2) Для всех n семейство $\mathscr{A}$ содержит множество всех слов длины n.

(3) Для некоторого полинома p, для каждого множества $A \in \mathscr{A}$ и для любых натуральных чисел n и $c < |A|$ выполнено следующее: существует покрытие множества всех слов длины n из A не более чем $p(n)|A|/c$ множествами из $\mathscr{A}$, причем мощность каждого из покрывающих множеств не превосходит c.

Поясним смысл последнего условия. Теорема 252 утверждает, что если слово имеет $(i * j)$-описание, то оно имеет и $((i + k) * (j - k))$-описание (для всякого $k \leqslant j$). В доказательстве этой теоремы мы разбивали данное $(i * j)$-описание на 2^k частей, каждая размером не больше 2^{j-k}. Третье условие гарантирует возможность такого разбиения, правда, на немного большее количество кусков (в полиномиальное от n количество раз).

Пусть дано слово x. По аналогии с множеством P_x, определённым на с. 476, обозначим через $P_x^{\mathscr{A}}$ множество пар $\langle i, j\rangle$, для которых x имеет $(i * j)$-описание, *принадлежащее семейству* $\mathscr{A}$. Если семейство $\mathscr{A}$ состоит из всех конечных подмножеств, то никаких ограничений нет, и тем самым $P_x^{\mathscr{A}} = P_x$. Для ограниченных семейств множество $P_x^{\mathscr{A}}$ всегда включено в множество P_x и чем меньше семейство $\mathscr{A}$, тем меньше множество $P_x^{\mathscr{A}}$.

Условия (1)–(3) гарантируют, что множества $P_x^{\mathscr{A}}$ обладают теми же свойствами, что и множества P_x. А именно, для любого слова длины n выполнено следующее:

- Множество $P_x^{\mathscr{A}}$ содержит пару, отстоящую от пары $\langle 0, n\rangle$ не более, чем на $O(\log n)$. Действительно, по свойству (2) семейство $\mathscr{A}$ содержит множество всех слов длины n, а значит любое слово длины n имеет $0 * n$-описание в $\mathscr{A}$ (игнорируя логарифмические добавки).
- Множество $P_x^{\mathscr{A}}$ содержит пару, отстоящую от пары $\langle KS(x), 0\rangle$ не более чем на константу. Действительно, из условия (3), примененного к $c = 1$ и множеству всех слов длины n в качестве A, следует, что $\mathscr{A}$ содержит все синглетоны, поэтому каждое слово имеет $(KS(x) * 0)$-описание (с точностью до $O(1)$-добавок).
- Имеет место аналог теоремы 252: если x имеет $(i * j)$-описание в $\mathscr{A}$, то для всех $k \leqslant j$ оно имеет и $((i + k) * (j - k))$-описание в $\mathscr{A}$. (В аккуратной формулировке к $i + k$ надо добавить еще $O(\log l(x))$.) Действительно, пусть x имеет $(i * j)$-описание $A \in \mathscr{A}$. Пусть нам дано $n = l(x)$, множество A и k. Будем порождать множества из семейства $\mathscr{A}$ до тех пор, пока не найдем покрытие множества всех слов длины n из A не более чем $p(n)2^k$ множествами мощности $|A|2^{-k}$ или менее (здесь p — полином из условия (3)). Сложность множества, покрывающего x, не превосходит $i + k + O(\log n + \log k)$, поскольку его можно найти, зная A, n, k и его порядковый номер среди множеств, покрывающих A. Без ограничения общности мы можем считать, что $k \leqslant n$, поскольку иначе утверждение очевидно — искомым описанием является синглетон $\{x\}$. Поэтому слагаемое $O(\log k)$ можно опустить.

Пример. Пусть семейство $\mathscr{A}$ состоит из всех *шаров Хемминга*, то есть множеств вида $\{x \mid l(x) = l(y), d(x,y) \leqslant r\}$, где y — некоторое слово (называемое центром шара), r — некоторое натуральное число, называемое радиусом шара, а $d(x,y)$ обозначает расстояние Хемминга (количество позиций, в которых x и y различны). Неформально говоря, нам известно, что для некоторого слова y и некоторого натурального r слово x было получено из y изменением не более чем r битов, при этом номера измененных битов выбирались случайно. Сами y и r при этом неизвестны. Например, слово y было передано нам по ненадежному каналу связи и полученное нами вместо y слово x отличается от y в небольшом количестве позиций.

355 Докажите, что для любого $r \leqslant n$ множество всех слов длины n можно покрыть $\mathrm{poly}(n)2^n/V$ шарами Хемминга радиуса r, где V обозначает мощность шара радиуса r. [Указание. Выберем случайным образом N шаров радиуса r. Для фиксированного слова x вероятность того, что оно не будет покрыто ни одним из выбранных шаров равна $(1 - V2^{-n})^N < e^{-V2^{-n}N}$. Если положить $N = n \ln 2 \cdot 2^n/V$, то эта верхняя оценка будет равна 2^{-n}. Значит при таком N вероятность того, что хотя бы одно слово длины n не будет покрыто, меньше 1.]

356 Докажите, что семейство всех шаров Хемминга удовлетворяет условиям (1)–(3). [Указание. Пусть дан шар A радиуса a и число $c < |A|$. Нам нужно покрыть A шарами мощности c или меньше. Поскольку множество всех слов длины n можно покрыть двумя шарами радиуса $n/2$, без ограничения общности мы можем считать, что $a \leqslant n/2$. Выберем наибольшее $b \leqslant n/2$ такое, что мощность $|B|$ шара радиуса b не превосходит c, и будем искать покрытие A шарами радиуса b. Мощности шаров Хемминга с радиусами, различающимися на 1, отличаются не более, чем в $n+1$ раз, поэтому $|B| \geqslant c/(n+1)$ и нам достаточно покрыть A не более чем $\mathrm{poly}(n)|A|/|B|$ шарами радиуса b.

Покроем все точки на расстоянии не более b от центра шара A одним шаром радиуса b (с тем же центром). Остаток шара разобьем на концентрические сферы: каждая сфера состоит из всех точек на расстоянии ровно d от центра шара A, где $d \in \{b+1, b+2, \ldots, a\}$ Количество сфер не превосходит n, поэтому достаточно научиться покрывать сферу S радиуса $d \in (b, n/2]$ не более чем $\mathrm{poly}(n)|S|/|B|$ шарами радиуса b.

Для этого в качестве f возьмём решение уравнения $b + f(1 - 2b/n) = d$, округлённое до ближайшего целого числа. Выберем случайным образом шар B радиуса b с центром на расстоянии f от центра S. Не менее $1/\mathrm{poly}(n)$-й части точек любого такого шара B принадлежат S. Действительно, обозначим через x и y центры S и B, соответственно. Выберем в слове y некоторую (b/n)-ю часть битов, в которых y совпадает с x, и некоторую (b/n)-ю часть битов, в которых y отличается от x, и изменим все выбранные биты. Любое слово, полученное таким образом, расположено на расстоянии b от y и на расстоянии $f - (b/n)f + (n-f)(b/n) = d$ от x. Общее количество таких слов равно $C_f^{f(b/n)} C_{n-f}^{(n-f)(b/n)}$, что с точностью до умножения на полином от n равно

$$2^{fh(b/n,1-b/n)+(n-f)h(b/n,1-b/n)} = 2^{nh(b/n,1-b/n)}.$$

Мощность $|B|$ шара радиуса b также равна этому числу (с той же точностью). Итак, каждый шар B радиуса b с центром на расстоянии f от x покрывает не менее $|B|/\operatorname{poly}(n)$ точек из S. При случайном выборе шара B каждое $z \in S$ имеет равные шансы быть покрытым. Поэтому вероятность того, что любое фиксированное $z \in S$ покрывается случайным шаром B, не меньше $|B|/(|S|\operatorname{poly}(n))$. Следовательно, для подходящего полинома с большой вероятностью случайно выбранные $\operatorname{poly}(n)|S|/|B|$ шаров радиуса b с центром на расстоянии f от x покроют все точки S.]

$\boxed{357}$ Пусть семейство $\mathscr{A}$ состоит из всех шаров Хемминга. Докажите, что существуют такие слова x, для которых множество $P_x^{\mathscr{A}}$ значительно меньше множества P_x: для некоторого положительного ε для всех достаточно больших n существует слово x длины n, для которого $P_x^{\mathscr{A}}$ отстоит от P_x более чем на εn. [Указание. Фиксируем любое $0 < \alpha < 1/2$ и обозначим через V мощность шара радиуса αn. Для данного n найдём перебором некоторое множество E, состоящее из $N = 2^n/V$ слов длины n, для которого любой шар Хемминга радиуса αn содержит не более n слов из E. (Существование такого множества можно доказать вероятностными рассуждениями: если выбрать случайно N слов длины n, то они с положительной вероятностью образуют нужное множество.) Возьмем в качестве x случайное слово из E, то есть, любое слово x в E сложности примерно $\log|E| = n - \log V$. Сложность любого шара A радиуса αn, содержащего x, не меньше $KS(x) = n - \log V$ (с точностью $O(\log n)$). В самом деле, зная шар A и номер x в $A \cap E$, можно найти x. Поэтому точка $(n - \log V, \log V)$ находится в логарифмической окрестности границы множества $P_x^{\mathscr{A}}$. С другой стороны, граница множества P_x находится ниже прямой $KS(A) + \log|A| = n - \log V$. Это следует из того, что x имеет $(0 * (n - \log V))$-описание (с логарифмической точностью) — это само множество E.]

$\boxed{358}$ Опишите множество $P_x^{\mathscr{A}}$ для x и $\mathscr{A}$ из предыдущей задачи. [Указание. Граница множества $P_x^{\mathscr{A}}$ состоит из вертикального отрезка $KS(A) = n - \log V$, $\log|A| \leqslant \log V$ и отрезка $KS(A) + \log|A| = n$, $\log V \leqslant \log|A|$, наклонённого под углом в 45° к осям координат.]

Пусть семейство $\mathscr{A}$ обладает свойствами (1)–(3). Тогда выполнен аналог теоремы 253 с немного худшей точностью: $O(\log n)$ надо заменить на $O(\sqrt{n \log n})$.

Теорема 257. *Пусть $k \leqslant n$ и пусть $n = t_0 > t_1 > \ldots > t_k = 0$ — последовательность натуральных чисел, имеющая сложность m. Тогда существует слово x сложности $k + O(\sqrt{n \log n}) + O(m)$ и длины $n + O(\log n)$, для которого множество $P_x^{\mathscr{A}}$ отстоит от множества $T = \{\langle i, j\rangle \mid (i \leqslant k) \Rightarrow (j \geqslant t_i)\}$ не более чем на $O(\sqrt{n \log n}) + O(m)$.*

◀ Доказательство похоже на доказательство теоремы 253. Вспомним, как проходило то доказательство. В качестве x мы брали первое в словарном порядке слово подходящей длины n', которое не принадлежит ни одному множеству сложности меньше i и мощности не больше 2^j, если точки (i, j) брать на границе множества T (будем называть такие множества «плохими»). Длина n' выбиралась так, чтобы общее количество слов в плохих множествах было строго меньше количества слов длины n'. Кроме того, для каждой из граничных точек (i, j) мы строили

множество A_{ij}, содержащее x, имеющее мощность 2^j и сложность, близкую к i. Множества A_{ij} строились «с нескольких попыток». А именно, сначала для каждого j мы заносили в A_{ij} первые в словарном порядке 2^j слов. Затем мы запускали процесс перечисления плохих множеств и выбрасывали из всех A_{ij} элементы всех обнаруженных плохих множеств. Как только в результате выбрасывания некоторое множество A_{ij} опустевало, мы заполняли его первыми в словарном порядке 2^j невыброшенными словами. Этим мы гарантировали, что в любой момент каждое из A_{ij} содержит первое в словарном порядке невыброшенное слово. В частности, это верно и после того, как обнаружены все плохие множества и изменение множеств A_{ij} прекратилось. Поэтому после стабилизации каждое из множеств A_{ij} содержит выбранное нами слово x. Каждое A_{ij} можно задать порядковым номером его последнего изменения плюс небольшой объем информации, нужный, чтобы задать j и запустить процесс перечисления плохих множеств. Поэтому нам было достаточно доказать, что логарифм количества изменений A_{ij} лишь ненамного превосходит i. Это следовало из того, что после каждого опустения множество A_{ij} заполнялось заново до нужной мощности. Чтобы опустошить его еще раз, нужно либо появление достаточного количества новых мелких плохих множеств (мощности меньше 2^j), либо появление одного нового крупного (мощности больше 2^j). Из формы границы T следовало, что общее количество элементов в мелких плохих множествах не больше 2^{i+j} (с точностью до полиномиального от n множителя), поэтому они могут опустошать множество A_{ij} не более 2^i раз. А количество крупных плохих множеств не превосходит 2^i, поскольку сложность каждого из них меньше i. (Точнее следует сказать, что количество интервалов между соседними изменениями множества A_{ij}, во время которых не появлялось крупных множеств, не больше 2^i.)

Почему это доказательство, как есть, не переносится на произвольное семейство множеств $\mathscr{A}$? Мы по-прежнему можем порождать плохие множества из нашего семейства, и общее количество элементов в них ограничено той же формулой, поэтому мы можем выбрать ту же длину n' искомого слова x. Причём перечислимость $\mathscr{A}$ нам даже не нужна — не будет никакой беды, если мы учтём и плохие множества вне $\mathscr{A}$. Проблема в том, что теперь мы не можем наполнить опустевшее A_{ij} первыми 2^j невыброшенными словами — полученное множество может не принадлежать семейству $\mathscr{A}$. Из-за этого мы не можем выбрать слово x заранее. Теперь x будет определяться не только множеством слов, входящих в объединение плохих множеств, но и порядком их появления. Нам придётся, наблюдая за их появлением, строить «хорошие» множества A_{ij} с непустым пересечением так, чтобы каждое из них изменялось не слишком много раз.[1] Для последнего важно, чтобы мощность пересечения хороших крупных множеств (мощности не менее 2^j) была близка к 2^j. В этом случае можно будет использовать те же рассуждения, что и раньше. В самом деле, чтобы это пересечение стало пустым, должно появиться много плохих мелких множеств или одно плохое крупное, поэтому пересечение будет редко опустевать, а значит, крупные множества будут меняться редко.

[1] Раньше непустота пересечения хороших множеств гарантировалась непустотой каждого из них — теперь так не получится.

При построении хороших множеств мы будем использовать следующий алгоритм. Допустим, имеются некоторые хорошие множества $A_1, \ldots, A_l$, причём наименьшее из них A_l имеет мощность не более 2^j, а их пересечение содержит не менее $\beta 2^j$ невыброшенных слов. Мы хотим добавить к ним еще одно хорошее множество A_{l+1} мощности не более $2^s < 2^j$ таким образом, чтобы пересечение $A_1, \ldots, A_{l+1}$ содержало побольше невыброшенных слов. Для этого мы покрываем A_l не более чем $\mathrm{poly}(n')2^{j-s}$ множествами из $\mathscr{A}$, мощности не более 2^s каждое, и выбираем среди них то, которое покрывает наибольшее количество невыброшенных слов из пересечения $A_1 \cap \ldots \cap A_l$. Возможность такого покрытия обеспечивается третьим из принятых требований к семейству $\mathscr{A}$, и его можно найти алгоритмически — это обеспечивается первым требованием к семейству. При таком алгоритме выбора A_{l+1} пересечение $A_1, \ldots, A_{l+1}$ содержит не менее

$$\frac{\beta 2^j}{\mathrm{poly}(n')2^{j-s}} = \frac{\beta 2^s}{\mathrm{poly}(n')}$$

невыброшенных слов.

Чтобы построить семейство хороших множеств, содержащее по одному множеству A_{it_i} для каждой точки (i, t_i) на границе T, нужно применить эту процедуру k раз. При каждом применении параметр β уменьшается в $\alpha = \mathrm{poly}(n')$ раз. Таким образом, мы можем гарантировать лишь, что размер пересечения первых i множеств не меньше $\alpha^{-i}2^{t_i}$. При больши́х i это во много раз меньше необходимого 2^{t_i}. Поэтому с помощью этого алгоритма (а другого у нас нет) не построить семейство хороших множеств, содержащее по одному множеству A_{ij} для каждой точки (i, j) на границе T (с соблюдением требования о мощности пересечений). Значит нам придётся рассматривать не все точки на границе, а только небольшую их часть.

Выберем $N = \sqrt{n/\log n}$ точек на границе T, идущих через равные интервалы по второй координате, начиная с точки $(0, n)$ и заканчивая точкой $(k, 0)$. А затем докажем, что найдётся x, для которого для каждой выбранной точки (i, j) наименьшая сложность множества из $\mathscr{A}$ размера 2^j (или меньше) близка к i. Почему этого достаточно? При выбранном нами N вторые координаты соседних выбранных точек отличаются на $\sqrt{n \log n}$. Поскольку граница T спускается под углом в 45° или более, разность первых компонент соседних выбранных граничных точек не больше разности их вторых компонент (а значит, не больше $\sqrt{n \log n}$). В частности, у выбранных точек при росте второй координаты сумма $i + j$ не возрастает (это нам понадобится в дальнейшем). Граница множества T монотонна, поэтому все её точки между любыми соседними выбранными точками близки к выбранным точкам (отстоят от них не более чем на $\sqrt{n \log n}$). То же самое верно и для границы множества $P_x^{\mathscr{A}}$: между соседними горизонталями (из выбранных) все её точки близки к крайним. А крайние точки по построению близки к границе T. Поэтому границы множеств T и $P_x^{\mathscr{A}}$ будут близки, следовательно, и сами множества будут близки.

Теперь перейдём к построению x и хороших множеств. У нас имеется N выбранных граничных точек (i, j), которые мы будем нумеровать i_l, j_l в порядке убывания второй координаты, так что $j_1 = n$ и $j_N = 0$. Сначала для каждого l в порядке возрастания найдём «хорошее» множество A_l из $\mathscr{A}$ мощности не более 2^{j_l} таким

образом, чтобы мощность пересечения первых l хороших множеств была не меньше $\alpha^{-l}2^{j_l}$. Для этого мы N раз применяем описанный выше алгоритм, взяв в качестве A_0 множество всех слов длины n' (по условию оно принадлежит семейству $\mathscr{A}$).

Затем мы начинаем перечислять «плохие» множества. При появлении очередного плохого множества мы выбрасываем все его элементы из множества всех слов длины n' (и из всех хороших множеств). Длину n' выбираем так, чтобы общее количество выброшенных слов никогда не превзошло половины от $2^{n'}$ (ясно, что нам хватит $n' = n + O(\log n)$).

Как только в результате выбрасывания пересечение первых l хороших множеств станет слишком маленьким для хотя бы одного $l \leqslant N$, мы находим наименьшее такое l и обновляем описанным выше алгоритмом все хорошие множества, начиная с l-го, в порядке возрастания номеров. (Нулевым хорошим множеством будем считать множество всех слов длины n' и положим $l_0 = n'$.) Будем называть эту процедуру *l-обновлением*: при l-обновлении хорошие множества с номерами меньше l не меняются, а все остальные обновляются.

Как выбрать порог для размера пересечения первых l множеств, при достижении которого мы производим l-обновление? Его выбор должен обеспечивать верхнюю оценку 2^{i_l} для числа l-обновлений. Для этого нужно, чтобы сразу после обновления l-го хорошего множества мощность пересечения первых l множеств была существенно (скажем, вдвое) больше этого порога. Вспомним, что при обновлении l-го множества порог для $(l-1)$-го множества ещё не перейдён. Поэтому наш алгоритм гарантирует, что пересечение первых l множеств имеет не меньше

$$\nu_{l-1}/\alpha 2^{j_{l-1}-j_l}$$

слов. Таким образом, нам достаточно, чтобы порог для l-го множества был, скажем, вдвое больше этой величины. Кроме того, нам нужно, чтобы порог ν_0 для нулевого хорошего множества A_0 никогда не был перейден. Этим требованиям очевидно удовлетворяют числа

$$\nu_l = (2\alpha)^{-l}2^{j_l}/2.$$

При таком выборе порога после каждого l-обновления для всех изменённых хороших множеств A_r мощность пересечения первых r хороших множеств не меньше $2\nu_r$. Следовательно, между любыми двумя последовательными l-обновлениями должно быть выброшено не менее ν_l слов (более того, для любого l-обновления в интервале между последним s-обновлением для $s \leqslant l$ перед ним и им самим должно быть выброшено не менее ν_l слов).

Убедимся, что это гарантирует желаемую верхнюю оценку для количества изменений l-го хорошего множества. Оно меняется только при s-обновлениях для $s \leqslant l$. Между любыми двумя последовательными s-обновлениями должно быть выброшено не менее ν_s слов, что не меньше ν_l. Выбрасывание ν_l слов может быть вызвано появлением нового «крупного» плохого множества (крупными мы называем множества мощности больше 2^{j_l}), либо порождением достаточного количества мелких плохих множеств (мощности не больше 2^{j_l}). Общее количество крупных множеств меньше $\sum_{r<l} 2^{i_r}$, а общее количество слов в мелких множествах меньше

$\sum_{r\geqslant l} 2^{i_r+j_r}$. Поэтому общее количество s-обновлений для любого фиксированного $s \leqslant l$ ограничено сверху величиной

$$\sum_{r<l} 2^{i_r} + \sum_{r\geqslant l} 2^{i_r+j_r}/\nu_l.$$

(Здесь первое слагаемое есть общее количество интервалов между последовательными s-обновлениями, во время которых появлялись крупные множества, а второе слагаемое есть общее количество интервалов между последовательными s-обновлениями, во время которых не появлялось крупных множеств.) В силу неубывания чисел i_r первое слагаемое не превосходит $N2^{i_l}$. Числа $i_r + j_r$ не возрастают с ростом r, поэтому второе слагаемое не больше $N2^{i_l+j_l}/\nu_l$. Вспомнив, что $\nu_l = (2\alpha)^{-l}2^{j_l-1}$ и $\alpha = \mathrm{poly}(n')$, мы видим, что второе слагаемое не больше

$$N2^{i_l+j_l+N\log 2\alpha-j_l+1} = N2^{i_l+O(N\log n')}.$$

Итак, количество s-обновлений для фиксированного $s \leqslant l$ меньше

$$N2^{i_l+O(N\log n')},$$

следовательно, количество изменений l-го хорошего множества меньше

$$N^2 2^{i_l+O(N\log n')}.$$

Число n' примерно равно n, а N есть $\sqrt{n/\log n}$. Поэтому сложность l-го хорошего множества не больше $i_l + O(\sqrt{n\log n}) + O(m + \log n)$ (мы добавили сложность описания множества T и l).

Когда новые плохие множества перестанут появляться, хорошие множества тоже перестанут изменяться. В качестве x выберем единственное слово из пересечения всех хороших множеств (напомним, что среди них имеется синглетон), полученных к этому моменту. Сложность этого слова находится в нужных пределах, поскольку по построению наименьшая сложность синглетона, содержащего x, примерно равна $i_N = k$. ▶

Замечание. В доказательстве мы выбрасывали слова из всех плохих множеств, а не только множеств из $\mathscr{A}$. Поэтому верхняя оценка множества $P_x^{\mathscr{A}}$ верна и для семейства, состоящего из всех конечных множеств. Иными словами, мы добились, что слово x имеет заданное множество $P_x^{\mathscr{A}}$ не только для исходного $\mathscr{A}$, но и для всех семейств конечных множеств, включающих $\mathscr{A}$ (не обязательно удовлетворяющим трём условиям на с. 484).

359 Проведите это рассуждение подробно.

360 (**а**) Пусть x — слово длины n, а r — натуральное число, не превосходящее $n/2$. Обозначим через $KS_r(x)$ наименьшую (простую) сложность слова y той же длины, что x, отличающегося от x не более чем в r позициях. Докажите, что $KS_r(x)$ с точностью до $O(\log n)$ равно наименьшему i, для которого x имеет $(i * \log V(r))$-описание, являющееся шаром Хемминга, где $V(r)$ обозначает мощность шара Хемминга радиуса r.

(**б**) Опишите все возможные формы функции $r \mapsto KS_r(x)$ с точностью до слагаемого $O(\sqrt{n \log n})$. [Указание. Для всех слов x длины n выполнено $KS_0(x) = KS(x)$, $KS_n(x) = O(\log n)$ и

$$0 \leqslant KS_a(x) - KS_b(x) \leqslant \log(V(b)/V(a)) + O(\log n)$$

для всех $a < b \leqslant n/2$. Обратно, для любых чисел $k \leqslant n$ и для любой функции $t\colon \{0, 1, \ldots, n\} \to \mathbb{N}$ сложности m, удовлетворяющей условиям $t(0) = k$, $t(n) = 0$ и $0 \leqslant t(a) - t(b) \leqslant \log(V(b)/\log V(a))$ для всех $a < b \leqslant n/2$, существует слово x длины n и сложности $k + O(\sqrt{n \log n}) + O(m)$, для которого $KS_a(x) = t(a) + O(\sqrt{n \log n}) + O(m)$.]

Сложность $KS_r(x)$ была впервые определена в работе [54]. В работе [42] изучены два варианта определения условной сложности с ошибками: равномерный и неравномерный варианты. Равномерная условная сложность с ошибками $KS^u_{rs}(x \mid y)$ определяется, как наименьшая сложность программы, которая по любому слову y' на расстоянии не более s от y выдаёт некоторое слово x' на расстоянии не более r от x (слово x' может зависеть от y'). Неравномерная сложность с ошибками $KS_{rs}(x \mid y)$ определяется как $\max_{y'} \min_{x'} KS(x' \mid y')$, если рассматривать x' не дальше r от x, а y' не дальше s от y. В определении неравномерной сложности программа переработки y' в x' может зависеть от y', а в определении равномерной сложности — не может. Поэтому неравномерная сложность меньше равномерной и может быть значительно меньше неё (последнего мы доказывать не будем, отсылая читателя к [42]).

Теорема 254 давала критерий того, что данное слово имеет $(i * j)$-описание (без ограничения класса описаний). Неясно, имеет ли она аналог для произвольного семейства множеств $\mathscr{A}$. Но утверждение следующей за ней теоремы 255 переносится почти без изменений на случай произвольного перечислимого семейства множеств.

Теорема 258. *Пусть семейство $\mathscr{A}$ перечислимо. Пусть слово x длины n имеет не менее 2^k множеств из $\mathscr{A}$, являющихся его $(i * j)$-описаниями. Тогда оно имеет $((i - k) * j)$-описание в $\mathscr{A}$ (и, следовательно, $(i * (j - k))$-описание, если семейство $\mathscr{A}$ удовлетворяет условию (3) на с. 484).*

В формулировке этой теоремы мы, как и раньше, опускаем слагаемые вида $O(\log(n + i + j + k))$, которые разрешается прибавлять к параметрам описаний.

◄ Пусть даны i, j, n, k. Будем перечислять все $(i * j)$-описания из $\mathscr{A}$, то есть множества мощности не более 2^j и сложности не более i, отбирая некоторые из них. При отборе мы стремимся обеспечить следующее: в любой момент каждое слово x длины n, которое имеет не менее 2^k описаний (среди уже обнаруженных), принадлежит хотя бы одному из отобранных множеств, а общее количество отобранных множеств не превосходит $2^{i-k} p(n, k, i, j)$, где p — некоторый полином. Нам достаточно доказать, что для некоторого полинома p существует стратегия отбора, гарантирующая выполнение этих двух требований. (Поскольку стратегия может быть найдена перебором по данным n, i, j, k, ее колмогоровская сложность ограничена величиной $O(\log(n + k + i + j))$.)

Существование такой стратегии можно доказать конструктивно и с помощью вероятностного рассуждения.

Вероятностное доказательство. Рассмотрим игру двух противников, в которой игроки ходят по очереди и каждый делает 2^i ходов. Первый игрок, совершая ход, указывает некоторое множество слов длины n, а второй на следующем за этим ходу сообщает, отбирает он это множество или нет. Второй игрок проигрывает, если после некоторого его хода количество отобранных множеств превысит

$$2^{i-k+1}(n+1)\ln 2$$

или найдется слово x длины n, принадлежащее не менее 2^k множествам первого игрока, но не принадлежащее ни одному из отобранных множеств.

В этой игре один из игроков имеет выигрышную стратегию. Мы утверждаем, что этим игроком является второй. Рассуждая от противного, допустим, что выигрышную стратегию имеет первый. Фиксируем такую стратегию и предполагаем, что первый игрок её придерживается.

Рассмотрим следующую вероятностную стратегию для второго игрока: после каждого хода A первого игрока отбираем множество A с вероятностью

$$p = 2^{-k}(n+1)\ln 2.$$

Чтобы получить противоречие, нам достаточно доказать, что эта стратегия выигрывает с положительной вероятностью.

Среднее количество отобранных множеств равно

$$p2^i = 2^{i-k}(n+1)\ln 2.$$

По неравенству Чебышёва вероятность того, что количество отобранных множеств вдвое превысит свое среднее значение, меньше $1/2$. Поэтому достаточно доказать, что второе условие выигрыша первого игрока (после некоторого хода найдётся x, принадлежащее не менее чем 2^k множествам первого игрока, но не принадлежащее ни одному из отобранных множеств) выполнено с вероятностью не более $1/2$.

Нам достаточно доказать, что для любого фиксированного x это условие выполняется с вероятностью меньше 2^{-n-1}. Для этого покажем индукцией по t, что вероятность события «после некоторого хода второго игрока слово x принадлежит не менее чем t множествам первого игрока, но не принадлежит ни одному из отобранных множеств» не превосходит $(1-p)^t$. Обозначим это событие через R_t. Для $t = 0$ утверждение очевидно. Чтобы сделать индуктивный переход, нам достаточно показать, что вероятность R_{t+1} при условии R_t не больше $1-p$.

Пусть $z = (z_1, z_2, \ldots, z_s)$ — произвольная последовательность первых s ходов второго игрока ($z_i = 1$, если i-е множество первого игрока отбирается, и $z_i = 0$ в противном случае). Будем называть последовательность z неудачной, если после s-го хода второго игрока слово x впервые оказалось покрыто t множествами первого игрока и не покрыто ни одним из отобранных множеств (из этого следует, что s-е множество первого игрока содержит x и $z_s = 0$). Неудачные последовательности попарно не согласованы и каждая последовательность ходов второго игрока, принадлежащая событию R_t начинается с некоторой неудачной последовательности. Поэтому события «стратегия второго игрока сделала ходы $z = (z_1, z_2, \ldots, z_s)$»

для неудачных z образуют разбиение события R_t. Значит нам достаточно доказать, что для любой неудачной последовательности z вероятность события R_{t+1} при условии, что второй игрок сделал ходы z, не больше $1-p$. Эта условная вероятность есть вероятность того, что выигрышная стратегия первого игрока в ответ на ходы $z_1,\ldots,z_s$ и случайные последующие ходы второго игрока в некоторый момент выдаст множество, содержащее x, а второй игрок не отберёт его. Поскольку решение не отбирать это множество принимается с вероятностью $1-p$ (независимо от предыстории игры), оцениваемая вероятность равна произведению $1-p$ и вероятности того, что первый игрок в ответ на ходы $z_1,\ldots,z_s$ и случайные последующие ходы второго игрока в некоторый момент выдаст множество, содержащее x. Следовательно, она не превосходит $1-p$.

Осталось заметить, что p было выбрано таким образом, чтобы при $t=2^k$ величина $(1-p)^t$ была меньше 2^{-n-1}.

Конструктивное доказательство. Рассмотрим ту же самую игру, только верхнюю границу количества отобранных множеств

$$2^{i-k+1}(n+1)\ln 2$$

заменим на

$$2^{i-k}i^2n\ln 2$$

и разрешим второму игроку на каждом ходу отбирать несколько множеств (из множеств, указанных первым игроком на предыдущих ходах). Укажем явно выигрышную стратегию второго игрока в этой игре. Она состоит из независимого применения i стратегий, занумерованных числами $1,2,\ldots,i$.

Стратегия номер s просыпается через каждые 2^s ходов первого игрока, а именно, после любого хода, номер которого кратен 2^s. Проснувшись, она формирует семейство S, состоящее из последних 2^s множеств первого игрока, и множество T, состоящее из всех слов, которые покрыты не менее $2^k/i$ множествами из S. Затем она отбирает некоторые множества из S таким образом, чтобы все слова из T оказались покрыты отобранными множествами. Применяя жадный алгоритм (берем множество из S, покрывающее наибольшее количество слов из T, затем множество, которое покрывает наибольшую часть остатка и т. д.), мы можем сделать это, отобрав не более

$$in2^{s-k}\ln 2$$

множеств. Действительно, каждое слово из T покрыто не менее, чем $2^k/i$ множествами из S. Поэтому некоторое множество из S покрывает не менее $|T|2^{k-s}/i$ слов из T. Значит, и жадный алгоритм выберет множество, покрывающее не менее этого количества слов. Аналогичные рассуждения применимы и к остатку T: второе множество, выбранное жадным алгоритмом, покрывает не менее $2^{k-s}/i$-й доли остатка T. Поэтому количество элементов T, оставшихся не покрытыми после выбора $in2^{s-k}\ln 2$ множеств, не превосходит

$$|T|(1-2^{k-s}/i)^{in2^{s-k}\ln 2} < 2^n e^{(-2^{k-s}/i)in2^{s-k}\ln 2} = 1,$$

то есть непокрытых элементов не останется. (Можно было бы применить и вероятностное рассуждение, показав, что случайно выбранные $in2^{s-k}\ln 2$ множеств из S

с положительной вероятностью покроют T. Но в этом случае стратегия стала бы менее явной.)

Таким образом, общее количество множеств, отобранных стратегией с номером s, не превосходит

$$in2^{s-k}(\ln 2)2^{i-s} = in2^{i-k}\ln 2,$$

что дает в целом не более $i^2n2^{i-k}\ln 2$ отобранных множеств.

Осталось доказать, что после каждого нашего хода любое слово, принадлежащее не менее чем 2^k множествам первого игрока, покрыто некоторым отобранным множеством. Рассмотрим ход номер t и любое непокрытое после этого хода слово x. Разложим t в двоичную запись: $t = 2^{s_1} + 2^{s_2} + \ldots$ (где $s_1 > s_2 > \ldots$). Поскольку x не принадлежит множествам, отобранным стратегиями с номерами $s_1, s_2, \ldots$, кратность x среди первых 2^{s_1} множеств меньше $2^k/i$, кратность x среди следующих 2^{s_2} множеств также меньше $2^k/i$ и т.д. Таким образом, после хода t слово x покрыто менее чем $i \cdot 2^k/i = 2^k$ множествами первого игрока. ►

Как и в случае неограниченного класса гипотез, из этой теоремы следует теорема об упрощении гипотез, содержащих «лишнюю» информацию.

Теорема 259. *Пусть семейство $\mathscr{A}$, из которого выбираются гипотезы, перечислимо. Если слово x имеет $(i*j)$-описание $A \in \mathscr{A}$, для которого $KS(A\,|\,x) \geqslant k$, то x имеет и $((i-k)*j)$-описание из $\mathscr{A}$ (и, следовательно, $(i*(j-k))$-описание, если семейство $\mathscr{A}$ удовлетворяет условию* (3) *на с.* 484).

(Как и раньше, мы опускаем логарифмические слагаемые, необходимые в точной формулировке.)

14.5. Дефект оптимальности и дефект случайности

Мы рассматривали две характеристики конечного множества A, содержащего слово x: (а) дефект случайности $d(x\,|\,A) = \log|A| - KS(x\,|\,A)$ и (б) разность $\log|A| + KS(A) - KS(x)$, которая показывает, насколько длина двухчастного описания слова x с помощью множества A отличается от минимально возможной. Эту разность мы будем называть *дефектом оптимальности* гипотезы A о слове x и обозначать $\delta(x\,|\,A)$.

Как связаны две эти характеристики? Начнём с очевидного наблюдения:

Теорема 260. *Дефект случайности слова x относительно конечного множества A, содержащего это слово, не превосходит дефекта оптимальности* (*с логарифмической точностью*):

$$d(x\,|\,A) \leqslant \delta(x\,|\,A) + O(\log l(x)).$$

◄ Нам нужно доказать, что

$$\log|A| - KS(x\,|\,A) \leqslant \log|A| + KS(A) - KS(x) + O(\log l(x)).$$

После сокращения $\log|A|$ мы получаем неравенство

$$KS(x) \leqslant KS(A) + KS(x\,|\,A) + O(\log l(x));$$

правая часть его (с логарифмической точностью) есть сложность пары $\langle x, A\rangle$ и потому не меньше $KS(x)$. ►

Как видно из этого рассуждения, разница между дефектом случайности и дефектом оптимальности равна $KS(x, A) - KS(x)$, то есть $KS(A\,|\,x)$ с точностью $O(\log l(x) + \log KS(A))$, что есть $O(\log l(x))$, если $KS(A) = O(KS(x))$. (Более сложных гипотез мы рассматривать не будем: какой смысл в объяснении, которое значительно сложнее объясняемого объекта?)

Легко привести пример гипотезы, для которой дефект оптимальности значительно больше дефекта случайности. Например, для случайного слова x длины n рассмотрим гипотезу B, состоящую из всех слов длины n, и добавим в B случайное слово y длины $n-1$, независимое от x. Тогда $KS(B\,|\,x)$ примерно равно n, и дефект оптимальности превосходит дефект случайности (который по-прежнему близок к нулю, несмотря на добавление слова y) примерно на n. Видно, что в этом случае гипотеза «плоха» с интуитивной точки зрения, поскольку содержит явно бессмысленный элемент y, никакого отношения к наблюдаемому x не имеющий; удаление y позволило бы улучшить гипотезу, приблизив дефект оптимальности к дефекту случайности (который практически не меняется).

Доказанная нами теорема 256 говорит, что нечто подобное можно сделать всегда: если для данной гипотезы B, объясняющей слово x, дефект оптимальности $\delta(x\,|\,B)$ больше дефекта случайности $d(x\,|\,B)$ (на величину $KS(B\,|\,x)$, как мы знаем), то можно найти другую гипотезу A (не большей сложности, чем B), дефект оптимальности $\delta(x\,|\,A)$ которой не превосходит $d(x\,|\,B)$.

Поэтому вопрос о том, можно ли для данного x найти гипотезу A с $KS(A) \leqslant \alpha$ и $d(x\,|\,A) \leqslant \beta$, который мы задавали в определении (α, β)-стохастичности, равносилен (с логарифмической точностью) вопросу о том, можно ли найти гипотезу A с $KS(A) \leqslant \alpha$ и $\delta(x\,|\,A) \leqslant \beta$.

361 Докажите, что (для данного слова x) устройство множества P_x (см. теорему 253) полностью определяет, при каких α и β слово x является (α, β)-стохастическим: это будет, если точка $(\alpha, KS(x) - \alpha + \beta)$ принадлежит P_x или $\alpha > KS(x)$ (как всегда, с логарифмической точностью).

362 Выведите отсюда обещанное выше (с. 474) утверждение о том, что если $\alpha + \beta < n - O(\log n)$, то существуют слова длины n, не являющиеся (α, β)-стохастическими. [Указание. Таковым будет первое (в алфавитном порядке) слово длины n, не имеющее $\alpha * (n - \alpha)$ описаний (если быть точным, то из параметров надо вычесть величины порядка $O(\log n)$). Его сложность примерно равна α, и можно сослаться на критерий из предыдущей задачи.]

363 Докажите, что среди слов длины n доля не-(α, β)-стохастических слов не меньше $2^{-\alpha-\beta-O(\log n)}$. [Указание. Таковыми будут первые $2^{n-\alpha-\beta}$ слова длины n (в алфавитном порядке), не имеющие $\alpha * (n - \alpha)$-описаний (мы опускаем логарифмические слагаемые). Сложность любого из этих слов не меньше α и не больше $\alpha + (n - \alpha - \beta) = n - \beta$. Из второго следует, что для каждого из этих слов x точка $(\alpha, KS(x) - \alpha + \beta)$ не принадлежит P_x.]

364 Докажите, что первое неравенство из условия теоремы 251 можно заменить на более слабое неравенство $\alpha + \beta < n - O(\log n)$. [Указание. Доказательство верхней оценки почти не изменяется: априорная вероятность слова, существующего по задаче 362, не меньше $2^{-\alpha}$. Доказательство нижней оценки в теореме 250 использовало только неравенство $\alpha < \beta - O(\log n)$.]

365 Для данного слова x рассмотрим множество Q_x, состоящее из пар $\langle\alpha, \beta\rangle$, для которых слово x является (α, β)-стохастическим. Докажите, что граница множества Q_x является (с логарифмической точностью) невозрастающей кривой, ведущей из точки некоторой точки $\langle 0, i\rangle$, где $i \leqslant l(x) - KS(x)$, в точку $\langle KS(x), 0\rangle$, и что любая простая невозрастающая кривая может являться такой границей.

366 Докажите, что если для данного слова x и данного α существует гипотеза, доставляющая минимум дефекта случайности среди гипотез сложности не выше α, но её дефект оптимальности превосходит её дефект случайности на γ, то граница множества P_x на участке от $\alpha - \gamma$ до α (считая по оси абсцисс) идёт под углом 45°.

[Указание. Воспользуйтесь более сильным утверждением теоремы 256.]

367 Пусть $\mathscr{A}$ — любое семейство конечных множеств слов, удовлетворяющее условиям (1) – (3) на с. 484. Докажите, что для любого x и любого $\alpha \leqslant KS(x)$ утверждения «существует множество $A \in \mathscr{A}$ сложности не более α с $d(x\,|\,A) \leqslant \beta$», «существует множество $A \in \mathscr{A}$ сложности не более α с $\delta(x\,|\,A) \leqslant \beta$» и «точка $\langle\alpha, KS(x) - \alpha + \beta\rangle$ принадлежит $P_x^{\mathscr{A}}$» эквивалентны (с логарифмической точностью).

368 Пусть $\mathscr{A}$ — произвольное семейство конечных множеств, перечисляемое программой p. Докажите, что для любого x длины n утверждения «существует множество $A \in \mathscr{A}$ с $d(x\,|\,A) \leqslant \beta$» и «существует множество $A \in \mathscr{A}$ с $\delta(x\,|\,A) \leqslant \beta$» эквивалентны (с точностью $O(l(p) + \log KS(A) + \log n + \log\log|A|)$).

14.6. Минимальные гипотезы

Пусть фиксировано некоторое слово x. Ему, как мы видели, соответствует множество P_x пар (α, β), для которых x имеет $(\alpha * \beta)$-описание. Такие описания, как мы говорили, естественно рассматривать как «статистические гипотезы», объясняющие происхождение слова x. Интересно понять, как именно устроены эти гипотезы. Оказывается, что можно более или менее явно описать некоторый класс гипотез, к которым в некотором смысле сводится любая гипотеза. Этот класс проистекает из доказательства теоремы 254.

Пусть l — произвольное число, большее $KS(x)$. Тогда перечень всех слов сложности не более l содержит x. Фиксируем такой перечень (простой алгоритм, порождающий все эти слова по одному разу). Пусть N_l — число слов в перечне. Запишем N_l в двоичной системе счисления, то есть представим его в виде суммы убывающих степеней двойки:

$$N_l = 2^{s_1} + 2^{s_2} + \ldots + 2^{s_t}, \text{ где } s_1 > s_2 > \ldots > s_t.$$

В соответствии с этим разложением и перечень можно разбить на группы в порядке порождения элементов: сначала идут 2^{s_1} элементов, затем 2^{s_2} элементов и так далее. Слово x попадает в одну из частей. Эту часть (соответствующее конечное множество) мы и будем рассматривать как описание слова x. Таким образом, мы получаем некоторое семейство описаний (при каждом $l > KS(x)$ — своё описание).

Следующие две теоремы устанавливают обещанные свойства этих описаний (во-первых, они являются минимальными, то есть лежат на границе множества P_x; во-вторых, всякое описание слова x в некотором смысле сводится к одному из них).

Теорема 261. *Пусть в описанной ситуации слово x попало в часть размера 2^s. Тогда эта часть является $((l-s)*s)$-описанием и точка $(l-s, s)$ лежит на границе множества P_x (точнее говоря, она является $((l-s+O(\log l))*s)$-описанием и соответствующая точка лежит в $O(\log l)$-окрестности границы множества P_x).*

◀ Чтобы задать часть, достаточно знать её размер, а также сколько элементов имеется перед ней, то есть достаточно знать s, l и все биты числа N_l, кроме s последних (то есть $l-s$ битов). Ещё нужно знать сам перечень, но он по предположению имеет логарифмическую сложность. Поэтому сложность рассматриваемой части есть $l-s+O(\log l)$, а число элементов в ней равно 2^s, что и требовалось.

Если бы точка $(l-s, s)$ не лежала бы на границе множества P_x, а входила бы в P_x вместе с некоторой окрестностью более чем логарифмического размера, то слово x имело бы заметно лучшие двухчастные описания, чем построенное нами (с той же или даже немного меньшей суммарной длиной и с бóльшим размером множества) и по теореме 254 (пункт (г)) слово x появлялось бы в перечне более чем за 2^s элементов до конца (что противоречит построению части). ▶

Следующая теорема показывает, что описаний рассмотренного вида в некотором смысле достаточно. Пусть даны произвольное слово x и произвольное конечное множество A, его содержащее. Пусть максимальная сложность слов из A равна l. Как и раньше, разобьём слова сложности не выше l (их число обозначим N_l) на части, размеры которых есть степени двойки, соответствующие единичным битам в двоичной записи N_l. Пусть B — часть, в которую попало слово x, и пусть она имеет размер 2^s.

Теорема 262. *Гипотеза B (как объяснение для слова x) не хуже гипотезы A с точки зрения зрения сложности и дефекта оптимальности*:

(а) $KS(B) \leqslant KS(A)$;

(б) $\delta(x\,|\,B) \leqslant \delta(x\,|\,A)$ *(с точностью $O(\log l)$). Кроме того,*

(в) $KS(B\,|\,A) = O(\log l)$ *(гипотеза B проста относительно A).*

◀ Зная A и значение l, мы можем дождаться, когда в перечне всех слов сложности не больше l появятся все слова из множества A. К этому моменту появится и слово x (входящее в часть размера 2^s), так что останутся необнаруженными не более $O(2^s)$ слов (из этой и следующих частей). Таким образом, мы знаем N_l с ошибкой не более чем $O(2^s)$ и тем самым знаем его первые $l-s$ его битов с ошибкой $O(1)$. А по этой информации (а также l и s) мы можем получить B. Таким образом, $KS(B\,|\,A) = O(\log l)$. Утверждение (в) (а тем самым и (а)) доказано.

Утверждение (б) очевидно по построению: если $KS(A) = \alpha$ и $\log|A| = \beta$, то все слова множества A имеют $(\alpha * \beta)$-описание и сложность не более $\alpha + \beta + O(\log\alpha)$, так что и максимальная сложность l не превосходит $\alpha + \beta + O(\log\alpha)$. Построенное двухчастное описание, как мы видели в предыдущей теореме, является $((l - s) * s)$-описанием, так что его суммарная длина (и потому дефект оптимальности) не больше, чем у A. ▶

Связь между параметрами гипотез A и B в этой теореме показана на рис. 14.3: точка изображает параметры гипотезы A, а серым цветом показана область, где могут находиться параметры гипотезы B.

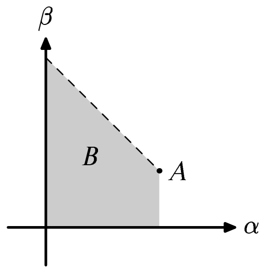


Рис. 14.3. Параметры гипотезы A и её «упрощения» B.

Что будет, если начальная гипотеза A сама уже была минимальной (близкой к границе множества P_x)? Можно ли утверждать, что B в этом случае имеет те же параметры, что и A? Вообще говоря, нет: возможно, что гипотеза B находится на пунктирной границе серой области (рис. 14.3). (Внутри серой области она находиться не может, поскольку в этом случае гипотеза A была бы, как легко проверить, внутренней для P_x.)

Можно сказать ещё и так: если представлять себе границу множества P_x как кривую, идущую поочерёдно вправо-вниз (под углом 45°) и вертикально вниз, то все точки поворота, в которых вертикальное движение сменяется наклонным, имеют соответствующие им гипотезы нашего семейства (поскольку для таких A серая область, в которой может находиться B, пересекается с P_x в единственной точке).

Заметим, что информация, содержащаяся в гипотезах построенного нами семейства, как это ни странно, почти не зависит от слова x: гипотеза B содержит ту же информацию, что и начало слова N_l длины $l - s$. Как мы видели в задаче 354 (с. 483), это начало можно заменить на слово N_{l-s} или (теорема 116, с. 195) на начальный отрезок числа Чейтина Ω. Тем самым наши гипотезы (по мере движения по границе P_x слева направо) содержат всё больше и больше битов числа Ω. Выходит, что гипотезы семейства содержат всё больше и больше информации не только об x (что хорошо), но и о числе Ω (которое от x не зависит).

Возможно, кроме логарифма размера и сложности, у гипотез имеются и другие неизученные нами параметры, с помощью которых можно разделить гипотезы одинакового размера и сложности на более и менее подходящие для объяснения x. В качестве такого параметра в работе [172] предложено рассмотреть тоталь-

ную сложность гипотезы A при известном x. Во всех примерах, рассмотренных нами, интуитивно правильные гипотезы для объяснения x имеют малую тотальную сложность относительно x. С другой стороны, можно показать, что гипотезы «универсального» из теоремы 261 имеют большую тотальную условную сложность относительно некоторых своих элементов (хотя и маленькую обычную условную сложность). Мы не будем доказывать этого утверждения, отсылая читателя к [172].

Ещё интересно отметить, что это наблюдение (о том, что одна гипотеза содержит часть информации из другой) относится лишь к построенным нами гипотезам, а не к произвольным гипотезам на границе множества P_x. Это показывает следующий пример. Пусть x — случайное слово длины n. Рассмотрим две следующих гипотезы: множество слов длины n, имеющих ту же первую половину, что у x, и множество слов длины n, имеющих ту же вторую половину, что у x. Обе гипотезы имеют совсем небольшой дефект оптимальности, однако содержат разную информацию. (Это не противоречит доказанной теореме, поскольку в качестве гипотезы B, которая лучше обеих, может фигурировать множество всех слов длины n.)

14.7. Немного философии

Задачу поиска двухчастного описания для данного слова x можно разукрашивать по-разному. Можно говорить о науке, которая наблюдает внешний мир в лице слова x и строит о нём разные гипотезы, пытаясь отобрать «наиболее правильную». При этом критерием отбора является простота гипотезы (измеряемая колмогоровской сложностью множества A; чем эта сложность меньше, тем гипотеза проще, то есть лучше), а также её «конкретность», или «объясняющая способность» (которая измеряется размером множества A; чем оно меньше, тем гипотеза конкретнее, то есть лучше). Под это дело можно даже подвести философскую базу в лице классика философии Оккама и его бритвы, согласно которой не следует умножать сущности без надобности. Ещё можно стремиться к тому, чтобы дефект случайности наблюдаемого слова x относительно предложенной гипотезы A был мал («не оставалось необъяснённых закономерностей»).

Можно изложить и более прагматичную версию. Колмогоровскую сложность можно рассматривать как предельный вариант науки о сжатии файлов: сложность слова x есть нижняя граница для всех способов его сжатия без потери информации. Программа-компрессор тем лучше, чем ближе она к этой границе подходит (для той или иной практически важной категории файлов).

Это относится к сжатию без потерь информации. Однако всё большее применение находят способы сжатия, где некоторая «несущественная» часть информации теряется и за счёт этого достигается большее сжатие.

Пусть, скажем, есть старая граммофонная пластинка, на которой от долгого хранения в пыльном чулане появились царапины (в случайных местах дорожки, содержащей запись). Царапины соответствуют различным паразитным пикам на осциллограмме сигнала (графике зависимости звукового давления от времени). Таким образом, изначальная информация подверглась случайному искажению. Сложность

записи при этом сильно возросла, если царапин много. Но если мы хотим передать лишь общее впечатление от проигрывания этой пластинки, то нам неважно, где именно находятся царапины, а важен лишь общий характер дефектов.

Другими словами, фактическая запись на пластинке является одним из элементов большого множества записей с «такого же типа царапинами». Тем самым получается двухчастное описание: сначала мы задаём это множество (скажем, указав начальную неискажённую запись и статистические характеристики наложенного шума), а затем указываем конкретный элемент этого множества (где именно царапины). Если при сжатии мы утратим вторую часть описания, сохранив лишь первую, то после восстановления мы получим запись, где на тот же исходный сигнал будет наложен шум с теми же статистическими характеристиками, но уже другой. Можно ожидать, что слушатель не заметит этой подмены. А если удастся шум вообще не накладывать, тем самым «очистив» запись, это ещё лучше (если нас интересует музыка, а не ностальгическое очарование старинного граммофона).

Пользуясь этой аналогией, можно следующим образом интерпретировать утверждение задачи 368. Пусть слово x было получено из неизвестного слова y той же длины наложением шума, то есть для некоторого известного нам r слово x было выбрано случайно в шаре Хемминга радиуса r с центром y. Желая удалить шум, мы ищем шар Хемминга радиуса r, дающий минимально короткое двухчастное описание для x (то есть шар минимальной сложности). Допустим нам это удалось, и мы в самом деле нашли шар минимально возможной сложности. С большой вероятностью дефект случайности x в исходном шаре мал, а значит, по задаче 368 (примененной к семейству, состоящему из множества шаров Хемминга радиуса r) мал будет и дефект оптимальности x в найденном нами шаре. Это означает, что в найденном двухчастном описании x вторая часть не содержит полезной информации. Иными словами, центр найденного шара является очищенной от шума версией x (в частности, и от того шума, который уже был в y).

Другой пример: в фотографии, скажем, песчаной гряды, сделанной с большим разрешением, зафиксированы положения отдельных песчинок, которые случайны, и потому сложность этой фотографии велика. Однако для зрителя она является всего лишь «типичным элементом» множества фотографий, где песчаная гряда находится на том же месте и сложена из такого же песка, хотя конкретные песчинки могут быть в разных местах. Если при сжатии сохранится только информация об этом множестве, а при восстановлении будет построен другой типичный его элемент, то разница будет незаметна.

Надо иметь в виду, что эта аналогия — всего лишь аналогия, причём довольно далёкая, и вряд ли математические результаты о колмогоровской сложности двухчастных описаний могут найти непосредственные «практические применения». (Хотя бы потому, что мы полностью игнорируем вычислительную сложность алгоритмов декодирования, а алгоритмы кодирования вообще не рассматриваем. Может быть, именно этим объясняется парадоксальная независимость оптимальных гипотез от слова x, которую мы отмечали выше.)

Приложение 1. Колмогоровская сложность и основания теории вероятностей

В этом разделе мы не формулируем и не доказываем никаких математических утверждений, а обсуждаем основания теории вероятностей, в том числе и роль алгоритмической теории информации, следуя [156].

Парадокс теории вероятностей

Часто говорят, что в естественных науках учёный выдвигает гипотезы, которые затем поверяет экспериментом: если эксперимент не согласуется с предсказаниями, полученными на основе гипотезы, то нужно её отбросить или изменить.

Можно ли эту схему применить к теории вероятностей? Предположим, что мы высказали гипотезу «честной монеты» (бросания независимы и орлы и решки равновероятны), а затем тысячу раз бросали монету, чтобы эту гипотезу проверить. Получилась последовательность орлов и решек (нулей и единиц) длины 1000. Но как можно о чём-то судить на основании этой последовательности, если гипотеза как раз и утверждает, что все 2^{1000} последовательностей длины 1000 равновероятны? Разве могут в этом случае одни из них опровергать гипотезу, а другие — подтверждать? Или все они равновероятны, но некоторые менее равновероятны (как мог бы сказать Оруэлл)?

Вот другая ситуация, где возникает тот же парадокс. Предположим (говорят, так и было в XIX веке), что колоды карт продаются запечатанными и предварительно перетасованными, чтобы можно было распечатать колоду и начать игру. Идея хорошая (если вы не шулер), но для её реализации нужно тасовать колоды перед отправкой их потребителям. Как и любой технологический процесс, тасование карт нуждается в контроле: ОТК должен просматривать колоды, и хорошо перетасованные пропускать, а плохо перетасованные отправлять на дотасовку. Но что такое «плохо» и «хорошо» перетасованные? Ведь идея тасования в том и состоит, чтобы сделать все $n!$ расположений карт равновозможными!

Более современный вариант того же парадокса. Составители теста для школьников, где в 20 вопросах надо отгадать один из двух правильных ответов[1] «а» или «б», хотят изготовить для каждого школьника индивидуальный вариант, случайно переставляя ответы. В одном из вариантов случайно оказалось, что правильны все двадцать ответов «а». Надо ли забраковать этот вариант?

[1] Мы оставляем в стороне вопрос об осмысленности таких тестов.

Сложившаяся практика

Что бы там ни говорили философы, статистики — люди практические и им работать надо. Попробуем описать, как такая работа происходит [167, 151, 156].

А. Применение статистических гипотез. Прежде всего надо признать, что теория вероятностей не *делает предсказаний*, а *даёт рекомендации*. Рекомендации эти основаны на следующем: *если рассчитанная* (*на основе принятой гипотезы*) *вероятность события A много меньше, чем вероятность события B, то возможность осуществления события B надо рассматривать в первую очередь и в гораздо большей степени, чем возможность осуществления события A*. (Мы предполагаем, что последствия двух этих событий одинаково серьёзны.) В частности, если вероятность A меньше вероятности попасть под метеорит, то скорее всего A можно игнорировать полностью (поскольку B так и так приходится игнорировать, иначе жить нельзя).

Борель [12] формулирует этот принцип так: «...В Париже меньше миллиона взрослых людей [это сейчас уже не так]; газеты ежедневно извещают о странных случаях или несчастьях, постигших одного из них; жизнь была бы невозможной, если бы каждый постоянно опасался за себя в отношении всех приключений, о которых можно прочесть в отделе разных происшествий; а это равносильно тому, чтобы сказать, что практически нужно пренебрегать вероятностями, которые меньше одной миллионной... Часто боязнь дурного приводит к ещё худшему. Чтобы уметь различить худшее, надо хорошо знать вероятности различных явлений» (с. 159 – 160).

Б. Отбор статистических гипотез. Здесь нельзя наивно утверждать, что если наблюдаемое событие имеет пренебрежимо малую вероятность согласно рассматриваемой гипотезе, то её нужно отбросить. В самом деле, это бы означало, что любой результат тысячи бросаний монеты опровергает её честность, поскольку эта конкретная последовательность орлов и решек имеет вероятность 2^{-1000}.

Именно в этом месте вступает в игру колмогоровская сложность. Мы отвергаем гипотезу, наблюдая *простое* событие, которому эта гипотеза приписывает *пренебрежимо малую* вероятность. Поэтому появление тысячи орлов подряд (ну, или чередование пяти сотен орлов с пятью сотнями решек) воспринимается нами как противоречащее честности монеты.

В этой формулировке важны обе части: «первым выпал орёл» — простое событие, но вероятность его достаточно велика, поэтому оно не противоречит честности монеты. С другой стороны, вероятность появления любой последовательности из тысячи битов пренебрежимо мала, но если эта последовательность не проста, это тоже не противоречит честности монеты.

Иногда правило отбора статистических гипотез и правило их применения объединяют в одной формулировке: «события с малой вероятностью достоверно не происходят». Вот как пишет про это Борель: «Не следует бояться применить слово *достоверность* для обозначения вероятности, которая отличается от единицы на достаточно малую величину» ([13], с. 7). Иногда эту формулировку называют *принципом Курно*. Но мы предпочитаем разделять эти две стадии, поскольку при

проверке гипотезы простота события играет первостепенную роль, а при её применении это уже не важно. (Впрочем, можно ожидать, что реально интересующие нас события будут скорее всего простыми.)

Простые или заранее указанные?

К сожалению, данное нами описание никак не назовёшь точным. И дело не только в том, что границу, начиная с которой надо считать вероятность пренебрежимо малой, трудно выбрать обоснованно, но и в том, что колмогоровская сложность, подразумеваемая в словах «простое событие», определена с точностью до константы и, строго говоря, не имеет смысла применительно к описанию одного события. С другой стороны, трудно описать сложившуюся практику в каких-то более точных терминах.

Один из возможных путей таков: можно сказать, что мы отбрасываем гипотезу, если мы наблюдаем весьма маловероятное событие, *указанное заранее* (до начала эксперимента). Но тогда возникает такой вопрос. Представим себе, что мы бросали монету и получили последовательность, которая на первый взгляд показалась нам случайной. Пришедший коллега, однако, обратил наше внимание на то, что если смотреть на биты через два на третий, то нули и единицы чередуются (??0??1??0??1 . . .). Будем ли мы после этого по-прежнему считать последовательность случайной? Видимо нет: хотя до эксперимента такая закономерность нам в голову не приходила, но она настолько проста, что она *могла бы придти*. Можно даже так сказать: рассмотрим объединение всех простых событий ничтожной вероятности (выбрав границу между простым и сложным так, чтобы это объединение оставалось очень маловероятным). Теперь мы можем утверждать, что это событие рассматривалось заранее (по крайней мере теми, кто прочёл этот раздел книги).

Другое соображение: если последовательность результатов испытаний в точности соответствует последовательности результатов вчерашнего эксперимента, или какой-то другой ранее полученной последовательности, то мы отвергаем гипотезу честной монеты, даже если эта ранее полученная последовательность сложна. Можно сказать, впрочем, что в этом случае полная последовательность наблюдений (как вчерашних, так и сегодняшних) попадает в простое маловероятное событие (повторение). Однако для вчерашних наблюдений у нас могло не быть никакой статистической гипотезы, так что лучше сказать, что наше сегодняшнее наблюдение имеет малую условную сложность относительно информации, имевшейся у нас до эксперимента.

Короче говоря, мы можем заменить одно расплывчатое требование («быть простым») на другое («быть известным заранее»). Вот как описывает ситуацию Борель ([12], с. 77–78):

> «Скажем несколько слов о мысли Бертрана по поводу равностороннего треугольника, который могли бы образовать три звезды; она связана с вопросом о круглом числе. Если рассматривать число, взятое наудачу между 1 000 000 и 2 000 000, то вероятность, что оно равняется 1342517, равна одной миллионной; вероятность, что оно равняется

1500000, равна тоже одной миллионной. Между тем эту последнюю возможность охотно будут считать менее вероятной, чем первую; это зависит от того, что никто и никогда не представляет себе индивидуально такое число, как 1542317; оно рассматривается как тип чисел того же вида, и если при его переписке изменяется одна цифра, то это едва замечается, и число 1324519 не различается от 1324517; читателю надо сделать усилие, чтобы убедиться, что четыре числа, написанные в предыдущих строках — все разные.

Когда число вроде предыдущего встречается как мера угла в десятичных долях центезимальных секунд, мы не задаёмся вопросом о том, какова была вероятность, чтобы данный угол равнялся именно $13^\circ 42' 51'',7$, ибо мы никогда не поставили бы себе такого вопроса до измерения угла. Этот угол должен, конечно, иметь какое-нибудь значение, и каково бы оно ни было с точностью до одной десятой секунды, можно было бы, измерив его, сказать, что вероятность *a priori*, что это значение будет именно таково, каково оно есть, равняется одной десятимиллионной и что это факт крайне удивительный...

Вопрос заключается в том, нужно ли делать те же оговорки, если установлено, что один из углов треугольника, образованного тремя звёздами, имеет замечательное значение и равен, например, углу равностороннего треугольника ⟨...⟩ или половине прямого угла... Вот что можно сказать по этому поводу: следует очень опасаться склонности считать замечательным обстоятельство, не определённое в точности перед опытом, так как число обстоятельств, могущих, с различных точек зрения, показаться замечательными, очень значительно.»

Частотный подход

Наиболее естественное и распространённое объяснения понятия «вероятность» такое: это предел частоты события, когда количество повторов эксперимента стремится к бесконечности. (Рихард фон Мизес отстаивал этот подход как единственно возможное обоснование теории вероятностей.)

Однако можно возразить, что никто не может наблюдать бесконечную последовательность испытаний, так что практическое применение этого подхода неизбежно должно иметь дело с конечным числом повторений. А для конечного числа повторений наше утверждение уже не так отчётливо: мы не говорим, что для честной монеты доля орлов в точности равна $1/2$ или даже гарантированно близка к $1/2$. Мы говорим лишь, что заметное отклонение от $1/2$ при большом числе испытаний весьма маловероятно. Употребление слова *маловероятно* возвращает нас к тому самому понятию вероятности, которые мы и пытались истолковать, и чтобы разорвать возникающий порочный круг, приходится вновь вернуться к тому или иному варианту «принципа Курно».

Отметим ещё, что частотный подход можно рассматривать как частный случай указанных выше принципов применения теории вероятностей. В самом деле, собы-

тие «число орлов при 1 000 000 бросаний отличается от 500 000 больше чем на 100 000» имеет простое описание и ничтожно малую вероятность. Поэтому мы отвергаем гипотезу честной монеты, наблюдая такое явление (а также не опасаемся такого события всерьёз, если верим в честность монеты).

Динамические и статистические закономерности

Мы описали типичные способы применения теории вероятностей. Но вопрос остаётся: ну хорошо, мы согласились, что теория вероятностей в какой-то степени описывает поведение симметричной монеты или игральной кости. Но что это: новый закон природы или следствие известных законов классической механики? Можем ли мы в каком-то смысле «доказать», что симметричный кубик имеет равные шансы упасть каждой из своих сторон (если бросать его с большой высоты и быстро крутить)?

Не очень понятно, правда, какого рода доказательства тут могут быть. Поэтому лучше сформулировать вопрос более практически. Пусть наш кубик не совсем симметричен, но мы знаем его форму и точное положение центра тяжести. Можем ли мы использовать законы механики, чтобы найти вероятности падения на каждую грань?

На этот вопрос, видимо, классическая механика даёт положительный ответ, хотя бы в принципе. Её законы позволяют рассчитать движение кубика (и конечное положение), если мы знаем начальные условия с достаточной точностью. Другими словами, множество начальных условий в фазовом пространстве разбито на шесть частей, которые соответствуют шести граням кубика. В этом смысле никакой случайности нет, всё детерминировано. Но эти шесть частей фазового пространства так плотно перемешаны, что в любой достаточно малой (но не слишком малой) части пространства в окрестности начального условия они встречаются в определённой пропорции. Эта пропорция и определяет вероятности исходов. Другими словами, какое бы достаточно плавно меняющееся распределение вероятностей на начальных условиях мы не взяли, возникающее распределение вероятностей на исходах будет близко к этой пропорции.

Последнее утверждение можно формализовать математически, введя некоторые оценки для отклонений, размера области в фазовом пространстве, начальных условий и пр. Вероятно, получится трудная (и до сих пор не решённая) математическая задача, но по крайней мере теоретически вероятности исходов могут быть предсказаны на основе детерминированных законов механики.

Будут ли сложными встречающиеся в природе последовательности?

Приведённые выше рассуждения вряд ли убедят философски мыслящих людей. Ну хорошо, скажут они, мы можем вычислить какие-то числа, которые можно интерпретировать как вероятности граней несимметричного кубика, исходя из законов механики и даже не зная распределения на начальных условиях (а лишь

предполагая, что его плотность не слишком резко меняется). Но тем не менее это распределение приносится в детерминированную картину классической механики извне.

Более наглядная формулировка того же самого вопроса такова. Давайте бросим (самую что ни на есть честную) монету миллион раз. Полученную последовательность нулей и единиц попробуем сжать с помощью какого-то из практически используемых компрессоров, скажем, с помощью программы `gzip`. (Для этого её надо вначале разбить на восьмибитовые байты.) Можно быть уверенным, что мы не получим сжатия больше чем на 1%, сколько раз ни повторяй этот опыт. В самом деле, такая степень сжатия возможна лишь для $2^{-10\,000}$-доли всех последовательностей.

Это утверждение (невозможность сжатия) вполне заслуживает названия «закон природы» — оно ничуть не хуже подтверждается экспериментально, чем многие другие законы. Внимание, вопрос: *является ли этот закон природы следствием известных законов классической механики или это некоторый дополнительный* (*по отношению к ней*) *закон*?

Чтобы осознать ситуацию, полезно её упростить. Представим себе, что время у нас дискретно, фазовое пространство представляет собой отрезок $[0, 1]$, а динамический закон (преобразование фазового пространства) задаётся формулой

$$x \mapsto T(x) = \textbf{if } 2x < 1 \textbf{ then } 2x \textbf{ else } 2x - 1.$$

Мы получаем последовательность состояний x_0 (начальное условие), $x_1 = T(x_0)$, $x_2 = T(x_1)$ и так далее. На каждом шаге мы замечаем, в какую половину отрезка попали и пишем 0, если точка будет в левой половине, и 1, если в правой.

Это преобразование обладает обсуждавшимся выше свойством «перемешивания»: если мы выберем начальное условие в каком-то не слишком малом интервале, и посмотрим, к чему это приведёт через какое-то достаточно большое время t, то увидим, что нуль будет получаться примерно в половине точек начального интервала (в смысле меры). В самом деле, наше преобразование в терминах двоичного разложения состоит просто в сдвиге на бит, так что наблюдаемая нами последовательность нулей и единиц представляет собой двоичное разложение начального условия. Действительные числа, в которых t-й бит равен единице, занимают примерно половину места в любом не слишком коротком отрезке. Теперь видно, что любое утверждение относительно получаемой таким образом последовательности (скажем, что её нельзя сильно сжать) является по сути утверждением о начальном условии.

Что же получается? Должны ли мы к динамическим законам добавить ещё один закон природы, говорящий о случайности (= несжимаемости) начальных условий? («Мир был создан в алгоритмически случайном состоянии», если говорить философски торжественно.) Ведь алгоритмические преобразования (соответствующие динамическим законам) не могут увеличивать сложность, поэтому если объекты большой сложности существуют в детерминированном (как мы временно предполагаем в этом обсуждении) мире, то они должны быть там с самого начала. (Отметим, что не удастся объяснить случайность тем, что мы наблюдаем состояние

мира в случайной точке в случайное время: количество битов, нужных для описания времени и места со всей мыслимой точностью, недостаточно для того, чтобы объяснить появление несжимаемых слов длиной 10^6.)

«Бог не играет в кости», говорил Эйнштейн? Что же, вместо этого мы должны предположить, что он заранее снабдил мир запасом концентрированной случайности при создании?

Случайность как незнание: генераторы псевдослучайных битов

Наше обсуждение становится слишком философским, и его трудно продолжать серьёзно. Но есть важные математические результаты, которые имеют прямое отношение к этому обсуждению, и мы сейчас о них немного поговорим. Вообще в этой книге мы избегали ограничений на время и память в рассматриваемых алгоритмах, оставаясь в рамках общей теории вычислимости. (Сложность с ограничениями на ресурсы — интересная, трудная и пока не очень развитая область, которая осталась целиком вне нашей книги.) Тем не менее при обсуждении природы вероятности не упомянуть об этих результатах нельзя.

Мы имеем в виду понятие генератора псевдослучайных битов (в смысле Блюма, Микэли и Яо; об этом центральном понятии современной сложностной криптографии можно прочесть, например, в книге Голдрайха [48]). Существование таких генераторов не доказано, но следует из некоторых сложностных предположений (существование односторонних функций), которые многие считают правдоподобными.

Объясним неформально, о чём идёт речь. Представим себе простую и быструю (полиномиальную) процедуру, которая берёт «зародыш» случайности (*seed* в англоязычной литературе), двоичное слово сравнительно небольшой длины, скажем, из тысячи битов. Эта процедура преобразует его в гораздо более длинное слово, скажем, в последовательность из 10^{10} битов. Естественно, что получившееся слово имеет малую сложность, поскольку оно может быть задано указанием самой процедуры (которая проста по нашему предположению) и зародыша. Более того, даже и сложность с ограничением на время мала, поскольку преобразование предполагается легко вычислимым.

Тем не менее, если нам покажут такое слово из 10^{10} битов, то мы можем и не отличить его от по-настоящему случайной последовательности. Это звучит парадоксально: мы только что сказали, что колмогоровская сложность такого слова мала, а для случайных слов она близка к длине, в каком же смысле мы говорим о неотличимости?

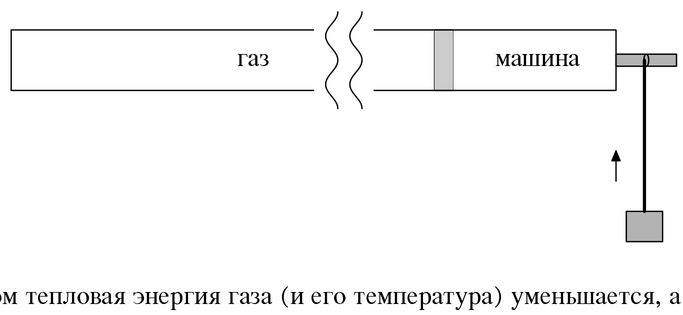
Смысл этот вполне практический — мы бы сразу согласились с тем, что это слово имеет малую сложность и потому нетипично, если бы нам показали зародыш. Но если нам дали это слово длины 10^{10} просто так, то перебрать все 2^{1000} возможных зародышей выше человеческих сил (даже если иметь в виду всё человечество со всеми его компьютерами), и никто так никогда и не узнает, что оно на самом деле было малой сложности.

Это соображение показывает, что в принципе мир может управляться простыми динамическими законами с достаточно простыми начальными условиями (описываемыми, скажем, тысячами битов). В таком мире «истинной» случайности не существует, и любая последовательность из 10^6 бросаний монеты (уже случившаяся, или которая случится в будущем) будет сильно сжимаемой. Тем не менее, вычислительно ограниченный наблюдатель (вроде нас самих) может никогда этого и не обнаружить.

Отступление: термодинамика

Соотношение между динамическими и статистическими законами часто обсуждается в связи со «вторым началом термодинамики». Однако далеко не всегда это обсуждение ведётся в достаточно точных терминах. Например, часто говорят, что второе начало термодинамики принципиально не может быть следствием динамических законов, поскольку те обратимы во времени, а второе начало — нет. С другой стороны, столь же часто говорят, что второе начало термодинамики имеет много эквивалентных формулировок, и одна из них утверждает, что «вечный двигатель второго рода» невозможен: не бывает машины, при циклической работе которой производится механическая работа только за счёт охлаждения некоторого резервуара.

Однако, как разъяснил авторам Никита Маркарян (частное сообщение), в такой формулировке второе начало термодинамики вполне может быть следствием динамических законов. Приведём набросок этого рассуждения. Представим себе вечный двигатель второго рода как машину, соединённую с длинным цилиндром. В этом цилиндре находится разогретый газ, и флуктуации давления газа на поршень передаются в машину (тем самым происходит теплообмен между газом и машиной). Затем эта машина каким-то образом преобразует эту тепловую энергию в механическую и вращает вал, наматывая верёвку и поднимая груз.



При этом тепловая энергия газа (и его температура) уменьшается, а груз поднимается. Само по себе это уменьшение температуры и подъём груза ещё не означает, что мы имеем дело с вечным двигателем второго рода. В самом деле, внутри машины вполне могут быть запасы льда, охлаждающие газ и пружина, поднимающая груз. От настоящего вечного двигателя требуют, чтобы угол поворота оси и переданная грузу энергия могли быть сколько угодно большими, если только запасы горячего газа (и длина цилиндра с газом, к нему подсоединённого) достаточно

велики. Кроме того, мы должны уточнить, при каких начальных условиях (координаты и скорости молекул газа) мы ожидаем такого поведения машины. Естественно потребовать, чтобы такое поведение происходило при большинстве начальных условий (с точки зрения естественной меры на точках фазового пространства, имеющих данную энергию).

После всех этих приготовлений уже можно объяснить, почему такого рода машина невозможна. Фазовое пространство всей системы (газ плюс машина) можно примерно представить как произведение двух компонент: фазового пространства самой машины и фазового пространства газа. Компоненты эти зависимы, и энергия каждой из них может меняться, но общая энергия сохраняется. Поскольку машина сама по себе имеет фиксированное число частей, не зависящее от объёма цилиндра, размерность её компоненты (число степеней свободы машины) пренебрежимо мала по сравнению с числом степеней свободы газа. Фазовое пространство для газа разбито на части в соответствии с уровнями энергии. При этом с ростом энергии газа объём соответствующей части в фазовом пространстве для газа растёт, а в фазовом пространстве машины — уменьшается. Но скорость роста сильно превосходит скорость уменьшения, поскольку показатель экспоненты определяется числом степеней свободы, а у газа оно гораздо больше. Таким образом, подавляющая часть фазового пространства всей системы сосредоточена в области, когда энергия газа велика, а энергия машины мала.

Остаётся вспомнить, что фазовый объём (в пространстве всей системы) сохраняется. Поэтому невозможно, чтобы траектория, начав движение в любой (или почти любой) точке в большой части, в итоге попала бы в малую часть. Вероятность этого события не превосходит отношения объёмов. Поэтому описанное выше поведение системы, которое позволило бы назвать её вечным двигателем второго рода, невозможно.

Это рассуждение, конечно, не является сколько-нибудь строгим и многого не учитывает. На самом деле мера на фазовом пространстве всей системы не является в точности произведением мер отдельно для газа и отдельно для машины (они взаимодействуют); начальные условия могут заранее лежать в малом множестве, если начальное состояние машины фиксировано с очень большой точностью (что не удивительно с точки зрения физика: это значит, что вначале машина очень холодная и потому может играть роль теплоприёмника). Тем не менее эти рассуждения показывают, что в принципе законы, традиционно относимые к статистическим, могут быть выведены из динамических законов без каких-либо дополнительных предположений.

Ещё одно отступление: квантовая механика

Другая физическая теория, постоянно упоминаемая как источник случайности — это квантовая механика. Принципиальные основы квантовой механики и её соотношение с классической и статистической физикой — предмет множества (не всегда понятных) обсуждений. Не вдаваясь в их детали, мы можем тем не менее отметить, что соотношение между квантово-механическими моделями и на-

блюдениями очень похоже на соотношение между статистическими гипотезами и наблюдениями.

Статистическая гипотеза приписывает возможным исходам вероятности; квантовая механика приписывает возможным исходам эксперимента не вероятности, а *амплитуды*. Эти амплитуды являются комплексными числами, а не вещественными, но это не мешает сформулировать «квантовый принцип Курно»: *события с малыми амплитудами не происходят*. Точнее говоря, следует сказать так: если амплитуда события A (по модулю) значительно меньше амплитуды события B, то реальность события A следует учитывать в гораздо меньшей степени (при одинаково тяжёлых последствиях). Объяснение, почему это даёт возможность вообще игнорировать события с малой амплитудой, остаётся прежним.

Интерпретация квадрата модуля амплитуды как вероятности (ожидаемой частоты) может быть тогда получена тем же способом, как и раньше. Представим себе систему, состоящую из N идентичных независимых систем, каждая из которых может находиться в двух состояниях, причём амплитуда состояния 1 равна z (в каждой из систем). Точнее говоря, состояние всей большой системы равно тензорному произведению описанных состояний отдельных систем. В этом случае амплитуда события «число единиц среди состояний сильно отличается от $N|z|^2$» (квадрат модуля проекции указанного состояния на описанное подпространство) мала — это по существу тот же самый закон больших чисел.

Подобным же образом, вероятно, можно анализировать конкретные механизмы измерения физических величин и «доказывать» (со ссылкой на тот же принцип Курно), что частота появления того или иного результата измерения будет близка к квадрату длины проекции состояния на соответствующее измерению подпространство, за исключением события малой амплитуды.

Приложение 2. Четыре алгоритмических лица случайности

В. А. Успенский[1]

Введение

Если кто-либо скажет нам, что он подбросил «честную» монету двадцать раз и, обозначив герб единицей, а решётку — нулём, получил такой результат:

$$10001011101111010000 \tag{I}$$

или такой:

$$01111011001101110001, \tag{II}$$

мы вряд ли будем удивлены. Однако если нам скажут, что результат бросаний был таким:

$$00000000000000000000 \tag{III}$$

или таким:

$$01010101010101010101, \tag{IV}$$

мы будем поражены и вообще не поверим или же усомнимся в корректности эксперимента. Возникает вопрос: почему?

По-видимому, цепочки (I) и (II) воспринимаются как случайные, а цепочки (III) и (IV) — как неслучайные.

Но что означают слова «воспринимается как случайная»? Классическая теория вероятностей не даёт ответа на этот важный вопрос. Не столь редко можно услышать следующее объяснение: вероятность каждой из цепочек (III) и (IV) слишком мала, она равна 2^{-20}, что меньше одной миллионной. Но ведь ровно такую же вероятность имеют цепочки (I) и (II)!

[1]Это приложение воспроизводит (с некоторыми изменениями) текст брошюры «Четыре алгоритмических лица случайности» (2-е изд., испр. — М.: МЦНМО, 2009; первое издание вышло в 2006 г.), написанной по материалам лекции, прочитанной автором брошюры, В. А. Успенским, 23 июля 2005 года в летней школе «Современная математика» в Дубне. Терминология и обозначения, использованные в брошюре, несколько отличаются от использованных в остальной части книги; эти различия отмечены в подстрочных примечаниях.

Сделаем три важных замечания.

- Во-первых, впечатление случайности зависит от распределения вероятностей. Если одна сторона монеты тяжелее другой или если человек в процессе бросания научится подкидывать монету так, чтобы она падала на нужную сторону, то появление цепочек (III) или (IV) может стать вполне ожидаемым. Сначала — для простоты — мы будем заниматься лишь независимыми бросаниями с вероятностями одна вторая для герба и одна вторая для решётки.
- Во-вторых, говорить о случайности имеет смысл лишь в применении к очень длинным цепочкам. Бессмысленно спрашивать, которая из цепочек 00, 01, 10, 11 более случайна, чем другие.
- В-третьих, точной границы между случайными и неслучайными (в интуитивном смысле) цепочками нет и не может быть. Ведь если в случайной цепочке заменить один знак, она остаётся случайной. Но, заменяя много раз, мы от любой цепочки можем прийти к (III) или (IV). Это известный *парадокс кучи*.

Итак, следует рассматривать только очень длинные цепочки, а в идеале — бесконечные (вообще, бесконечность — это такое полезное приближение сверху к очень большому конечному). Бесконечные цепочки принято называть *последовательностями*. Оказывается, последовательности уже можно довольно осмысленно разделить на случайные и неслучайные. Иными словами, можно не без успеха пытаться найти строгое математическое определение для понятия ***случайная последовательность нулей и единиц***. В настоящей брошюре мы изложим результаты таких попыток, предпринятых различными авторами. Однако следует честно признать, что для практических приложений интерес представляют именно конечные случайные цепочки и потому идеализация, происходящая при переходе к цепочкам бесконечной длины, неизбежно сопряжена с «отрывом от жизни». Впрочем, аналогичный отрыв возникает и при изучении совокупности *всех* конечных цепочек, поскольку в жизни встречаются лишь цепочки ограниченной длины, а в очень длинных случайных цепочках возникают такие эффекты, которые могут и не соответствовать наивным представлениям о случайности. Сделав это признание, приступим к изложению.

Нам будут полезны некоторые термины и обозначения.

Всякая конечная цепочка $\langle i_1, \ldots, i_n \rangle$ нулей и единиц называется также *двоичным словом*, а число n — *длиной* этого слова. Длина слова x обозначается так: $|x|$.[1] Длина слова может быть равна нулю; слово длины нуль ничего не содержит и называется *пустым*. Пустое слово обозначается буквой Λ.

Множество всех двоичных слов обозначается буквой Ξ. Множество всех последовательностей нулей и единиц обозначается буквой Ω.

Для последовательности $\langle a_1, a_2, \ldots \rangle$ каждое из двоичных слов $\langle a_1, a_2, \ldots, a_n \rangle$ является её *началом*. Пусть $x \in \Xi$ — двоичное слово. Множество всех тех последовательностей из Ω, для которых x служит началом, обозначим через Ω_x. Каждое

[1] В основном тексте книги длина слова x обозначалась $l(x)$.

множество вида Ω_x, где $x \in \Xi$, принято называть *шаром*. Шару Ω_x приписывается его *объём*[1] $\mathbf{v}(x)$, равный $2^{-|x|}$.

Содержательно, каждую последовательность из Ω мы рассматриваем как серию результатов бросаний монеты. Ещё раз подчеркнём, что для наглядности мы ограничиваемся на первых порах ситуацией, когда результаты всех бросаний равновероятны. На математическом языке это означает, что на Ω задано так называемое *равномерное распределение вероятностей*, а это, в свою очередь, означает следующее: для каждого шара вероятность того, что случайно выбранная последовательность из Ω попадёт в этот шар, равна объёму этого шара.

Наша цель — попытаться выделить из Ω некоторое точно описанное подмножество, претендующее на роль множества всех случайных последовательностей. Традиционная теория вероятностей не только не приближается к решению этой задачи, но даже не может её сформулировать в своих терминах. На помощь приходит теория алгоритмов. Может показаться парадоксальным, что понятие случайности уточняется на основе такого чуждого случайности понятия, как алгоритм, — тем не менее, дело обстоит именно так: все известные до сих пор определения случайности индивидуального объекта (в нашем примере — индивидуальной последовательности нулей и единиц) опираются на понятие алгоритма.

Чтобы найти требуемое определение, обычно поступают так: формулируют некое характеристическое свойство, которым обладают случайные (в неформальном, интуитивном смысле) последовательности, а затем последовательности, обладающие этим свойством, и объявляют — по определению — случайными.

Какими же свойствами обладает случайная последовательность нулей и единиц?

Во-первых, она *частотоустойчива*. Вот что это означает для того простейшего случая, когда нули и единицы равновероятны: частота нулей, как и частота единиц, стремится к одной второй (*частота нулей* — это их доля в начальном отрезке последовательности). Но более того: в случайной последовательности указанная устойчивость частот выполняется не только для последовательности в целом, но и для любой её законной, разумной подпоследовательности.

Во-вторых, она *хаотична*. Это означает, что она сложно устроена и не может иметь разумного описания. Психологическому эксперименту, с которого мы начали, Колмогоров дал такое объяснение. Цепочки (I) и (II) потому воспринимаются как случайные, что они сложны, их устройство нельзя коротко описать. А вот цепочки (III) и (IV) имеют простое, легко описываемое устройство.

В-третьих, она *типична*. Это означает, что она принадлежит любому разумному большинству.

В-четвёртых, она *непредсказуема*. Это означает, что, играя против неё на деньги (т. е. пытаясь угадать члены последовательности и делая ставки), последовательность невозможно обыграть вне зависимости от разумности стратегии.

Конечно, слово «разумный», встречающееся в объяснениях перечисленных четырёх свойств, нуждается в уточнении. Теория алгоритмов как раз и предлагает такие уточнения, наполняя это слово точным смыслом — своим для каждого

[1] В основном тексте книги мы говорили об *интервалах* и их *длинах*.

из наших четырёх свойств. Тем самым, возникают четыре алгоритмических свойства: *частотная устойчивость*, *хаотичность*, *типичность*, *непредсказуемость*. Каждое из них представляет своё собственное алгоритмическое лицо случайности, и каждое из них, с бо́льшими или меньшими основаниями, может претендовать на роль строгого математического определения понятия случайности. Можно сказать и так: возникают четыре точно очерченных класса последовательностей, каждый из которых претендует на то, чтобы служить «истинным» классом случайных последовательностей; некоторые из этих претензий более оправданы, чем другие.

Итак, цель дальнейшего изложения такова. Во-первых, предложить строгие определения для перечисленных выше четырёх свойств и, тем самым, для возникающих на основе этих свойств четырёх классов последовательностей. Во-вторых, сформулировать известные на сегодня взаимоотношения между названными свойствами и, следовательно, между соответствующими классами последовательностей.

Лицо первое: Частотоустойчивость и стохастичность

По-видимому, одним из первых поставил вопрос о том, что такое отдельно взятая случайная последовательность, замечательный немецкий математик Рихард фон Мизес в начале XX века — в 1919 г. Во всяком случае, именно он первым предложил сравнительно удачное (хотя и нестрогое) определение, послужившее отправной точкой для дальнейшего развития.

Мизес исходил из того, что случайной последовательности должна быть присуща устойчивость частот, а именно: доля единиц (как и доля нулей) в начальном отрезке случайной последовательности должна стремиться к одной второй при неограниченном увеличении длины начального отрезка. Но этого недостаточно. Например, этим свойством устойчивости частот обладает заведомо неслучайная последовательность

$$0, 1, 0, 1, 0, 1, 0, 1, \ldots .$$

Очевидным образом необходимо, чтобы устойчивостью частот обладала не только вся последовательность целиком, но и её подпоследовательности. Однако устойчивостью частот не могут обладать *все* подпоследовательности. В самом деле, возьмём самую что ни на есть случайную последовательность и отберём в подпоследовательность те её члены, которые равны нулю... Значит, подпоследовательность, в которой должна соблюдаться устойчивость частот, должна быть *разумной*, или *допустимой*. Например, если отобрать в подпоследовательность все те члены, чьи номера суть простые числа, или же все те члены, которые непосредственно следуют за нулевыми, то полученная в каждом из этих двух вариантов подпоследовательность является допустимой.

Последовательности, в которых устойчивость частот соблюдается во всех её допустимых подпоследовательностях, Колмогоров предложил называть *стохастическими*.

Сам Мизес объяснял, какие последовательности следует считать допустимыми, в расплывчатых терминах. Первое уточнение предложил в 1940 г. американский

математик Алонзо Чёрч — один из создателей теории алгоритмов. Чёрч предложил, чтобы допустимые подпоследовательности строились на основе определённых алгоритмов. Последовательности, в которых устойчивость частот наблюдается во всех допустимых по Чёрчу подпоследовательностях, получили название *стохастических по Чёрчу*.[1] Однако оказалось, что определение Чёрча слишком широко: существует, например, стохастическая по Чёрчу последовательность, перестающая быть таковой после вычислимой перестановки её членов.[2]

В 1963 г. Колмогоров усовершенствовал определение Чёрча, предложив более обширный класс допустимых алгоритмов отбора членов исходной последовательности в подпоследовательность и, тем самым, заведомо не меньший (а на деле более обширный) класс допустимых подпоследовательностей; в частности, колмогоровский алгоритм разрешает членам подпоследовательности иметь иной порядок следования, чем в исходной последовательности. Поэтому класс всех последовательностей, *стохастических по Колмогорову* (т. е. таких, у которых устойчивость частот наблюдается во всех допустимых по Колмогорову подпоследовательностях),[3] является подклассом класса последовательностей, стохастических по Чёрчу, а на самом деле, даже строгим подклассом, т. е. не совпадающим с объемлющим классом. Для дальнейших ссылок класс всех последовательностей, стохастических по Колмогорову, обозначим буквой **S**.

Выяснилось, что класс **S** также слишком широк. Оказалось, например, что может быть построена такая стохастическая по Колмогорову последовательность, в каждом начальном отрезке которой единиц больше, чем нулей; а это противоречит как интуиции, так и законам теории вероятностей (например, закону возвратности состояний при случайном блуждании).[4] Таким образом, даже наиболее продвинутое из известных на сегодняшний день математически строгих уточнений идей Мизеса не даёт полного отражения интуитивного представления о случайности (хотя и это неполное отражение может оказаться полезным).

Для большей ясности приведём определения Чёрча и Колмогорова. Центральным для обоих определений является указание тех правил, согласно которым из членов рассматриваемой последовательности составляются допустимые подпоследовательности. Поскольку каждое такое правило производит отбор членов последовательности для включения их в допустимую подпоследовательность, сами эти правила принято называть *допустимыми правилами отбора*.[5]

Для наглядности представим себе, что члены исследуемой последовательности написаны на картах. Карты лежат одна за другой, лицевой стороной вниз, так что мы не видим, что на них написано. Наша цель — отобрать некоторые из этих карт и составить из них другую последовательность — допустимую подпоследовательность. (Как мы увидим, в случае колмогоровского определения термин «подпоследовательность» имеет более широкий смысл, чем это обычно принято.) Допустимое правило отбора представляет собою алгоритм, который на каждом этапе постро-

[1] В основном тексте книги они называются *случайными по Мизесу – Чёрчу*.

[2] См. теорему 203(г) на с. 345.

[3] В основном тексте они называются *случайными по Мизесу – Колмогорову – Лавлэнду*.

[4] См. теорему 203(б) на с. 345.

[5] В основном тексте они были названы допустимыми правилами *выбора*.

ения подпоследовательности предписывает, во-первых, которую из карт надлежит открыть и, во-вторых, следует или нет включать эту открываемую карту в подпоследовательность. Входом алгоритма служит информация о всех уже открытых к этому моменту картах. Может случиться, что алгоритм отберёт лишь конечное число членов исходной последовательности; в этом случае считается, что никакой допустимой подпоследовательности не образовалось. (Напомним, что исходная последовательность признаётся стохастической при наличии устойчивости частот в любой её допустимой подпоследовательности.)

Переходим к более точному описанию.

Функция называется *вычислимой*, коль скоро существует *вычисляющий* её алгоритм. При этом говорят, что алгоритм *вычисляет* функцию f, коль скоро он перерабатывает всякий её аргумент x, на котором функция определена, в соответствующее значение $f(x)$ и не выдаёт никакого результата в применении ко всякому такому аргументу, на котором функция не определена.

Исследуемую на устойчивость последовательность обозначим через $a_1, a_2, \ldots$, так что на n-й карте записана цифра a_n, равная 0 или 1.

Допустимое по Чёрчу правило отбора является произвольной вычислимой функцией G, определённой на множестве Ξ всех двоичных слов и принимающей одно из двух значений: *Да* и *Нет*. Карты открываются последовательно, одна за другой, начиная с первой, и каждый раз — до открытия очередной карты — решается вопрос, включать ли эту карту в подпоследовательность. Вопрос этот решается следующим образом. Пусть уже открыты n карт, на которых записаны, соответственно, цифры $a_1, \ldots, a_n$. Если $G(a_1, \ldots, a_n)$ есть *Нет*, то следующая, $(n+1)$-я, карта не включается в подпоследовательность. Если $G(a_1, \ldots, a_n)$ есть *Да*, то следующая, $(n+1)$-я, карта включается в подпоследовательность в качестве очередного члена. Включать или не включать в подпоследовательность член a_1, зависит от значения $G(\Lambda)$. Таким образом, правилу G отвечает бесконечная подпоследовательность (в случае, если таковая образовалась)

$$a_{n(1)}, a_{n(2)}, a_{n(3)}, \ldots,$$

где числа $n(1), n(2), n(3), \ldots$ образуют возрастающую последовательность, составленную из всех чисел n, для которых $G(a_1, \ldots, a_{n-1}) = \text{\textit{Да}}$.

Изложению определения Колмогорова предпошлём обобщение понятия подпоследовательности. *Обобщённой подпоследовательностью* последовательности $a_1, a_2, \ldots, a_n, \ldots$ назовём всякую последовательность вида

$$a_{\varphi(1)}, a_{\varphi(2)}, \ldots, a_{\varphi(k)}, \ldots,$$

для которой выполнено условие

$$i < j \;\Rightarrow\; \varphi(i) \neq \varphi(j).$$

Для обычной подпоследовательности (которая является частным случаем обобщённой) это условие заменено на более сильное — с правой частью $\varphi(i) < \varphi(j)$.

Каждое допустимое по Колмогорову правило отбора имеет целью создать некоторую обобщённую подпоследовательность исходной последовательности. Мы

потому написали «имеет целью создать», а не «создаёт», что процесс создания может прерваться, и тогда в результате применения правила возникает не бесконечная последовательность, а *кортеж* (т. е. конечный набор), составленный из членов исходной последовательности. Для стохастичности по Колмогорову требуется устойчивость частот в каждой из допустимых, т. е. образованных каким-либо допустимым по Колмогорову правилом, обобщённых подпоследовательностей.

Само *колмогоровское* (т. е. допустимое по Колмогорову) *правило* состоит из двух вычислимых функций: F и G. Первая служит для образования некой вспомогательной обобщённой подпоследовательности. Окончательная же обобщённая подпоследовательность исходной последовательности строится при помощи функции G в качестве обычной подпоследовательности вспомогательной обобщённой подпоследовательности. Аргументами функций F и G служат двоичные слова, так что каждая из этих функций определена на своём подмножестве множества Ξ. Значениями функции F служат целые положительные числа, у функции G два возможных значения: *Да* и *Нет*. Сперва строится последовательность натуральных чисел $n(1), n(2), n(3), \ldots$:

$$n(1) = F(\Lambda),\ \ n(2) = F(a_{n(1)}),\ \ \ldots,\ \ n(k+1) = F(a_{n(1)}, \ldots, a_{n(k)}).$$

Построение этой последовательности прекращается (и возникает конечный кортеж), как только наступает один из трёх случаев:

- значение $F(a_{n(1)}, \ldots, a_{n(k)})$ не определено;
- значение $G(a_{n(1)}, \ldots, a_{n(k)})$ не определено;
- значение $F(a_{n(1)}, \ldots, a_{n(k)})$ совпадает с одним из чисел $n(1), \ldots, n(k)$.

Если же остановки в построении не произошло и возникла бесконечная последовательность номеров $n(1), n(2), n(3), \ldots$, то далее строится вспомогательная обобщённая подпоследовательность $a_{n(1)}, a_{n(2)}, a_{n(3)}, \ldots$. Из неё, наконец, отбираются — в порядке возрастания k — те её члены $a_{n(k)}$, для которых выполнено равенство $G(a_{n(1)}, \ldots, a_{n(k-1)}) = $ *Да*.

Лицо второе: Хаотичность

Вернёмся к примерам конечных цепочек из Введения и вспомним объяснение, предложенное Колмогоровым: цепочки (I) и (II) воспринимаются как случайные, потому что они *сложны*; цепочки (III) и (IV) воспринимаются как неслучайные, потому что они *просты*. По-видимому, мы ожидаем, что результат случайного эксперимента окажется сложным, и удивляемся, когда получаем что-то простое.

Некоторые объекты можно квалифицировать как большие или маленькие, другие — как тяжёлые или лёгкие, третьи — как сложные или простые. В начале 60-х годов Колмогоров наметил математическую теорию, позволяющую оценивать сложность объектов. Сейчас эта теория называется **теорией колмогоровской сложности**.

В основе теории колмогоровской сложности лежит следующая простая и естественная идея: ***сложность объекта измеряется длиной его кратчайшего описания***.

В самом деле, каждый объект может иметь сколь угодно сложные описания, но сложный объект невозможно описать коротко.

Пусть Y — множество всех объектов, которые мы рассматриваем, а X — множество всех мыслимых описаний этих объектов. Через $|x|$ обозначим длину описания x. Сложность объекта y обозначим $\mathrm{Comp}(y)$. В соответствии со сказанным,

$$\mathrm{Comp}(y) = \min_x\{|x|\colon x \text{ есть описание для } y\}.$$

Если объект y неописуем, т. е. для него не существует описания, то, естественно, его сложность равна бесконечности.

Разумеется, длины надо мерить единообразным способом, чтобы не получилось, что описание одного и того же объекта имеет на китайском языке длину единица (один иероглиф), а на русском — сóрок (сóрок букв). Поэтому все описания кодируются в двоичном алфавите — в виде двоичных цепочек. Вот эти-то двоичные коды мы и будем отныне считать описаниями, так что отныне $X = \Xi$.

Множество всех таких пар $\langle x, y\rangle$, что x служит описанием для y, естественно называть *языком описания.*

Заметим, что и один и тот же объект может иметь в данном языке много описаний, и одно и то же описание может служить описанием многих объектов. Таково, например, описание «двоичное слово, состоящее только из нулей», а, скажем, выражение «двоичное слово» годится в качестве описания для любого двоичного слова.

Всё сказанное носило предварительный характер, чтобы оправдать нижеследующее формальное изложение.

В декартовом произведении $\Xi \times Y$ рассматриваем произвольное подмножество E, которое называем *языком описания.* Если $\langle x, y\rangle \in E$, будем говорить, что x является *описанием* объекта y. *Сложность* Comp_E объекта y *относительно* языка E определяется так:

$$\mathrm{Comp}_E(y) = \min_x\{|x|\colon \langle x, y\rangle \in E\}.$$

Напомним, что минимум по пустому множеству принято считать равным бесконечности.

Для языка $E = \Xi \times Y$, в котором каждое x служит описанием для каждого y, сложность любого объекта y равна нулю, поскольку одним из описаний этого y является пустое слово; такой язык мыслим, но не будет встречаться в тех семействах языков, которые мы будем рассматривать.

Представим себе два языка, причём описание какого-либо объекта на втором языке происходит путём удвоения описания этого же объекта на первом языке. Ясно, что второй язык «хуже» первого. Предпочтительнее те языки, которые в состоянии давать более короткие описания.

Будем говорить, что язык A *не хуже* языка B и писать $A \leqslant B$, если существует такая константа c, что для всякого y справедливо неравенство $\mathrm{Comp}_A(y) < \mathrm{Comp}_B(y) + c$.

Если принять, что для так называемых естественных языков, т. е. для тех, которыми пользуется человечество, могут быть сформулированы правила перевода с одного языка на другой, то становится очевидным, что каждый из этих языков

не хуже другого. Ведь, скажем, турецкое описание какого-либо объекта можно составить так: взять произвольное японское описание этого объекта и дополнить его правилами японо-турецкого перевода. И турецкое, и японское описания, и правила перевода считаются закодированными в виде двоичных слов. Поэтому длина такого турецкого описания равна длине японского описания плюс не зависящая от выбора объекта длина правил перевода. В качестве японского описания выберем кратчайшее. В итоге получаем, что турецкий язык не хуже японского.

Встаёт вопрос о выборе оптимального языка — такого, который не хуже любого другого. Пусть $\mathscr{L}$ — некоторое языково́е семейство, т. е., попросту, некоторое множество языков. Язык A из семейства $\mathscr{L}$ называется *оптимальным* (для $\mathscr{L}$), если он не хуже любого другого языка из этого семейства, т. е. если

$$\forall B \in \mathscr{L} \text{ выполнено } A \leqslant B.$$

Если оптимальный язык существует, то именно с его помощью следует измерять сложность. Сложность объекта относительно одного из оптимальных языков называется *алгоритмической энтропией* этого объекта.[1] Энтропию и рассматривают как окончательную искомую меру сложности в рамках заданного языково́го семейства.

Для некоторых важных языковы́х семейств имеет место теорема о существовании оптимального языка, а тем самым и энтропии. Эта теорема называется *теоремой Соломонова – Колмогорова.*

Для данного семейства может существовать много оптимальных языков и тем самым много энтропий. Однако, в силу определения оптимальности, любые две энтропии (взятые для одного и того же фиксированного языково́го семейства) различаются не более чем на аддитивную константу. Иными словами, если A и B суть два оптимальных для $\mathscr{L}$ языка, то существует такая константа c, что для всех y

$$|\mathrm{Comp}_A(y) - \mathrm{Comp}_B(y)| < c.$$

Замечание. Конечно, неприятно, что алгоритмическая энтропия, претендующая на роль «истинной сложности», определяется не однозначно, а всего лишь с точностью до аддитивной добавки, не превышающей константы. Однако попытки найти среди энтропий «наиболее оптимальную» пока что ни к чему хорошему не привели.

Из уважения к Колмогорову алгоритмическую энтропию обозначают обычно[2] буквой K; иногда к этой букве K добавляют ещё вторую букву, указывающую то языково́е семейство, к которому относится рассматриваемая энтропия. Если K' и K'' суть две энтропии, относящиеся к одному и тому же языково́му семейству, то, как отмечалось,

$$|K' - K''| < c.$$

[1] В основном тексте мы пишем «сложность», оставляя слово «энтропия» для шенноновской энтропии.

[2] В англоязычной литературе буква K обычно обозначает префиксный вариант сложности; простая колмогоровская сложность обычно обозначается буквой C. Сам Колмогоров использовал буквы K и H (в разных публикациях — разные).

Колмогоров не только сформулировал понятие алгоритмической энтропии объекта, но и применил его к исследованию случайных последовательностей. Наблюдение Колмогорова состояло в том, что в случайной последовательности энтропия начального отрезка растёт достаточно быстро при неограниченном увеличении длины этого отрезка. (Ничего не поделаешь: случайная последовательность может начинаться с миллиона нулей, но и этот миллион — ничто перед бесконечностью, и если взять все начальные отрезки в совокупности, то их энтропия неизбежно будет расти.)

Итак, в качестве описываемых объектов мы будем рассматривать двоичные цепочки, такие, как (I), (II), (III), (IV) и т. п. Поэтому отныне $Y = \Xi$.

Если язык содержит пару $\langle z, z\rangle$, то это означает попросту, что цепочка z описывает самоё себя. Язык D, состоящий в точности из пар такого вида, называется *диагональным* (термин из математики), или *автонимным* (термин из лингвистики). Для такого языка $\mathrm{Comp}_D(y) = |y|$. Ограничимся языковы́ми семействами, содержащими автонимный язык (таким будет, в частности, семейство монотонных языков, описанное ниже). Тогда для каждой энтропии K, соответствующей этому семейству, и подходящей константы c будет выполнено неравенство

$$K(y) < |y| + c.$$

Таким образом, если пренебречь аддитивной константой, максимально возможное значение для энтропии цепочки равно длине цепочки. Колмогоров предположил, что у случайной последовательности энтропия начального отрезка достигает этого максимума, т. е. равна длине этого отрезка — опять-таки, при пренебрежении аддитивной константой. В этом состояла основная идея Колмогорова относительно хаотичности.

Итак, фиксируем некоторое языково́е семейство и одну из соответствующих этому семейству энтропий K. Назовём последовательность

$$a_1, a_2, \ldots, a_n, \ldots$$

хаотической, если существует такая константа c, что

$$K(a_1, a_2, \ldots, a_n) > n - c.$$

Очевидно, свойство хаотичности не зависит от выбранной энтропии, а только от выбранного семейства.

Оказалось, что при подходяще выбранном языковом семействе строгое понятие хаотичности хорошо отражает интуитивное представление о случайности.

Создавая теорию сложности объектов, Колмогоров придал соотношению ***описание – объект*** алгоритмический характер. Следуя Колмогорову, ограничимся *перечислимыми* языками. Понятие *перечислимого множества* — это одно из основных понятий теории алгоритмов (да и математики в целом). Этому понятию можно дать такое наглядное объяснение. Представим себе безостановочно работающий принтер, последовательно печатающий слова. После напечатания слова принтер делает

пробел, так что слова отделены одно от другого. Перед напечатанием очередного слова принтер берёт время на размышление. Это время может оказаться бесконечным, и тогда слово не печатается вовсе; в таком случае напечатанным будет лишь конечное множество слов — в частности, пустое множество, если принтер в самом начале своей работы задумался навсегда. Так вот, множество всех когда-либо напечатанных таким принтером слов окажется перечислимым — и всякое перечислимое множество может быть получено таким образом при подсоединении принтера к «идеальному компьютеру». Перечислимо и множество теорем любой формальной теории. Здесь нет места разъяснять, ни что такое идеальный компьютер, ни что такое формальная теория. Но определение понятия ***перечислимое множество*** мы сейчас приведём.

Для наглядности начнём с термина «счётное множество». Термин этот имеет два варианта толкования. В первом, более узком (и более распространённом), толковании счётное множество — это такое множество, которое можно поставить во взаимно однозначное соответствие с натуральным рядом. Во втором, более широком, толковании счётное множество — это такое множество, которое можно поставить во взаимно однозначное соответствие с каким-либо начальным отрезком натурального ряда. Напомним, что *начальным отрезком натурального ряда* называется произвольное множество M натуральных чисел, содержащее вместе со всяким своим элементом и все меньшие числа. Таким образом, и весь натуральный ряд, и пустое множество являются начальными отрезками. Поэтому при втором понимании счётности все конечные множества, в том числе пустое множество, оказываются счётными. Именно это второе, более широкое понимание счётности удобно для наших целей; его мы и примем. А тогда счётному множеству можно дать и такое определение: *множество называется счётным, если оно либо пусто, либо может быть расположено в последовательность* (*т. е. является множеством членов какой-либо последовательности*). Например, конечное множество $\{a, b, c\}$ можно расположить в последовательность $a, b, c, c, c, c, \ldots$. Заменяя в приведённом определении счётного множества термин «последовательность» на термин «вычислимая последовательность», мы получаем определение перечислимого множества. Что касается определения термина «вычислимая последовательность», то оно сейчас будет дано.

Последовательность $w_1, w_2, \ldots, w_n, \ldots$ называется *вычислимой*, если существует алгоритм, вычисляющий её n-й член w_n по его номеру n. Поэтому понятие вычислимой последовательности называют *алгоритмическим аналогом*, или *эффективным аналогом*, понятия последовательности, а понятие перечислимого множества — алгоритмическим аналогом, или эффективным аналогом, понятия счётного множества. Перечислимые множества называются также *эффективно счётными*. Повторим определение: *множество называется перечислимым, или эффективно счётным, если оно либо пусто, либо может быть расположено в вычислимую последовательность* (*т. е. является множеством членов какой-либо вычислимой последовательности*).

Все рассматриваемые нами языки суть подмножества декартова произведения $\Xi \times \Xi$ и, следовательно, счётны. Идеология Колмогорова состояла в том, чтобы рассматривать только эффективно счётные (они же — перечислимые) языки.

Окончательно надлежащий выбор языково́го семейства произвёл колмогоровский ученик Леонид Левин. Именно, в 1973 г. он ввёл в рассмотрение семейство монотонных языков и изучил соответствующее этому семейству понятие хаотичности.[1] Приведём необходимые определения.

Будем говорить, что двоичные слова u и v *согласованы*, и писать $u \approx v$, если одно из этих слов есть начало другого.

Язык E называется *монотонным*, если он перечислим и выполнено условие:

$$\big(\langle x_1, y_1\rangle \in E \& \langle x_2, y_2\rangle \in E \& (x_1 \approx x_2)\big) \ \Rightarrow \ (y_1 \approx y_2).$$

Можно показать, что в семействе монотонных языков имеется оптимальный. Алгоритмическая энтропия, соответствующая семейству монотонных языков, называется *монотонной энтропией*[2] и обозначается KM.

Всякую последовательность, являющуюся хаотической для семейства монотонных языков, будем называть просто *хаотической*.[3] Условие хаотичности последовательности $\langle a_1, a_2, a_3, \ldots\rangle$ можно записать так:

$$\exists c \ \forall n \text{ выполнено } KM(a_1, a_2, \ldots, a_n) > n - c.$$

Класс всех хаотических последовательностей обозначается буквой **C**.

Считается, что строгое понятие хаотической последовательности может служить хорошим отражением интуитивного понятия случайной последовательности. К тому есть два основания.

Во-первых, для каждой отдельно взятой хаотической последовательности выполняются основные законы теории вероятностей.

Во-вторых, класс хаотических последовательностей совпадает с другим «претендентом» на роль строгого аналога расплывчатого класса случайных последовательностей, а именно с классом **T** всех типических последовательностей (о нём будет рассказано далее):

$$\mathbf{C} = \mathbf{T}.$$

Поэтому хаотические последовательности можно было бы называть *типическо-хаотическими* или *хаотическо-типическими*, а сам класс обозначать **TC** или **CT**. Этот класс у́же класса **S** всех последовательностей, стохастических по Колмогорову (последний, как уже́ отмечалось, слишком широк):

$$\mathbf{TC} \subset \mathbf{S}, \quad \mathbf{TC} \neq \mathbf{S}.$$

Лицо третье: Типичность

Сказать про какой-либо объект, что он типичен — это значит сказать, что он принадлежит к любому подавляющему большинству. Например, типичный человек,

---

[1] Близкое понятие ввёл в 1972 г. Шнорр, см. об этом ниже.

[2] В основном тексте используется термин «монотонная сложность». Приведённое только что определение монотонной сложности рассматривается в разделе 6.2.

[3] Поскольку это свойство оказывается равносильным случайности по Мартин-Лёфу (типичности), отдельного названия для него в основном тексте не используется.

встреченный на московской улице, имеет рост менее двух метров (т. е. принадлежит к подавляющему большинству людей, имеющих рост менее двух метров), возраст более трёх лет (т. е. принадлежит к подавляющему большинству людей, имеющих возраст более трёх лет) и т. д. Конечно, любое из большинств, о которых идёт речь, должно быть разумным, ведь какой объект ни возьми, он заведомо не принадлежит к подавляющему большинству, образованному всеми остальными, отличными от данного, объектами.

Мы исходим из интуитивного представления о том, что случайный объект должен обладать свойством типичности. Наша цель — дать строгое математическое определение этого свойства для последовательностей нулей и единиц при равномерном распределении вероятностей на пространстве Ω таких последовательностей. Последовательности, удовлетворяющие этому строгому определению, мы будем называть *типическими*, оставляя слово *типичный* для употребления на интуитивном уровне. В силу сказанного выше, нам надо определить сперва, что есть подавляющее большинство последовательностей, а затем — какие большинства следует считать разумными. А тогда класс типических последовательностей автоматически определится как пересечение всех разумных большинств.

Оставляя термин «подавляющее большинство» для интуитивного употребления, мы будем говорить о *больших* множествах последовательностей. Дополнение к большому множеству до всего Ω будем называть *малым*. Очевидно, достаточно определить, что такое малое множество, а тогда большие множества определятся как дополнения к малым.

Итак, пусть дано какое-то множество последовательностей $Q \subset \Omega$. Мы хотим определить, когда его следует считать *малым*. В терминах теории вероятностей мы сказали бы, что Q малое, коль скоро вероятность попадания в Q равна нулю. В терминах теории меры мы сказали бы, что Q малое, коль скоро оно имеет меру нуль. Мы, однако, постараемся определить, что такое малое множество, в более простых терминах.

Множество Q называется *малым*, если его можно покрыть шарами, сумма объёмов которых сколь угодно мала. Определению малого множества можно дать такую формулировку: множество Q называется *малым*, если для каждого натурального числа m найдётся такая последовательность двоичных слов

$$\langle x(1), x(2), \ldots, x(n), \ldots \rangle,$$

что

$$Q \subset \bigcup_n \Omega_{x(n)},$$

$$\sum_n \mathrm{v}(x(n)) = \sum_n 2^{-|x(n)|} < \frac{1}{m}.$$

Очевидно, каждая отдельная последовательность из Ω образует малое множество, а потому понятие большого множества не может претендовать на роль уточнения понятия *разумного* большинства. Пересечение всех больших множеств пусто.

«Разумность» вводится путём следующей корректировки определения.

Во-первых, потребуем, чтобы упоминаемая в определении последовательность двоичных слов $\langle x(1), x(2), \ldots, x(n), \ldots\rangle$ была вычислимой. Другими словами, мы требуем, чтобы существовал алгоритм, вычисляющий её n-й член $x(n)$ по его номеру n.

Во-вторых, потребуем, чтобы такая вычислимая последовательность не просто существовала для каждого m, но строилась бы по этому m *эффективно*. Слово «эффективно» означает ***с помощью алгоритма***. Здесь необходимы разъяснения. Дело в том, что алгоритм, получающий на вход m и выдающий на выходе требуемую последовательность, невозможен просто потому, что последовательность — это бесконечный объект, а алгоритмы оперируют лишь с конечными объектами. Однако в нашем случае, поскольку последовательность вычислима, то у неё есть алгоритм, который её вычисляет, даже очень много таких алгоритмов. Алгоритмы (в другой системе терминов — программы алгоритмов) являются конечными объектами, и потому вполне осмысленно говорить об алгоритме, который по числу m, поступившему на его вход, выдаёт на выходе один из алгоритмов (в другой системе терминов — одну из программ), вычисляющих последовательность $\langle x(1), x(2), \ldots, x(n), \ldots\rangle$.

Внеся в определение малого множества эти два добавления, мы получаем определение *эффективно малого*[1] множества, а тем самым, и определение *эффективно большого* множества. Пересечение всех эффективно больших множеств оказывается непустым. Более того, оно само является эффективно большим. Это наименьшее среди эффективно больших множеств и есть искомое множество **T** всех *типических* последовательностей.

Типические последовательности называют также *случайными по Мартин-Лёфу* — по имени ученика Колмогорова, замечательного шведского математика Пера Мартин-Лёфа, который в 1966 г. сформулировал только что изложенное определение типичности в качестве строгого уточнения понятия случайности.

Как уже говорилось, класс **T** всех типических последовательностей совпадает с классом **C** всех хаотических последовательностей:

$$\mathbf{T} = \mathbf{C}.$$

Поэтому, как уже говорилось, типические последовательности можно именовать также *хаотическо-типическими* или *типическо-хаотическими*, а сам класс всех таких последовательностей обозначать **CT** или **TC**.

Как мы уже знаем,

$$\mathbf{CT} \subset \mathbf{S}, \quad \mathbf{CT} \neq \mathbf{S}.$$

Лицо четвёртое: Непредсказуемость

Вот начало «Двенадцати стульев» Ильфа и Петрова: «В уездном городе N было так много парикмахерских заведений и бюро похоронных процессий, что казалось, жители города рождаются лишь затем, чтобы побриться, остричься, освежить голову вежеталем и сразу же умереть». Суждение остаётся верным, если заменить N на

[1] В основном тексте такие множества называются *эффективно нулевыми*.

Москву, парикмахерские на залы игровых автоматов, похоронные бюро на казино, а цель рождения на «играть». Этот печальный факт, однако, позволяет наглядно объяснить то свойство случайных последовательностей, которое мы называем непредсказуемостью.

Интуитивно ясно, что всякая случайная последовательность является *непредсказуемой* в том смысле, что в каком бы порядке мы ни выбирали её члены, знание значений уже выбранных членов не позволяет предсказать значение того следующего члена, который мы намереваемся выбрать. Таким образом, Казино, обладающее такой последовательностью и предлагающее Игроку угадывать её члены и делать при этом денежные ставки, не разорится; говоря более точно, Казино уверено, что никакой Игрок не может обладать такой стратегией игры, которая приведёт к разорению Казино, каким бы капиталом оно ни обладало.

Непредсказуемость какой-либо последовательности, таким образом, определяется в терминах игры, которую Игрок ведёт против обладающего этой последовательностью Казино, или, для краткости, *против данной последовательности*.

Итак, представим себе, что Игрок приходит в Казино. Каждая из сторон — и Казино, и Игрок — обладает своим начальным капиталом. Казино располагает некоторой фиксированной, но неизвестной Игроку последовательностью нулей и единиц и предлагает Игроку предсказывать её члены — не обязательно в монотонном порядке их следования и даже не обязательно все её члены.

Для наглядности представим себе, что члены последовательности написаны на картах, которые лежат рубашками вверх, так что Игрок не видит, что там написано. Последовательность предстаёт перед Игроком в виде бесконечного ряда таких карт. Игра состоит в том, что Игрок на каждом своём ходу указывает ту карту, которая должна быть открыта, одновременно предсказывая значение, которое обнаружится на этой карте, и объявляя размер денежной ставки. Если предсказание окажется правильным, Казино выплачивает Игроку сумму ставки, если неправильным — Игрок выплачивает эту сумму Казино. Считается, что Игрок *выиграл*, если он сумел разорить Казино. Разумеется, если Игроку открыт неограниченный кредит, он всегда может разорить Казино, удваивая ставки. Но игра идёт на наличные, так что величина ставки ограничена текущим капиталом Игрока.

Последовательность называется *предсказуемой*, если существует выигрывающий алгоритм игры. *Выигрывающим* мы называем алгоритм со следующим свойством: каким бы начальным капиталом ни обладало Казино, оно рано или поздно будет разорено, если Игрок применит этот алгоритм. Последовательность называется *непредсказуемой*, если она не является предсказуемой.

На математическом языке ситуация описывается так. Рассматривается бесконечная последовательность нулей и единиц

$$\mathbf{a} = \langle a_1, a_2, a_3, \ldots \rangle.$$

При каждом ходе Игрока возникает тройка чисел

$$\langle n, i, v \rangle,$$

где

$$n \in \mathbb{N}, \quad i \in \{0, 1\}, \quad v \in \mathbb{Q}, \quad v \geqslant 0;$$

здесь, как обычно, $\mathbb{N}$ обозначает натуральный ряд[1], а $\mathbb{Q}$ — множество всех рациональных чисел. Содержательно: натуральное число n есть номер того члена последовательности, на который делается ставка; i есть предсказываемое значение этого члена; неотрицательное рациональное число v есть размер ставки. Ходы делаются друг за другом, начиная с первого; тройка, возникающая на k-м ходу, обозначается $\langle n(k), i(k), v(k)\rangle$. Более математически грамотно было бы сказать, что каждый ход *есть* тройка чисел и что ходы не делаются, а *предъявляются*.

Капитал Игрока перед k-м ходом обозначается $V(k-1)$. Без ограничения общности можно считать, что начальный капитал Игрока равен единице: $V(0)=1$.

После каждого хода капитал Игрока меняется (если только ставка не была нулевой). А именно:

- если $i(k)=a_{n(k)}$ (Игрок угадал), то $V(k)=V(k-1)+v(k)$;
- если $i(k)\neq a_{n(k)}$ (Игрок не угадал), то $V(k)=V(k-1)-v(k)$.

И ещё два прибавления к сказанному.

Во-первых, ходы бывают *корректные* и *некорректные*, и чтобы игра продолжалась, необходимо, чтобы ход был корректным. А именно, k-й ход считается *корректным*, если выполнены оба нижеследующие требования:

1) номер открываемой карты *корректен*: это значит, что подлежащая открытию карта не была уже открыта ранее, т. е. $n(k)$ не совпадает ни с одним из чисел $n(1),\ldots,n(k-1)$;
2) делаемая Игроком ставка *корректна*: это значит, что она меньше его текущего капитала, т. е. $v(k)<V(k-1)$.

Если же хотя бы одно из этих требований не выполнено, ход считается *некорректным.*

Основное правило остановки. Если Игрок совершает некорректный ход, игра останавливается. При этом Игрок остаётся при имеющемся у него к данному моменту капитале — и, тем самым, заведомо не выигрывает.

Во-вторых, не исключается возможность того, что Игрок вообще не делает очередного хода (даже самого первого хода!), и в этом случае Игрок также остаётся при имеющемся у него к данному моменту капитале и, как и в случае некорректного хода, заведомо не выигрывает. Однако в этом случае — в отличие от случая некорректного хода — мы избегаем выражения «игра останавливается». Дело в том, что ситуацию неделания хода можно наглядно представить себе следующим образом. Перед каждым своим ходом Игрок решает, какой ход ему следует сделать. Решение требует обдумывания, и Игрок берёт время на обдумывание. Время это ничем не ограничено, и процесс размышления может затянуться до бесконечности. В течение времени обдумывания хода капитал Игрока не меняется. Поэтому если Игрок ни за какое конечное время не принимает решения о своём ходе, его капитал застывает. В этом случае выигрыша Игрока произойти не может. Однако в этом случае (в отличие от случая некорректного хода) не наступает момента остановки игры,

[1] Натуральный ряд иногда начинают с единицы, а иногда (в частности, в большинстве исследований по теории вычислимых функций) с нуля. В данном случае мы не включаем нуль в натуральный ряд.

ибо никогда не поступает сигнала об остановке. Таким образом, у игры есть три возможных сценария развития: 1) Игрок делает бесконечное число ходов; 2) Игрок делает лишь конечное число ходов, и причиной этого служит то, что был сделан некорректный ход; 3) Игрок делает лишь конечное число ходов, и причиной этого служит то, что на каком-то этапе игры Игрок не в состоянии прийти к решению об очередном ходе. Разумеется, сказанное носит иллюстративный характер, и математическое описание не включает в себя ссылку на такие понятия, как «решать», «обдумывать» и т. п.

По определению, Игрок *выигрывает* (при игре против **a**), если

$$\sup_k V(k) = +\infty,$$

т. е. если

$$\forall W \ \exists k \ \ V(k) > W.$$

Содержательно это означает, что Игрок разоряет Казино, каким бы исходным капиталом W Казино ни обладало. Очевидно, что Игрок в состоянии выиграть лишь при условии, что всякий раз, когда ему предстоит делать ход, он его делает и этот ход оказывается корректным.

Как выглядит игра, мы описали. Перейдём теперь к понятию системы игры, или *стратегии*. Смысл стратегии в том, чтобы избавить Игрока от необходимости самостоятельно принимать решения: стратегия берёт эту функцию на себя. Стратегия есть правило, для каждого хода указывающее Игроку, какой на этом ходу он должен сделать ход (т. е. какую тройку предъявить). Разумеется, стратегия выдаёт такое указание лишь в том случае, если ход должен быть сделан. Выше уже отмечалась возможность того, что никакого хода не делается; в этом случае, естественно, стратегия не выдаёт никакого указания. При указании хода стратегия опирается на всю предшествующую историю игры. История же игры состоит из всех уже сделанных к рассматриваемому моменту ходов и из всех ставших уже известными членов последовательности. Мы лишь потому не включаем в историю игры информацию о капитале Игрока на каждый момент, что эта информация легко вычисляется из только что перечисленных сведений.

Таким образом, историю игры перед k-м ходом можно записать в виде таблицы:

$$\begin{array}{cccc}
n(1) & n(2) & \ldots & n(k-1) \\
i(1) & i(2) & \ldots & i(k-1) \\
v(1) & v(2) & \ldots & v(k-1) \\
a_{n(1)} & a_{n(2)} & \ldots & a_{n(k-1)}
\end{array}$$

(если $k = 1$, таблица пуста).

На математическом языке стратегия есть функция, которая каждой подобной таблице (в том числе пустой) либо ничего не ставит в соответствие, либо ставит в соответствие некоторый ход, т. е. тройку $\langle n, i, v \rangle$. Под «каждой подобной таблицей» мы понимаем отнюдь не только такую таблицу, которая отражает реальное течение игры, а произвольную таблицу, в которой в первой строке стоят положительные целые числа, в третьей — неотрицательные рациональные числа, а во второй и четвёртой — нули и единицы.

Если таблица реально встретилась в процессе игры (в качестве истории игры на каком-то этапе) и если задана стратегия, то первые три строки таблицы однозначно восстанавливаются по её четвёртой строке. В самом деле, применение стратегии к пустой таблице даёт нам первый ход $\langle n(1), i(1), v(1)\rangle$. Тем самым — поскольку четвёртая строка предполагается известной — мы получаем историю игры перед вторым ходом в виде таблицы

$$\begin{array}{c} n(1) \\ i(1) \\ v(1) \\ a_{n(1)}. \end{array}$$

Теперь к этой таблице снова применяем стратегию, которая указывает второй ход $\langle n(2), i(2), v(2)\rangle$; получаем таблицу

$$\begin{array}{cc} n(1) & n(2) \\ i(1) & i(2) \\ v(1) & v(2) \\ a_{n(1)} & a_{n(2)}, \end{array}$$

и так далее.

Сказанное даёт нам право при определении стратегии брать в качестве аргумента не всю таблицу в целом, а лишь её четвёртую строку. Заметим, что в этой четвёртой строке стоит двоичное слово, т. е. элемент множества Ξ. Стратегия должна, имея этот элемент, или не выдавать ничего, или выдавать ход, который есть тройка, т. е. элемент декартова произведения $\mathbb{N} \times \{0,1\} \times \mathbb{Q}_+$. Здесь символом $\mathbb{Q}_+$ обозначено множество всех неотрицательных рациональных чисел.

Мы приходим к окончательному определению понятия стратегии: *стратегия* есть отображение некоторого подмножества множества Ξ всех двоичных слов в декартово произведение $\mathbb{N} \times \{0,1\} \times \mathbb{Q}_+$:

$$\Xi \to \mathbb{N} \times \{0,1\} \times \mathbb{Q}_+.$$

Для наших целей особый интерес представляет случай, когда указанное отображение задаётся каким-то алгоритмом. Поясним, что это значит. Пусть **A** — алгоритм, на вход которого могут подаваться элементы из множества X, а на выходе получаются элементы из множества Y. В множестве X выделяется *область результативности* алгоритма **A**, состоящая из тех и только тех элементов, в применении к которым **A** даёт результат. Алгоритм **A** *задаёт* следующее отображение: область определения этого отображения совпадает с областью результативности алгоритма, и для каждого элемента этой области значение отображения на этом элементе совпадает с тем результатом, который получается при применении алгоритма к этому элементу.

Если стратегия задана каким-то алгоритмом, она называется *вычислимой.* (Если подать на вход задающего стратегию алгоритма такую историю игры, для которой очередной ход не определён, то алгоритм работает на этом входе бесконечно долго, не приходя ни к какому результату, но и не выдавая сообщение об отсутствии

такового. Эта бесконечная работа алгоритма как раз и происходит за то взятое Игроком на обдумывание бесконечное время, о котором говорилось выше.)

Стратегия называется *выигрывающей для последовательности* **a**, если Игрок, применяющий эту стратегию в игре против **a**, выигрывает.

Последовательность называется *предсказуемой*, если для неё существует выигрывающая вычислимая стратегия, и *непредсказуемой* — в противном случае. Класс всех непредсказуемых последовательностей обозначается буквой **U**.[1]

Известно, что всякая непредсказуемая последовательность является стохастической по Колмогорову (принадлежит классу **S**) и что всякая типическо-хаотическая последовательность (последовательность класса **CT**) непредсказуема:

$$\mathbf{CT} \subset \mathbf{U} \subset \mathbf{S}.$$

Известно также, что класс стохастических по Колмогорову последовательностей существенно шире класса непредсказуемых:

$$\mathbf{S} \neq \mathbf{U}.$$

Открытым остаётся вопрос о совпадении классов хаотических (он же класс типических) и непредсказуемых последовательностей:

$$\mathbf{CT} \stackrel{?}{=} \mathbf{U}.$$

Эта важная проблема ждёт своего решения.

О безостановочных стратегиях

Если игра никогда не останавливается, она называется *безостановочной*. Стратегия называется *безостановочной*, если, какова бы ни была последовательность, применение против неё этой стратегии приводит к безостановочной игре. В определении предсказуемости можно ограничиться безостановочными стратегиями и дать такую равносильную формулировку: последовательность называется *предсказуемой*, если для неё существует выигрывающая вычислимая безостановочная стратегия. Чтобы убедиться в равносильности, достаточно объяснить, как можно алгоритм **A**, задающий выигрывающую стратегию, переделать в алгоритм **B**, задающий выигрывающую безостановочную стратегию. Такая переделка осуществляется весьма просто. Сперва по поступившему на вход алгоритма **A** двоичному слову восстанавливается история игры, что позволяет знать как номера всех открытых ранее карт, так и величину текущего капитала Игрока. Затем всякий ход, предписываемый исходным алгоритмом **A**, проверяется на корректность, и если он оказывается некорректным, то новый алгоритм **B** никакого хода не выдаёт, а объявляется не определённым на указанном двоичном слове.

[1] В основном тексте такие последовательности упоминались как «случайные относительно немонотонных вычислимых мартингалов», см. обсуждение на с. 347.

Обобщение на вычислимые распределения вероятностей

До сих пор, для простоты и большей наглядности, мы ограничивались ситуацией, когда на пространстве Ω всех двоичных последовательностей было задано *равномерное распределение вероятностей.* Все основные идеи были видны на примере этой простейшей ситуации. Для полноты картины мы намерены рассмотреть теперь общую ситуацию, когда на пространстве Ω задано произвольное *вычислимое распределение вероятностей.* Что это такое, будет разъяснено позже. А сейчас мы хотим сказать несколько слов, направленных на то, чтобы сделать изложение доступным и такому Читателю, который не знаком с общим понятием *распределения вероятностей*, или *вероятностной меры.*

Говорят, что на множестве M задана *мера* μ, коль скоро, во-первых, выделен некоторый класс подмножеств множества M, называемых *измеримыми*, и, во-вторых, каждому измеримому подмножеству A отнесено неотрицательное число $\mu(A)$, называемое *мерой* этого подмножества. При этом требуется, чтобы выполнялись некоторые аксиомы, приводить которые мы здесь не будем; отметим только, что следствием этих аксиом является такой факт: объединение непересекающихся измеримых множеств измеримо и его мера равна сумме мер этих множеств. Если $\mu(M) = 1$, мера называется *вероятностной*, или *распределением вероятностей.* В этом случае $\mu(A)$ содержательно трактуется как вероятность того, что случайно выбранный элемент множества M попадает в A.

Каждая мера на Ω характеризуется мерами шаров. Так, для равномерного распределения (и только для него)

$$\forall x \in \Xi \text{ выполнено } \mu(\Omega_x) = 2^{-|x|}.$$

Равномерное распределение вероятностей на пространстве Ω отвечает тому сценарию, когда нули и единицы возникают в последовательности с равными вероятностями. Ближайшим обобщением равномерного распределения является *распределение Бернулли*, или *бернуллиевское распределение* (называемое также *биномиальным*), при котором нули и единицы возникают независимо с вероятностью p для единицы и вероятностью $1 - p$ для нуля; это число p условимся называть *параметром* бернуллиевского распределения. Если параметр равен одной второй, получаем равномерное распределение. Говоря формально, распределение Бернулли с параметром p задаётся формулой

$$\mu(\Omega_x) = p^k (1 - p)^{|x|-k},$$

где k — количество единиц в слове x.

Следующим обобщением служит класс распределений, которые мы будем называть *квазибернуллиевскими.* На содержательном уровне квазибернуллиевость распределения вероятностей означает, что единицы и нули появляются в качестве членов последовательности по-прежнему независимо, но с переменной вероятностью, зависящей от номера члена. Вероятность того, что член с номером k есть единица, условимся обозначать через $p(k)$; тогда вероятность того, что этот член есть

нуль, окажется равной $1 - p(k)$. Формальное определение квазибернуллиевского распределения таково. Пусть дана последовательность действительных чисел

$$\mathbf{p} = \langle p(1), p(2), \ldots, p(k), \ldots \rangle, \quad 0 \leqslant p(k) \leqslant 1.$$

Про распределение μ будем говорить, что оно *квазибернуллиевское с параметром* $\mathbf{p}$, если для всякого двоичного слова $x = \langle x_1, \ldots, x_n \rangle$ имеет место равенство

$$\mu(\Omega_x) = \prod_{i=1}^{n} r_i,$$

где $r_i = p(i)$ при $x_i = 1$ и $r_i = 1 - p(i)$ при $x_i = 0$. Для ясности: бернуллиевское распределение с параметром p является квазибернуллиевским с параметром $\mathbf{p} = \langle p, p, \ldots, p, \ldots \rangle$.

В этой главке мы укажем, как определения стохастичности, хаотичности, типичности и непредсказуемости должны быть расширены для общей ситуации *вычислимого распределения вероятностей* (что это такое, будет разъяснено ниже). Объявим наперёд, что при этих расширенных определениях для произвольного вычислимого распределения μ сохраняются те же соотношения, которые были выписаны выше для случая равномерного распределения:

$$\mathbf{C}(\mu) = \mathbf{T}(\mu) \subset \mathbf{U}(\mu) \subset \mathbf{S}(\mu),$$
$$\mathbf{S}(\mu) \neq \mathbf{U}(\mu)$$

(последнее выполнено при условии, что мера любого шара положительна).

Здесь через $\mathbf{C}(\mu)$, $\mathbf{T}(\mu)$, $\mathbf{U}(\mu)$, $\mathbf{S}(\mu)$ соответственно обозначены те классы последовательностей, хаотических, типических, непредсказуемых, стохастических по Колмогорову относительно распределения μ, которые будут определены ниже. (В наших прежних обозначениях $\mathbf{C} = \mathbf{C}(\eta)$, $\mathbf{T} = \mathbf{T}(\eta)$, $\mathbf{U} = \mathbf{U}(\eta)$ и $\mathbf{S} = \mathbf{S}(\eta)$, где η есть равномерное распределение.)

Эта главка — для любителей обобщений, и мы призываем уважаемого Читателя подумать, стоит ли её читать. Во-первых, она несколько труднее предыдущих главок. Во-вторых, сама задача о формулировке строгого определения случайной последовательности делается тем менее ясной, чем более обширным становится класс рассматриваемых распределений. Ведь само представление о случайности индивидуальной последовательности имеет сколько-нибудь ясный интуитивный смысл лишь для простейших распределений вероятностей, таких, как равномерное и его ближайшие обобщения.

Вычислимые меры и вычислимые распределения

Хотелось бы называть меру μ на пространстве Ω вычислимой, если существует такой алгоритм, который по каждому двоичному слову x даёт меру $\mu(\Omega_x)$ шара Ω_x. Однако алгоритмы не могут иметь дело с действительными числами, а только с числами целыми, рациональными и т. п. — строго говоря, даже не с числами, а с их именами в виде слов в каком-либо фиксированном конечном алфавите (*словом*

в данном алфавите называется любая конечная цепочка, составленная из букв этого алфавита). Дать же имена всем действительным числам невозможно, поскольку для каждого конечного алфавита все слова в нём образуют лишь счётное множество. Поэтому можно требовать лишь наличие алгоритма, дающего не самоё меру шара, а сколь угодно точное рациональное приближение к ней.

Окончательное определение таково. Меру μ называют *вычислимой*, коль скоро существует алгоритм, который для каждого двоичного слова x и каждого положительного рационального числа ε выдаёт такое рациональное число, которое отличается от меры $\mu(\Omega_x)$ шара Ω_x не более, чем на ε.

Иногда в определение вычислимости включают ещё и требование о существовании такого алгоритма, который для каждого двоичного слова x определяет, имеет ли место равенство $\mu(\Omega_x) = 0$. Вычислимые меры, удовлетворяющие этому дополнительному требованию (которое мы не включаем в определение вычислимости), будем называть *сильно вычислимыми*. Заметим, что из вычислимости меры не вытекает, как могло бы показаться, её сильная вычислимость.

Важным подклассом класса сильно вычислимых мер является класс вычислимо-рациональных мер. Будем называть меру *вычислимо-рациональной*, если мера каждого шара есть рациональное число и существует алгоритм, вычисляющий эту меру по заданному шару, т. е., на более точном языке, алгоритм, который для каждого двоичного слова x выдаёт дробь, выражающую меру $\mu(\Omega_x)$ шара Ω_x. Заметим для ясности, что если мера вычислима, а её значения на шарах рациональны, то отсюда ещё не следует, что она вычислимо-рациональна: умение находить для данного рационального числа сколь угодно точные рациональные приближения к нему ещё не означает умения выписать само это число в виде отношения двух целых чисел!

Теперь нет нужды отдельно определять, что такое вычислимое распределение вероятностей: это просто-напросто вычислимая вероятностная мера. Именно на вычислимые распределения естественно обобщаются многие определения и факты, изложенные нами для равномерного распределения. В частности, для произвольного вычислимого распределения верны как *теорема Мартин-Лёфа*, утверждающая, что пересечение всех эффективно больших множеств само является эффективно большим, так и *теорема Левина*, утверждающая, что типичность равносильна хаотичности, определённой на основе монотонной энтропии.[1]

Стохастичность

Для какой-либо последовательности **e** её k-й член обозначаем e_k или, дабы избежать многоэтажных индексов, $e(k)$.

В случае равномерного распределения стохастичность последовательности понималась как *глобальная устойчивость частот*, т. е. как устойчивость частот в *ка-*

[1]В основном тексте книги она называется *теоремой Левина – Шнорра*; в работе Шнорра, появившейся раньше, рассматривается некоторая особая разновидность сложности, которую автор назвал "процессной". Числовые значения монотонной сложности и процессной сложности могут различаться достаточно сильно (уточнение дано ниже, в разделе "История и библиография"), процессную сложность можно рассматривать как особый вариант монотонной сложности и доказательства Левина и Шнорра сходны.

ждой из допустимых подпоследовательностей. Допустимые же подпоследовательности возникали путём применения допустимых по Колмогорову правил отбора. В ситуации произвольной вероятностной меры общая схема сохраняется, только устойчивость частот заменяется на некоторое более общее свойство, а именно на так называемый *закон больших чисел.*

Определение стохастичности для бернуллиевского распределения с параметром p очевидно: надо потребовать, чтобы в каждой допустимой подпоследовательности соблюдалась устойчивость частот с одной и той же предельной частотой p. Иными словами, доля единиц в начальном отрезке всякой допустимой подпоследовательности должна, при неограниченном удлинении отрезка, стремиться к p. Случаи, когда $p = 0$ или $p = 1$, являются вырожденными; для первого из них не считается стохастической последовательность, в которой встречается хотя бы одна единица, для второго — последовательность, в которой встречается хотя бы один ноль.

Фон Мизес понимал вероятность как предельную частоту. Поэтому его идеология не простирается за пределы бернуллиевских распределений. Мостом к общему случаю произвольного распределения служат квазибернуллиевские распределения. Ими и займёмся.

Ясно, что для квазибернуллиевского распределения предельной частоты в подпоследовательности может и не быть, а если она и есть, то, вообще говоря, в каждой подпоследовательности — своя. Приходится поэтому говорить о стохастичности относительно данного правила отбора. Пусть $\mathbf{p}$ — параметр квазибернуллиевского распределения. Прежде всего, заявим, что не считается стохастической последовательность, в которой значение хотя бы одного её члена имеет нулевую вероятность. Или, говоря на формальном языке: последовательность $\mathbf{a}$ не считается стохастической, если существует такое k, что либо $a(k) = 0$ и $p(k) = 1$, либо $a(k) = 1$ и $p(k) = 0$. Для остальных последовательностей определение таково: последовательность $\mathbf{a}$ будем называть *стохастической относительно правила отбора* Θ, коль скоро её обобщённая подпоследовательность

$$\mathbf{b} = \langle a(m_1), a(m_2), \ldots, a(m_k), \ldots \rangle,$$

полученная по этому правилу, удовлетворяет требованию закона больших чисел:

$$\frac{a(m_1) + \ldots + a(m_k)}{k} - \frac{p(m_1) + \ldots + p(m_k)}{k} \to 0$$

при $k \to \infty$. Далее, последовательность $\mathbf{a}$ называется *стохастической* (*по Колмогорову*), если для любого колмогоровского правила, приводящего к бесконечной обобщённой подпоследовательности, она является стохастической относительно этого правила.

Замечание. Слово «бесконечной» в предыдущей фразе избыточно, так как всякая обобщённая подпоследовательность является последовательностью, а следовательно, бесконечна по определению. Тем не менее мы будем иногда использовать это формально избыточное упоминание о бесконечности, дабы подчеркнуть, что имеется в виду именно последовательность, а не кортеж.

Чтобы объяснить, как понятие стохастичности по Колмогорову обобщается на более широкий класс вероятностных распределений, введём несколько обозначений.

Пусть $n(1), n(2), \ldots, n(k)$ суть натуральные числа и пусть

$$i(1), i(2), \ldots, i(k) \in \{0, 1\}.$$

Через $A^{n(1), \ldots, n(k)}_{i(1), \ldots, i(k)}$ обозначим множество всех таких последовательностей $\mathbf{a} \in \Omega$, у которых

$$a_{n(1)} = i(1), \;\; a_{n(2)} = i(2), \;\; \ldots, \;\; a_{n(k)} = i(k). \qquad (*)$$

Дробь

$$\frac{\mu\left(A^{n(1), \ldots, n(k), m}_{i(1), \ldots, i(k), 1}\right)}{\mu\left(A^{n(1), \ldots, n(k)}_{i(1), \ldots, i(k)}\right)}$$

обозначим символом

$$\mu\left(\begin{array}{c|c} m & n(1), \;\ldots,\; n(k) \\ 1 & i(1), \;\ldots,\; i(k) \end{array}\right);$$

это есть условная вероятность того, что m-й член последовательности $\mathbf{a}$ будет равен 1 при условии $(*)$; эта величина не определена, коль скоро знаменатель дроби обращается в нуль.

Рассмотрим два произвольных, но фиксированных объекта: последовательность $\mathbf{a} \in \Omega$ и допустимое по Колмогорову правило отбора Θ. Наша цель — придать смысл высказыванию «последовательность $\mathbf{a}$ стохастична относительно правила Θ». Процесс, посредством которого правило Θ отбирает из последовательности $\mathbf{a}$ члены обобщённой подпоследовательности $\mathbf{b}$, состоит из двух этапов. На первом этапе строится вспомогательная обобщённая подпоследовательность $\mathbf{c}$, из некоторых членов которой на втором этапе и строится $\mathbf{b}$. Более подробно,

$$\mathbf{c} = \langle a_{n(1)}, a_{n(2)}, \ldots, a_{n(k)}, \ldots \rangle,$$

где номер $n(k)$ вычисляется алгоритмически по кортежу

$$\langle a_{n(1)}, a_{n(2)}, \ldots, a_{n(k-1)} \rangle.$$

Имея на входе этот же кортеж, правило Θ решает, включать или нет член $a_{n(k)}$ в окончательную последовательность $\mathbf{b}$. Таким образом,

$$\mathbf{b} = \langle a(n(k_1)), a(n(k_2)), \ldots, a(n(k_j)), \ldots \rangle.$$

На каждом из обоих этапов может наступить момент, когда правило Θ не даст никакого результата, т. е. не выдаст номера на первом этапе или не примет решения на втором этапе. Если такое произойдёт, последовательность $\mathbf{b}$ окажется конечной, т. е., говоря более строго, окажется не последовательностью, а кортежем. В этом случае никакого требования к $\mathbf{b}$ не выдвигается и, говоря формально, $\mathbf{a}$ признаётся стохастической относительно Θ. Если же $\mathbf{b}$ оказалось бесконечной, то для признания последовательности $\mathbf{a}$ стохастической относительно Θ требуется выполнение нижеследующего свойства последовательности $\mathbf{a}$.

Обозначим через r_j величину

$$\mu\left(\begin{matrix} n(k_j) \\ 1 \end{matrix}\;\middle|\; \begin{matrix} n(1), & n(2), & \ldots, & n(k_j-1) \\ a(n(1)), & a(n(2)), & \ldots, & a(n(k_j-1)) \end{matrix}\right).$$

Рассмотрим разность

$$\delta_j = \frac{r_1 + r_2 + \ldots + r_j}{j} - \frac{a(n(k_1)) + a(n(k_2)) + \ldots + a(n(k_j))}{j}.$$

Величина δ_j не определена, если не определена хотя бы одна из величин $r_1, \ldots, r_j$. Скажем, что $\mathbf{b}$ *подчиняется закону больших чисел*, если все величины δ_j определены и $\delta_j \to 0$ при $j \to \infty$. Последовательность $\mathbf{a}$ назовём *стохастической относительно правила* Θ, если обобщённая подпоследовательность $\mathbf{b}$, полученная из $\mathbf{a}$ согласно правилу Θ, подчиняется закону больших чисел.

Наконец, последовательность $\mathbf{a}$ называется *стохастической* (*по Колмогорову*), если для любого допустимого по Колмогорову правила отбора она является стохастической относительно этого правила — при условии, что это правило строит бесконечную обобщённую подпоследовательность (а не конечный кортеж).

Заметим, что в самом определении вычислимость меры не используется. Однако без предположения о вычислимости едва ли возможно сравнивать частотоустойчивость (она же стохастичность) с другими алгоритмическими лицами случайности.

Хаотичность

Определение последовательности, хаотической относительно меры μ, таково: последовательность $\mathbf{a} = \langle a_1, a_2, a_3, \ldots\rangle$ называется *хаотической относительно* μ, если существует такая константа c, что для всех n выполняется неравенство

$$KM(a_1, a_2, \ldots, a_n) > -\log \mu(\Omega_{a_1, a_2, \ldots, a_n}) - c,$$

где, как всегда, знак логарифма без нижнего индекса означает логарифм по основанию два.

Мотивировка этого определения, имеющего содержательный смысл для вычислимых вероятностных мер, такая же, как и в случае равномерного распределения вероятностей. Дело в том, что для всякой вычислимой вероятностной меры выполняется соотношение:

$$\exists c\ \forall x \in \Xi \text{ выполнено } KM(x) < -\log \mu(\Omega_x) + c.$$

Говоря неформально, для любой вычислимой меры μ можно подобрать язык описания, приспособленный к этой мере в том смысле, что снабжает слова x с бо́льшим $\mu(\Omega_x)$ короткими описаниями (и написанное соотношение указывает длину этих коротких описаний). Определение хаотичности требует, чтобы полученные таким способом короткие описания нельзя было бы ещё укоротить (более чем на константу).

Типичность

Понятие типичности обобщается на произвольную меру достаточно очевидным образом: надо просто в том определении типической последовательности, которое давалось для случая равномерного распределения, заменить объём шара на его меру.

Сперва определяется понятие эффективно малого множества. Множество $Q \subset \Omega$ называется *эффективно малым относительно меры* μ, коль скоро существует алгоритм $\mathbf{A}$, удовлетворяющий нижеследующему требованию. При поступлении на вход алгоритма $\mathbf{A}$ натурального числа m на его выходе возникает алгоритм построения такой последовательности двоичных слов $\langle x(1), x(2), \ldots, x(n), \ldots \rangle$ (т. е. алгоритм вычисления такой функции $n \mapsto x(n)$), что

$$Q \subset \bigcup_n \Omega_{x(n)},$$

$$\sum_n \mu(\Omega_{x(n)}) < \frac{1}{m}.$$

Далее, *эффективно большое относительно меры* μ множество определяется как дополнение (до Ω) к эффективно малому.

Для вычислимых мер имеет место *теорема Мартин-Лёфа*: объединение всех эффективно малых множеств является эффективно малым, а пересечение всех эффективно больших — эффективно большим. Существующее согласно этой теореме наименьшее эффективно большое множество называют *конструктивным носителем меры* μ. В случае, если μ — вычислимое распределение вероятностей, последовательности, принадлежащие конструктивному носителю распределения μ, и называют *типическими* относительно этого распределения. Таким образом, $\mathbf{T}(\mu)$ есть не что иное как конструктивный носитель распределения μ.

Непредсказуемость

Здесь мы укажем, что́ нужно добавить к определениям, сформулированным для равномерного случая, при переходе к произвольным распределениям. Таких добавлений будет два: некий мультипликативный коэффициент (для равномерного распределения он равен единице и потому не нужен) и дополнительное правило остановки (в случае равномерного распределения оно потому не нужно, что не возникает такой ситуации, при которой оно могло бы быть применено).

Распределение вероятностей влияет на правило изменения капитала Игрока после очередного хода. Если Игрок не угадал, его капитал уменьшается на сумму сделанной им ставки — так же, как и в случае равномерного распределения. Но если Игрок угадал, то прирост его капитала равняется ставке, умноженной на некоторый коэффициент. Этот коэффициент сравнительно велик, если вероятность угадать была низка, и сравнительно мал, если вероятность угадать была высока. Для равномерного распределения вероятность угадать всегда равна одной второй, а коэффициент всегда равен единице. Точная формулировка правила прироста капитала будет сейчас изложена.

Для последовательности $\mathbf{a}$ её k-й член обозначаем a_k, для последовательности $\mathbf{a}'$ её k-й член обозначаем a'_k и т. д.

Ход, делаемый на j-м ходу, есть тройка $\langle n(j), i(j), v(j)\rangle$.

Пусть последовательность, которой располагает Казино, есть $\mathbf{a}$. Положим

$$A(k-1) = \{\mathbf{a}' \in \Omega\colon\ a'_{n(j)} = a_{n(j)} \text{ при } j = 1, 2 \ldots, k-1\}$$

(так что $A(0) = \Omega$),

$$A_i(k) = \{\mathbf{a}' \in A(k-1)\colon\ a'_{n(k)} = i\} \text{ для } i = 0, 1.$$

Разумеется, эти обозначения имеют смысл в предположении, что все участвующие в их определениях номера $n(l)$ определены. Полезно заметить, что

$$\Omega = A(0) \supset A(1) \supset A(2) \supset \ldots, \tag{1}$$

$$1 = \mu(A(0)) \geqslant \mu(A(1)) \geqslant \mu(A(2)) \geqslant \ldots; \tag{2}$$

если Игрок на k-м ходу угадал, то

$$i(k) = a_{n(k)}, \quad A_{i(k)}(k) = A(k); \tag{3}$$

если Игрок на k-м ходу не угадал, то

$$i(k) \neq a_{n(k)}, \quad A_{1-i(k)}(k) = A(k); \tag{4}$$

кроме того,

$$A(k-1) = A_0(k) \cup A_1(k). \tag{5}$$

В случае, если $i(k) = a_{n(k)}$ (прогноз Игрока на его k-м ходе оправдался — он угадал), капитал Игрока увеличивается по формуле

$$V(k) = V(k-1) + v(k) \cdot \frac{\mu(A_{1-i(k)}(k))}{\mu(A_{i(k)}(k))}. \tag{6}$$

Наша формула (6) гарантирует «честность» игры: математическое ожидание выигрыша, т. е. прироста капитала, за один ход равно нулю. Однако на пути применения этой формулы нас подстерегает неприятность, невозможная для равномерного распределения, как и для любой позитивной меры (мера называется *позитивной*, если мера любого шара положительна). Неприятность заключается в том, что стоящая в знаменателе вероятность $\mu(A_{i(k)}(k))$ может оказаться равной нулю. Эта неприятность устраняется путём введения Дополнительного правила остановки, которое мы сейчас приведём. В случае равномерного распределения это правило было излишним, поскольку предусмотренная в нём ситуация не могла встретиться.

Дополнительное правило остановки. Оно применяется тогда, когда в ходе игры *впервые* наступает такая ситуация, что $\mu(A(k)) = 0$ (просим Читателя взглянуть на формулу (2)). Пусть $\mu(A(k-1)) \neq 0$, $\mu(A(k)) = 0$. Последний ход, который был сделан, имел номер k. Делая этот ход, Игрок объявил свой прогноз $i(k)$.

Если прогноз оказался верным, т. е. оказалось, что $i(k) = a_{n(k)}$, игра останавливается, а капитал Игрока считается возросшим до бесконечности: $V(k) = +\infty$. Таким образом, в этом случае Игрок объявляется выигравшим. Если же прогноз оказался неверным, т. е. оказалось, что $i(k) \neq a_{n(k)}$ (или, что то же самое, $1 - i(k) = a_{n(k)}$), игра опять-таки останавливается, но с тем капиталом Игрока, который у него был к этому моменту, так что величина $V(k)$ оказывается равной величине $V(k-1)$. В этом случае, следовательно, Игрок не выигрывает.

Ясно, что проблема с нулевым знаменателем в формуле (6) устраняется этим правилом. Действительно, формула (6) применяется лишь в случае, когда $i(k) = a_{n(k)}$. В этом случае, согласно (3), $A_{i(k)}(k) = A(k)$. Поэтому интересующий нас знаменатель может оказаться нулевым лишь в ситуации, когда $\mu(A(k)) = 0$. Но именно эта ситуация как раз и регулируется не формулой (6), а нашим Дополнительным правилом. (Можно, впрочем, считать, что и формулой (6), если разрешить деление на нуль положительного числа $\mu(A_{1-i(k)}(k))$ и принять бесконечность в качестве результата такого деления.)

Определения стратегии, а также вычислимой стратегии, безостановочной стратегии, выигрывающей стратегии остаются, с учётом сделанных добавлений, для общего случая теми же, как и для частного случая равномерного распределения.

История и библиография

[1] А. Н. Колмогоров, В. А. Успенский. *Алгоритмы и случайность // Теория вероятностей и её применения*, 1987. Т. 32, вып. 3, с. 425–455.

[2] В. А. Успенский, А. Л. Семёнов. *Теория алгоритмов: основные открытия и приложения.* — М.: Физматлит, 1987. — 288 с.

[3] В. А. Успенский, А. Л. Семёнов, А. Х. Шень. *Может ли индивидуальная последовательность нулей и единиц быть случайной? // Успехи математических наук*, 1990. Т. 45, вып. 1, с. 105–162.

[4] V. A. Uspensky, A. Shen. *Relations between varieties of Kolmogorov complexities // Mathematical Systems Theory*, 1996. Vol. 29, no. 3, p. 271–292.

[5] An. A. Muchnik, A. L. Semenov, V. A. Uspensky. *Mathematical metaphysics of randomness // Theoretical Computer Science*, 1998. Vol. 207, p. 263–317.

[6] А. Шень. *О соотношениях между различными алгоритмическими определениями случайности // Доклады Академии наук СССР*, 1988. Т. 302, № 3, с. 548–552.

[7] В. В. Вьюгин. *Алгоритмическая энтропия (сложность) конечных объектов и её применение к определению случайности и количества информации // Семиотика и информатика. Выпуск 16.* — М: ВИНИТИ, 1981. — С. 14–43.

[8] M. Li, P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* — New York e. a.: Springer-Verlag, 1993. — xx+546 pp., 38 illustrations;

Second ed. — New York e. a.: Springer-Verlag, 1997. — xx+637pp., 41 illustrations.

Сам этот список из восьми названий никоим образом не претендует на полноту. Однако в этих публикациях (особенно в [8]) можно найти дальнейшие ссылки, и вкупе с этими ссылками некое подобие полноты уже достигается.

В книге [2] следует обратить внимание на § 2.6 «Приложения к теории вероятностей: определения случайной последовательности». Следует также иметь в виду, что терминология этой книги несколько архаична: хаотические последовательности называются там *случайными по Колмогорову*, а типические — *случайными по Мартин-Лёфу*. Частотоустойчивые, они же стохастические, последовательности называются в [2] *случайными по Мизесу*, стохастические по Чёрчу — *случайными по Мизесу – Чёрчу*, стохастические по Колмогорову — *случайными по Мизесу – Колмогорову – Лавлэнду* (а в [3] — *стохастическими по Колмогорову – Лавлэнду*; Д. Лавлэнд (D. Loveland) открыл эти последовательности независимо от Колмогорова, хотя и позже: соответствующая статья Колмогорова была опубликована в 1963 г., статья Лавлэнда — в 1966 г.). Что касается непредсказуемых последовательностей — в том точном понимании, как было сформулировано выше, — то они в указанной книге отсутствуют, поскольку появились в литературе лишь в 1998 г. в статье [5].

Пример стохастической по Чёрчу последовательности, перестающей быть таковой после подходяще выбранной вычислимой перестановки её членов, опубликовал в 1966 г. Д. Лавлэнд. Значение этого примера состоит не только в том, что он показывает неадекватность определения Чёрча, но и в том, что он открывает новое качество случайности, интуитивно очевидное, но ранее не замеченное: случайная последовательность должна оставаться случайной после любой вычислимой (т. е. задаваемой каким-то алгоритмом) перестановки её членов.

Теория сложности объектов (в дополнение к уже развивавшейся в то время теории сложности вычислений) была основана Колмогоровым на его семинарах в Московском университете в начале 60-х годов XX века и имела своей главной целью перестройку понятийной базы теории информации (на основе представления о том, что чем сложнее объект, тем больше информации он содержит) с последующим приложением к теории случайности. В опубликованной в 1969 г. (английский вариант в 1968 г.) статье Колмогоров писал:

> 1) основные понятия теории информации должны и могут быть обоснованы без помощи обращения к теории вероятностей и так, что понятия «энтропия» и «количество информации» оказываются применимы к индивидуальным объектам;
> 2) введённые таким образом понятия теории информации могут лечь в основу новой концепции случайного, соответствующей естественной мысли о том, что случайность есть отсутствие закономерности.

Предложение измерять сложность объекта длиной его кратчайшего описания и понятие оптимального языка были изложены Колмогоровым в статье 1965 г.; за год до того сходные идеи были опубликованы американским исследователем Рэем Соломоновым (Ray Solomonoff), о работах которого Колмогоров узнал лишь позже. Ввиду этого теорему о существовании оптимального языка мы называем *теоре-*

мой Соломонова – Колмогорова. В эти же годы (т. е. в середине 60-х годов XX в.) Колмогоров высказывает на своих семинарах предположение о том, что быстрота роста энтропии начальных отрезков последовательности может служить критерием случайности рассматриваемой последовательности. Однако введённое им в рассмотрение языковóе семейство приводило к энтропии, непригодной для поставленной цели. Как уже говорилось выше, в главке о хаотичности, годное для этой цели семейство обнаружил в 1973 г. Леонид Левин, введя в рассмотрение монотонную энтропию.

Типические последовательности (под названием “random”, то есть «случайные») были, как уже отмечалось в главке о типичности, открыты в 1966 г. Пером Мартин-Лёфом (Per Martin-Löf).

Утверждение о том, что класс последовательностей, стохастических по Колмогорову, шире класса типических (они же — хаотические) последовательностей, доказал Александр Шень (см. [6], а также [2, п. 6.2.4])[1].

Пусть K — энтропия в каком-то из вариантов этого понятия. (В литературе исследованы не менее шести таких вариантов: простая, априорная, монотонная, процессная, префиксная и энтропия разрешения; все эти варианты существенно различны в том точном смысле, что разность энтропий, принадлежащих любым двум из перечисленных вариантов, не ограничена.) Скажем, что последовательность $a_1, a_2, a_3, \dots, a_n, \dots$ является *хаотической относительно K* (при равномерном распределении!), если

$$\exists c \ \forall n \text{ выполнено } K(a_1, a_2, a_3, \dots, a_n) > n - c.$$

(Заметим для ясности, что ни для простой энтропии, ни для энтропии разрешения последовательностей с таким свойством нет вовсе, а для каждой из остальных четырёх энтропий хаотичность оказывается равнообъёмной типичности.) В той же статье Левина 1973 г., в которой была введена монотонная энтропия, содержалось и доказательство того факта, что понятие хаотичности относительно монотонной энтропии равнообъёмно понятию типичности. Независимо от Левина в том же 1973 г. Клаус-Петер Шнорр (Claus-Peter Schnorr; предварительная публикация в трудах конференции в 1972 г.) ввёл в рассмотрение свой вариант энтропии, так называемую *процессную энтропию* (у Шнорра — *process complexity*), и доказал (независимо от Левина, но очень похожим способом), что понятие хаотичности относительно процессной энтропии также равнообъёмно понятию типичности. Хотя процессная энтропия и монотонная энтропия существенно различаются (как показал Владимир Вьюгин [7, строки 6 – 4 снизу], их разность не ограничена никакой константой) и хотя впоследствии сам Шнорр перестал пользоваться своей процессной энтропией, фактически от неё отказавшись, вышеуказанную *теорему Левина* иногда называют *теоремой Левина – Шнорра.*

Префиксную энтропию ввели Левин (в диссертации на соискание учёной степени кандидата физико-математических наук, представленную к защите в 1971 г.;

[1] Основная идея этого доказательства была предложена М. Ламбальгеном для монотонных правил выбора и легко обобщается на произвольные.

её, однако, защитить не удалось, и определение префиксной энтропии было опубликовано лишь в 1974 г.)[1] и позже (но независимо от Левина) — Грегори Чейтин (Gregory J. Chaitin) (см. его статью *A theory of program size formally identical to information theory // Journal of the Association of Computing Machinery*, 1975, v. 22, no. 3, p. 329 – 340). В той же статье Чейтин ввёл понятие хаотичности относительно префиксной энтропии и объявил (без доказательства), что этот вариант хаотичности равносилен типичности; первое опубликованное доказательство этого факта появилось в статье Вьюгина [7]: следствие 3.2 на с. 38. *Префиксная энтропия* может быть определена как энтропия для семейства префиксных языков. Язык E называется *префиксным*, если он перечислим и выполнено условие:

$$\big(\langle x_1, y_1\rangle \in E \ \&\ \langle x_2, y_2\rangle \in E \ \&\ (x_1 \approx x_2)\big) \ \Rightarrow\ (y_1 = y_2).$$

Заметим ещё, что в литературе термин «сложность» (“complexity”) часто употребляется в смысле ***энтропия***, т. е. в смысле ***сложность относительно оптимального языка***.

Непредсказуемые последовательности (в смысле настоящей статьи) впервые возникли весной 1991 г. в совместном докладе под названием «Randomness and Lawlessness» («Случайность и беззаконность»), который Андрей Мучник, Алексей Семёнов и автор этих строк сделали на конференции в Калифорнии. Конференция была посвящена основаниям теории случайности и проходила с 4 по 7 марта в Институте математических исследований в социальных науках (Institute for Mathematical Studies in the Social Sciences) Станфордского университета. Содержание доклада было опубликовано в 1998 г. в виде статьи [5]. В этой статье приведены, в частности, теоремы о соотношении непредсказуемости с другими алгоритмическими лицами случайности.[2]

Отметим, что определение непредсказуемости в нашем настоящем очерке отличается одной деталью от определения в [5]. Именно, в [5] для корректности ставки требовалось выполнение не строгого неравенства $v(k) < V(k-1)$, а более слабого нестрогого неравенства $v(k) \leqslant V(k-1)$. Класс непредсказуемых последовательностей остаётся одним и тем же при обоих пониманиях корректности. Замена в определении корректности нестрогого неравенства на строгое вызвана двумя причинами. Во-первых, сама игра становится более содержательной: ведь в случае равенства ставки текущему капиталу стоит Игроку ошибиться в своём предсказании, как его капитал обнулится и он будет вынужден в дальнейшем делать лишь нулевые ставки. Во-вторых, именно строгое неравенство позволяет при переносе определений с рационально-вычислимых мер (каковые только и рассматривались в [5]) на любые вычислимые меры сохранить эффективность действий игрока. Ведь каждый раз Игрок должен удостоверяться в корректности хода. Получить же такое удостоверение алгоритмическим путём возможно лишь в варианте строгого неравенства: существует алгоритм, в том и только в том случае дающий положительный

[1] См. работу [76] из библиографии к основному тексту.

[2] Для случая, когда члены последовательности просматриваются слева направо, игровой подход к случайности и соответствующее понятие мартингала было предложено Ж. Виллем в 1930-е годы в качестве альтернативы подходу фон Мизеса, основанному на устойчивости частот.

ответ на вопрос «$v(k) < V(k-1)$?», когда и в самом деле $v(k) < V(k-1)$; не существует алгоритма, в том и только в том случае дающего положительный ответ на вопрос «$v(k) \leqslant V(k-1)$?», когда и в самом деле $v(k) \leqslant V(k-1)$.

Сама идея о связи случайности с невозможностью гарантированного выигрыша достаточно очевидна: ещё фон Мизес, не давая точных формулировок, говорил о «невозможности системы игры». Впоследствии встречались и строгие определения, однако воспроизводимая здесь (с косметическим ремонтом) формулировка из [5] кажется более близкой к интуитивному представлению о случайности. Дело в том, что предшествующие игровые определения либо использовали стратегии, в которых вычислимость (означающая наличие алгоритма, указывающего игроку его очередной ход) заменялась на другое (хотя и связанное с понятием алгоритма, но, видимо, менее естественное) требование, либо заведомо не приводили к классу последовательностей, равнообъёмному классу типическо-хаотических последовательностей. Для определения непредсказуемости из [5] остаётся надежда на указанную равнообъёмность. Ясно, что чем бóльшим количеством «лиц случайности» характеризуется какой-то точно очерченный класс последовательностей, тем обоснованнее право этого класса служить формальным аналогом расплывчатого интуитивного представления о случайности.

Используемые понятия и обозначения

Множество целых чисел обозначается $\mathbb{Z}$, множество натуральных чисел (включая нуль) — $\mathbb{N}$ множество действительных чисел — $\mathbb{R}$. Множество рациональных чисел обозначается $\mathbb{Q}$, среди них выделяются *двоично-рациональные числа* (конечные двоичные дроби), имеющие вид $m/2^n$ при целых m и n.

Число элементов в конечном множестве A обозначается $|A|$.

Если основание логарифмов не указано явно, $\log x$ обозначает логарифм по основанию 2 (как обычно, $\ln x$ — натуральный логарифм).

В некоторых оценках используется обозначение $\lfloor x \rfloor$ для целой части числа x (наибольшего целого числа, не превосходящего x), а также обозначение $\lceil x \rceil$ для наименьшего целого числа, большего или равного x.

Как обычно, запись $f \leqslant g + O(1)$ (где f и g — выражения, которые могут содержать переменные) означает, что существует число c, для которого $f \leqslant g + c$ при всех значениях переменных. Аналогичным образом $f \leqslant g + O(h)$ (при неотрицательном h) означает, что при некотором c и при всех значениях переменных выполняется неравенство $f \leqslant g + ch$. Запись $f = g + O(h)$ (при неотрицательном h) означает, что при некотором c и при всех значениях переменных выполняется неравенство $|f - g| \leqslant ch$. В частности, $f = O(h)$ означает, что $|f| \leqslant ch$ для некоторой константы c; запись $f = \Omega(h)$ означает, что $|f| \geqslant ch$ для некоторой константы $c > 0$ (обычно выражение f положительно), а $f = \Theta(h)$ означает, что $c_1 h \leqslant |f| \leqslant c_2 h$ (также обычно используется для неотрицательного f).

Через $\mathbb{B}$ мы обозначаем множество $\{0, 1\}$. Конечные последовательности нулей и единиц называются *двоичными словами*, их множество мы обозначаем Ξ. Для любого конечного множества (*алфавита*) A через A^n обозначается множество всех *слов алфавита* A длины n, то есть множество всех последовательностей длины n, составленных из элементов множества A (*букв* алфавита A). Через A^* обозначается множество слов всех длин (включая *пустое* слово Λ длины 0), так что, например, $\Xi = \mathbb{B}^*$. Длина слова x обозначается $l(x)$. Через ab мы обозначаем *конкатенацию* слов a и b, то есть результат приписывания слова b справа к слову a. Говорят, что слово a является *началом*, или *префиксом*, слова b, если $b = ax$ для некоторого слова x. Говорят, что a является *концом*, или *суффиксом*, слова b, если $b = xa$ для некоторого слова x. Говорят, что a является *подсловом* слова b, если $b = xay$ для некоторых слов x и y (другими словами, если a является началом конца b или концом начала b).

Мы рассматриваем также бесконечные последовательности нулей и единиц. Множество таких последовательностей мы называем Ω. Для всякого двоичного слова x можно рассмотреть множество всех бесконечных последовательностей, на-

чинающихся на x. Это множество обозначается Ω_x. Множества такого вида называются *интервалами*. Естественным образом определяется конкатенация двоичного слова и бесконечной последовательности нулей и единиц. Рассматриваются также бесконечные последовательности букв произвольного алфавита A; их множество обозначается A^∞.

Иногда полезно рассматривать вместе конечные и бесконечные последовательности. Мы используем обозначение Σ для множества конечных и бесконечных последовательностей нулей и единиц (так что $\Sigma = \Xi \cup \Omega$). Через Σ_x мы обозначаем множество всех конечных и бесконечных продолжений (конечного) слова x.

Мы рассматриваем вычислимые функции, аргументами и значениями которых являются двоичные слова. Функции считаются частичными (не обязательно всюду определёнными), если иное не оговорено специально. Функция f является *вычислимой*, если существует машина (программа, алгоритм), которая останавливается на тех и только тех входах, где $f(x)$ определено, и выдаёт в качестве результата $f(x)$.

Вместо двоичных слов в качестве аргументов и значений можно использовать и другие конструктивные объекты (натуральные числа, целые числа, конечные множества слов, графы и так далее) — достаточно, чтобы их можно было закодировать двоичными словами и чтобы разные такие кодировки отличались вычислимыми функциями (по слову можно алгоритмически определить, является ли оно кодом какого-либо объекта в данной кодировке, а также найти код того же объекта в другой кодировке).

Можно говорить также и о вычислимости для других объектов (действительных чисел, мер), но это каждый раз требует особого определения (см. ниже).

Множество конструктивных объектов (двоичных слов, натуральных чисел и др.) называется *перечислимым*, если существует алгоритм без входа, который печатает на выходе элементы этого множества и только их (с произвольными промежутками; алгоритм не обязан завершать работу, даже если множество конечно). Порядок, в котором напечатаны элементы, может быть любым.

Действительное число α называется *вычислимым*, если существует алгоритм, которые указывает его рациональные приближения с любой заданной точностью (то есть по любому $\varepsilon > 0$ указывает приближение к α с погрешностью не более ε; в этом случае мы говорим, что алгоритм *вычисляет число*). Действительное число α называется *перечислимым снизу*, если оно является пределом неубывающей последовательности рациональных чисел (вариант: если множество рациональных чисел, меньших α, перечислимо). Можно также говорить о *вычислимой последовательности* действительных чисел, или о *перечислимой снизу последовательности* действительных чисел, имея в виду, что каждое из чисел вычислимо (соотв. перечислимо снизу), и соответствующие алгоритмы могут быть эффективно указаны по номеру члена.

Мы рассматриваем *меры* (распределения вероятностей) на пространстве Ω; они задаются своими значениями на интервалах Ω_x и потому мы отождествляем их с неотрицательными действительными функциями p на словах, для которых $p(\Lambda) = 1$, а $p(x) = p(x0) + p(x1)$ для любого x, и говорим о мерах на двоичном дереве. Мы рассматриваем также *полумеры на двоичном дереве*, то есть распре-

деления вероятностей на конечных и бесконечных последовательностях. Они соответствуют функциям p, для которых $p(\Lambda) = 1$ и $p(x) \geqslant p(x0) + p(x1)$. Мы также говорим о *полумерах на натуральных числах*, имея в виду последовательности p_i неотрицательных действительных чисел, для которых $\sum_{i\in\mathbb{N}} p_i \leqslant 1$. Можно отождествить такие последовательности с распределениями вероятностей на множестве $\mathbb{N}_\perp$, содержащим натуральные числа и дополнительный символ $\perp$ («неопределённое значение»).

Среди полумер (на дереве и на натуральных числах) выделяются перечислимые снизу. В каждом из этих двух классов существует наибольшая полумера (с точностью до умножения на константу). Она называется *априорной вероятностью* (на дереве или на натуральных числах). Априорная вероятность натурального числа n обозначается обычно $m(n)$; априорная вероятность вершины дерева x (двоичного слова x) обозначается обычно $a(x)$. Иногда мы говорим также о значении $m(x)$ для двоичного слова x, имея в виду вычислимое взаимно однозначное соответствие между двоичными словами и натуральными числами.

Для различных видов колмогоровской сложности используются такие обозначения: $KS(x)$ обозначает простую колмогоровскую сложность (в англоязычной литературе обычно пишут $C(x)$). Префиксная сложность обозначается $KP(x)$ или $KP'(x)$, если мы хотим подчеркнуть, что мы используем определение с беспрефиксными способами описания. (В англоязычной литературе обычно используется обозначение $K(x)$.) Те же буквы используются для сложности пар, троек и т. д., а также условной сложности. Например, $KS(x\,|\,y)$ есть условная сложность (обычная, не префиксная) слова x при известном y, а $m(x, y\,|\,z)$ есть априорная вероятность пары (x, y) (точнее, натурального числа, соответствующего ей) при известном z. Для монотонной сложности используется обозначение KM, априорная сложность (минус логарифм априорной вероятности на дереве) обозначается KA. (В англоязычной литературе используются обозначения Km и даже K_m для монотонной сложности и KM для априорной сложности.) Наконец, сложность разрешения мы обозначаем KR.

Максимальное время работы оптимального декомпрессора на программах длины не более n мы обозначаем $BB(n)$ (в префиксном варианте $BP(n)$), эта величина тесно связана с максимальным натуральным числом, имеющим сложность не выше n, которое обозначается $B(n)$.

Мы используем также некоторые топологические понятия. Пространство $\mathbb{N}_\perp$ состоит из натуральных чисел с добавленным к ним элементом $\perp$ (неопределённость); открытым считается всё пространство, а также любые множества, не содержащие $\perp$. Это пространство, наряду с пространством Σ (открытыми множествами являются объединения множеств вида Σ_x) используется в общей схеме классификации сложностей. В пространствах последовательностей и действительных чисел мы говорим об *эффективно открытых* множествах, имея в виду объединения перечислимых семейств интервалов (или множествах вида Σ_x и интервалов с рациональными концами).

Большинство понятий теории вычислимости (в том числе и связанных с колмогоровской сложностью) допускает *релятивизацию*, когда во всех алгоритмах разрешается обращение к внешней процедуре, *оракулу*. Эта процедура отвечает на

вопросы о принадлежности к некоторому множеству A, которое тоже называют оракулом. Соответствующие понятия называют «разрешимостью с оракулом A», «перечислимостью с оракулом A»и т. д.; в обозначениях для сложности добавляют верхний индекс, скажем, простая сложность слова x с оракулом A обозначается $KS^A(x)$.

Из классической теории информации мы используем понятие *шенноновской энтропии*; для случайной величины ξ, принимающей k значений с вероятностями $p_1, \ldots, p_k$, её энтропия $H(\xi)$ равна $-\sum_k p_k \log p_k$. Это определение имеет смысл и для пар величин, и с его использованием *условная энтропия* величины ξ при известной η определяется как $H(\xi, \eta) - H(\eta)$. Разность $H(\xi) + H(\eta) - H(\xi, \eta)$ называется *взаимной информацией* в случайных величинах ξ и η и обозначается $I(\xi : \eta)$. То же обозначение $I(x : y)$ употребляется для взаимной информации в алгоритмической теории, но поскольку она не в точности симметрична, то обычно читают его как «информация в x об y» и определяют как $KS(y) - KS(y \mid x)$.

Литература

[1] Ahlswede R., Cai N., Li S. R., Yeung R., Network Information Flow, *IEEE Transactions on Information Theory*, v. 46 (2000), no. 4, p. 1204 – 1216.

[2] Ahlswede R., Körner J., On common information and related characteristics of correlated information sources, Preprint, Presented at the 7th Prague Conference on Information Theory (1974).

[3] Alon N., Newman I., Shen A., Tardos G., and Vereshchagin N., Partitioning multi-dimensional sets in a small number of "uniform" parts, *European Journal of Combinatorics,* v. 28 (2007), p. 134 – 144.
Предварительная версия: ECCC[1] TR05-095 (2005).

[4] Becher V., Figueira S., and Picchi R., Turing unpublished algorithm for normal numbers, *Theoretical Computer Science*, v. 377 (2007), no. 1 – 3, p. 126 – 138.

[5] Bennett C. H., Gács P., Li M., Vitányi P. M. B., and Zurek W., Information distance, *IEEE Trans. Information Theory*, v. 44 (1998), no. 4, p. 1407 – 1423.
Предварительная версия: Thermodynamics of computation and information distance, *Proc. 25th ACM Symp. on Theory of Computing* (STOC 1993), p. 21 – 30.

[6] Bienvenu L., *Game-theoretic characterization of randomness: unpredictability and stochasticity*, Ph.D. thesis, University of Marseille, 2008.

[7] Bienvenu L., Downey R., Kolmogorov complexity and Solovay functions, *Electronic Proc. 26th Symp. on Theoretical Aspects of Computer Science* (STACS 2009), stacs2009.informatik.uni-freiburg.de/proceedings.php

[8] Бьенвеню Л., Гач П., Хойруп М., Рохас К., Шень А., Алгоритмические тесты и случайность относительно классов мер. *Труды МИАН*[2], т. 274 (2011), с. 41-102.
Английский вариант: Bienvenu L., Gács P., Hoyrup M., Rojas C., and Shen A., Algorithmic tests and randomness with respect to a class of measures, *Proc. of the Steklov Institute of Mathematics*, v. 274 (2011), p. 41 – 102. См. также: arXiv:1103.1529v2.

[9] Bienvenu L., Muchnik An., Shen A., Vereshchagin N.K., Limit complexities revisited, *Theory of Computing Systems*, v. 47 (2010), no. 3, p. 720 – 736.
См. также: arXiv:0802.2833. Исправленная версия (включая более простые доказательства более сильных утверждений): arXiv:1204.0201

[10] Bienvenu L., Sablik M., The dynamics of cellular automata in shift-invariant topologies, *Proc. 11th Conference on Developments in language theory* (DLT 2007), LNCS[3] v. 4588, p. 84 – 95.

[1] ECCC = Electronical Colloquium on Computational Complexity. Опубликованные в ECCC статьи выложены в eccc.hpi-web.de.

[2] МИАН = Математический институт им. В. А. Стеклова Академии Наук

[3] LNCS = Lecture Notes in Computer Science.

[11] Bienvenu L., Shafer G., Shen A., On the history of martingales in the study of randomness, *Electronic Journal for History of Probability and Statistics*, v. 5 (2009), no. 1, p. 1 – 40, www.jehps.net/juin2009/BienvenuShaferShen.pdf
Более подробное изложение: Bienvenu L., Shen A., Algorithmic information theory and martingales, arXiv:0906.2614.

[12] Борель Э., *Случай*, Перевод с франц., М., Петроград: Госиздат, 1923, 215 с. (Современные проблемы естествознания. Кн. 8.)

[13] Борель Э., *Вероятность и достоверность*, Перевод с франц., 3-е изд. М.: Наука, 1969. 112 с.

[14] Buhrman H., Fortnow L., Laplante S., Resource-bounded Kolmogorov complexity revisited, *SIAM Journal on Computing*, v. 31 (2002), no. 3, p. 887-905.

[15] Buhrman H., Fortnow L., Newman I., Vereshchagin N., Increasing Kolmogorov complexity. *Proc. 22nd Symp. on Theoretical Aspects of Computer Science* (STACS 2005), LNCS v. 3404, p. 412 – 421.
Предварительная версия: ECCC TR04-081 (2004).

[16] Calude C. S., *Information and randomness: an algorithmic perspective.* 2nd ed., Springer-Verlag, 2002 (first edition, 1994), 450 pp. ISBN 3-540-43466-6.

[17] Calude C. S., Hertling P., Khoussainov B., Wang, Y., Recursively enumerable reals and Chaitin Omega numbers, *Proc. 15th Symp. on Theoretical Aspects of Computer Science* (STACS 1998), LNCS v. 1373, p. 596 – 606.

[18] Calude C. S., Staiger L. and Terwijn S., On partial randomness. *Annals of Pure and Applied Logic*, v. 138 (2006), no. 1-3, p. 20 – 30.

[19] Chaitin G. J., On the length of programs for computing binary sequences, *Journal of the ACM*, v. 13 (1966), no. 4, p. 547 – 569.

[20] Chaitin G. J., On the length of programs for computing binary sequences: statictical considerations, *Journal of the ACM*, v. 16 (1969), no. 1, p. 145 – 159.

[21] Chaitin G. J., Computational complexity and Gödel's incompleteness theorem, *ACM SIGACT News*, no. 9 (1971), p. 11 – 12.

[22] Chaitin G. J., Information-theoretic limitations of formal systems, *Journal of the ACM*, v. 21 (1974), no. 3, p. 403 – 424.

[23] Chaitin G. J., A theory of program size formally identical to information theory, *Journal of the ACM*, v. 22 (1975), no. 3, p. 329 – 340.

[24] Chaitin G. J., Incompleteness theorems for random reals, *Advances in Applied Mathematics*, v. 8 (1987), p. 119 – 146.

[25] Chaitin G. J., *Algorithmic information theory*, Cambridge University Press, 1987. Third printing, 1990.

[26] Champernowne D. G., The construction of decimals normal in the scale of ten, *Journal of the London Mathematical Society*, v. 8 (1933), p. 254 – 260.

[27] Chan T. H., Yeung R. W., On a relation between information inequalities and group theory, *IEEE Transactions on Information Theory*, v. 48 (2002), no. 7, p. 1992 – 1995.

[28] Chernov A., Muchnik An. A., Romashchenko A., Shen A., Vereshchagin N. K. Upper semi-lattice of binary strings with the relation "x is simple conditional to y". *Theoretical Computer Science*, v. 271 (2002), no. 1 – 2, p. 69 – 95.
Предварительная версия: *Proc. 14th IEEE Conference on Computational Complexity* (CCC 1999), p. 114 – 122.

[29] Chernov A., Hutter M., Schmidhuber J., Algorithmic complexity bounds on future prediction errors, *Information and Computation*, v. 205 (2007), p. 242 – 261. DOI 10.1016/j.ic.2006.10.004, см. также arXiv:cs/0701120.

[30] Chernov A., Shen A., Vereshchagin N., Vovk V., On-line Probability, Complexity and Randomness, *Proc. 19th Conference on Algorithmic Learning Theory* (ALT 2008), p. 138 – 153.

[31] Chung F. R. K., Graham R. L., Frankl P., Shearer J. B., Some intersection theorems for ordered sets and graphs. *Journal of Combinatorial Theory, A*, v. 43 (1986), p. 23 – 37.

[32] Church A., On the concept of a random sequence. *Bull. Amer. Math. Soc*, v. 46 (1940), no. 2, p. 130 – 135.

[33] Cormen T. H., Leiserson C. E., Rivest, R. L., Stein C., *Introduction to Algorithms*, 3 ed., Cambridge, MIT Press, 2009.

[34] Daley R.P., Minimal-program complexity of pseudo-recursive and pseudo-random sequences, *Mathematical Systems Theory* (now *Theory of Computing Systems*), v. 9 (1975), no. 1, p. 83-94.

[35] David A. P., de Rooij S., Shafer G., Shen A., Vereshchagin N. K., and Vovk V., Insuring against loss of evidence in game-theoretic probability, *Statistics & Probability Letters*, v. 81 (2011), no. 1, p. 157 – 162. См. также: arXiv:1005.1811.

[36] Day, A., Increasing the gap between descriptional complexity and algorithmic probability, *Proc. 24th IEEE Conference on Computational Complexity* (CCC 2009), p. 263 – 273. Более подробное изложение:
homepages.mcs.vuw.ac.nz/~adam/papers/day_monotone_a_priori.pdf.

[37] Downey R., Hirschfeldt D., *Algorithmic randomness and complexity*, Springer-Verlag, 2010, 855 pp. ISBN 978-0387955674.

[38] Downey R., Hirschfeldt D., Nies A., Terwijn S., Calibrating randomness., *The Bulletin of Symbolic Logic*, v. 12 (2006), no. 3, p. 411 – 491.

[39] Durand B., Levin L., Shen A., Complex tilings, *Journal of Symbolic Logic*, v. 73 (2007), no. 2, 593 – 613. См. также: arXiv:cs/0107008.

[40] Durand D., Shen A., and Vereshchagin N., Descriptive complexity of computable sequences, *Theoretical Computer Science*, v. 171 (2001), p. 47 – 58.
Предварительные версии: *Proc. 16th Symp. on Theoretical Aspects of Computer Science* (STACS 1999), LNCS v. 1563, p. 153 – 162 и ECCC TR01-087 (2001).

[41] Durand B. and Vereshchagin N., Kolmogorov – Loveland stochasticity for finite strings, *Information Processing Letters,* v. 91 (2004), p. 263–269.

[42] Fortnow L., Lee T., Vereshchagin N., Kolmogorov complexity with error, *Proc. 23rd Symp. Theoretical Aspects of Computer Science* (STACS 2006), LNCS v. 3884, p. 137 – 148.
Предварительная версия: ECCC TR04-080 (2004).

[43] Гач П., О симметрии алгоритмической информации. *Доклады АН СССР*, т. 218 (1974), № 6, с. 1265 – 1267.
Английский перевод: Gács P., On the symmetry of algorithmic information. *Soviet Math. Dokl*, v. 15 (1974), no. 5, p. 1477-1480.

[44] Gács P., Exact expressions for some randomness test, *Zeitschrift für Math. Logik und Grundlagen d. Math.*, v. 26 (1980), p. 385 – 394.

[45] Gács P., On the relation between descriptional complexity and algorithmic probability, *Theoretical Computer Science*, 1983, v. 22, p. 71 – 93.
Предварительная версия: *Proc. 22nd Symp. on Foundations of Computer Science* (FOCS 1981), p. 296 – 303.

[46] Gács P., Every sequence is reducible to a random one, *Information and Control* (now *Information and Computation*), v. 70 (1986), no. 2 – 3, p. 186 – 192.

[47] Gács, P., and Körner, J. Common information is far less than mutual information, *Problems of Control and Information Theory,* v. 2 (1973), no. 2, p. 149 – 162.

[48] Goldreich O., *Foundations of Cryptography. V. 1. Basic Tools.* Cambridge University Press, Cambridge, 2007.

[49] Gorbunov K. Yu., On a complexity of the formula $A \vee B \Rightarrow C$, *Theoretical Computer Science*, v. 207 (1998), no. 2, p. 383 – 386.

[50] Халмош П., *Теория меры*, М.: ИЛ, 1953, 292 с.

[51] Hammer D., Romashchenko A., Shen A., Vereshchagin N., Inequalities for Shannon entropies and Kolmogorov complexities, *Journal of Computer and System Sciences*, v. 60 (2000), p. 442 – 464.
Предварительная версия: *Proc. 12th IEEE Conference on Computational Complexity* (CCC 1997), p. 13 – 23.

[52] Hammer D., Shen A., A strange application of Kolmogorov complexity, *Theory of Computing Systems*, v. 31 (1998), no. 1, p. 1 – 4.

[53] Hölzl R., Kräling T., Merkle W., Time-bounded Kolmogorov complexity and Solovay functions, *Proc. 34th Symp. Mathematical Foundations of Computer Science* (MFCS 2009), LNCS v. 5734, p. 392 – 402.

[54] Impagliazzo R., Shaltiel R., and Wigderson A., Extractors and pseudo-random generators with optimal seed length, *Proceedings of the 32nd ACM Symp. on the Theory of Computing* (STOC 2000), p. 1 – 10.

[55] Kakutani S., On equivalence of infinite product measures, *Annals of Mathematics*, Second Series, v. 49 (1948), no. 1, p. 214 – 224.

[56] Kalinina E., Prefix-free and prefix-correct complexities with compound conditions, *Proc. 5th Computer Science Symp. in Russia* (CSR 2010), LNCS v. 6072, p. 259 – 265.

[57] Karpovich P., Monotone complexity of a pair, *Proc. 5th Computer Science Symp. in Russia* (CSR 2010), LNCS v. 6072, p. 266 – 275.

[58] Kjos-Hanssen B., The probability distribution as a computational resource for randomness testing, *Open access Journal of Logic and Analysis*, v.2 (2010), `logicandanalysis.org/index.php/jla/article/view/78`.

[59] Kjos-Hanssen B., Merkle W., Stephan F., Kolmogorov complexity and the recursion theorem, *Transaction of the American Mathematical Society*, v. 363 (2010), p. 5465 – 5480.
Предварительная версия: *Proc. 23rd Symp. on Theoretical Aspects of Computer Science* (STACS 2006), LNCS v. 3884, p. 149 – 161.

[60] Kleene S. C., On the interpretation of intuitionistic number theory. *Journal of Symbolic Logic*, v. 10 (1945), pp. 109 – 124.

[61] Kolmogoroff A., Zur Deutung der intuitionistishen Logik. *Mathematische Zeitschrift*, Bd. 35 (1932), H. 1, S. 58 – 65.
Русский перевод: К толкованию интуиционистской логики, в сборнике: Колмогоров А. Н., *Избранные труды. Математика и механика*, М.: Наука, 1985, с. 142 – 148.

[62] Kolmogorov A. N., On tables of random numbers, *Sankhyā, The Indian Journal of Statistics, Ser. A*, v. 25 (1963), no. 4, p. 369 – 376. Перепечатано в: *Theoretical Computer Science*, v. 207 (1998), no. 2, p. 387 – 395.
Русский перевод: О таблицах случайных чисел. *Семиотика и информатика*, 1982, вып. 18, с. 3 – 13, М.: ВИНИТИ. Перепечатано в: Колмогоров А. Н., *Теория информации и теория алгоритмов*, М.: Наука, 1987, с. 204 – 213.

[63] Колмогоров А. Н., Три подхода к определению понятия «количество информации», *Проблемы передачи информации*, т. 1 (1965), вып. 1, с. 3 – 11
Английский перевод: Kolmogorov A. N., Three approaches to the quantitative definition of information. *Problems Inform. Transmission*, v. 1 (1965), no. 1, p. 1 – 7.

[64] Колмогоров А. Н., К логическим основам теории информации и теории вероятностей, *Проблемы передачи информации*, т. 5 (1969), вып. 3, с. 3 – 7.
Несколько раньше опубликован английский перевод (переводчик не указан): Kolmogorov A. N., Logical basis for information theory and probability theory, *IEEE Trans. Inform. Theory*, v. 14 (1968), p. 662 – 664.

[65] Колмогоров А. Н., Комбинаторные основания теории информации и исчисления вероятностей [доклад на Международном математическом конгрессе (Ницца, 1970)], *Успехи математических наук*, т. 38 (1983), №4, с. 27 – 36.

[66] Колмогоров А. Н., Фомин С. В. *Элементы теории функций и функционального анализа*, 4-е изд., М.: Наука, 1976, 544 с.

[67] Колмогоров А. Н., Успенский В. А., Алгоритмы и случайность, *Теория вероятностей и её применения*, т. 32 (1987), вып. 3, с. 425 – 455.
Английские переводы: Kolmogorov A. N. and Uspensky V. A., Algorithms and randomness, *SIAM J. Theory Probab. Appl.* v. 32 (1987) p. 389 – 412 [with annoying translation errors[1]]. Without annoying translation errors: Prokhorov Yu. V. and Sazonov V. V., Eds., *Proc. 1st World Congress of the Bernoulli Society* (*Tashkent 1986*), v. 1: Probab. Theory and Appl., VNU Science Press, Utrecht 1987, p. 3 – 55.

[68] *Колмогоров и кибернетика.* Под ред. Д. А. Поспелова, Я. И. Фета. Новосибирск: издательство ИВМ и МГ СО РАН, 2001. (Вопросы истории информатики, вып. 2). Стенограмма доклада Колмогорова в Институте философии АН СССР (23 апреля 1965 года) опубликована на с. 118—137. Её также можно найти (октября 2014) по адресу `http://cshistory.nsu.ru/?int=VIEW&el=1832&templ=INTERFACE`.

[69] Kučera A., Measure, Π^0_1-classes and complete extensions of PA, в сборнике: Ebbinghaus H.-D., Müller G. H. and Sacks G. E. (Eds.), *Recursion Theory Week* (*Oberwolfach, 1984*), Lecture Notes in Mathematics, v. 1141 (1985), p. 245 – 259.

[70] Kučera A., Slaman T., Randomness and recursive enumerability, *SIAM Journal on Computing*, v. 31 (2001), no. 1, p. 199 – 211.

[71] Kuipers L., Niederreiter H., *Uniform distribution of sequences*, Wiley-Interscience, 1949.

[72] Kummer M., On the complexity of random strings, *Proc. 13th Symp. on Theoretical Aspects of Computer Science* (STACS 1996), LNCS v. 1046, p. 25 – 36.

[73] van Lambalgen M., *Random sequences*, Ph. D. Thesis, University of Amsterdam, 1987.

[74] de Leeuw K., Moore E. F., Shannon C. E., and Shapiro N., Computability by probabilistic machines, в сборнике: C. E. Shannon and J. McCarthey (Eds.), Automata Studies, Princeton University Press, Princeton, New Jersey, 1956, p. 183 – 212.

[1]Some of those errors drastically distort meaning; e. g. Russian term перечислимый, which should be translated as enumerable, or recursively enumerable, was translated as countable.

[75] Левин Л. А., О понятии случайной последовательности, *Доклады АН СССР*, т. 212 (1973), № 3, с. 548 – 550.
Английский перевод: Levin L.A., On the notion of a random sequence, *Soviet Math. Dokl.*, v. 14 (1973), p. 1413 – 1416.

[76] Левин Л. А., Законы сохранения (невозрастания) информации и вопросы обоснования теории вероятностей, *Проблемы передачи информации*, т. 10 (1974), вып. 3, с. 30 – 35.
Английский перевод: Levin L. A., Laws of information conservation (nongrowth) and aspects of the foundation of probability theory, *Problems of Information Transmission*, v. 10 (1974), p. 206 – 210.

[77] Левин Л. А., О различных мерах сложности конечных объектов (аксиоматическое описание), *Доклады АН СССР*, т. 227 (1976), № 4, с. 804 – 807.
Английский перевод: Levin L. A., Variuos measures of complexity for finite objects (axiomatic description), *Soviet Math. Dokl.*, v. 17 (1976), p. 522 – 526.

[78] Левин Л. А. О принципе сохранения информации в интуиционистской математике, *Доклады АН СССР*, т. 227 (1976), № 6, с. 1293-1296.
Английский перевод: Levin L. A., On the principle of conservation of information in intuitionistic mathematics, *Soviet Math. Dokl.*, v. 17 (1976), no. 2, p. 601 – 605

[79] Левин Л. А., Равномерные тесты случайности, *Доклады АН СССР*, т. 227 (1976), № 1, с. 33-35.
Английский перевод: Levin L.A., Uniform tests of randomness, *Soviet Math. Dokl.*, v. 17 (1976), issue 2, p. 337–340.

[80] Левин Л. А., Об одном конкретном способе задания сложностных мер, *Доклады АН СССР*, т. 234 (1977), № 3, с. 536 – 539.
Английский перевод: On a concrete method of assigning complexity measures, *Soviet Math. Dokl.*, v. 18 (1977), no. 3, p. 727 – 731.

[81] Levin L. A., A concept of independence with application in various fields of mathematics, MIT Technical Report, MIT/LCS/TR-235, 1980, 21 p.

[82] Levin L. A., Randomness conservation inequalities: information and independence in mathematical theories, *Information and Control*, v. 61 (1984), no. 1 – 2, p. 15 – 37.

[83] Levin L. A., Vyugin V. V., Invariant properties of informational bulks, *Proc. 6th Symp. on Mathematical Foundations of Computer Science* (MFCS 1977), LNCS v. 153, p. 359 – 364.

[84] Levin L. A., Forbidden information, 2002, 8 pp., arXiv:cs/0203029.
Предварительная версия: *Proc. 43th IEEE Symp. on Foundations of Computer Science* (FOCS 2002), p. 761 – 768.

[85] Li M., Vitányi P., *An Introduction to Kolmogorov complexity and its applications*, 3rd ed., Springer, 2008 (1 ed., 1993; 2 ed., 1997), 792 pp. ISBN 978-0-387-49820-1.

[86] Li S. R., Yeung R. W., Cai N., Linear network coding, *IEEE Transactions on Information Theory*, v. 49 (2003), no. 2, p. 371 – 381.

[87] Loomis L. H., Whitney H., An inequality related to the isoperimetric inequality, *Bulletin American Mathematical Society*, v. 55 (1949), p. 961 – 962.

[88] Loveland D. W., A new interpretation of von Mises' concept of a random sequence, *Z. Math. Logik und Grundlagen d. Math.*, v. 12 (1966), p. 279 – 294.

[89] Loveland D. W., The Kleene hierarchy classification of recursively random sequences, *Trans. Amer. Math. Soc.*, v. 125 (1966), p. 497 – 510.

[90] Loveland D. W., A Variant of the Kolmogorov concept of complexity, *Information and Control* (now *Information and Computation*), v. 15 (1969), p. 510 – 526.

[91] Loveland D. W., On minimal-program complexity measures, *Proc. 1st ACM Symp. on Theory of Computing* (STOC 1969), p. 61 – 65.

[92] Lutz J., Dimension in complexity classes, *SIAM Journal on Computing*, v. 32 (2003), p. 1236 – 1259.
Предварительная версия: *Proc. 15th IEEE Conference on Computational Complexity* (CCC 2000), p. 158 – 169.

[93] Lutz J. H., The dimensions of individual strings and sequences, *Information and Computation*, v. 187 (2003), no. 1, p. 49 – 79.
Предварительная версия: Gales and the constructive dimension of individual sequences, *Proc. 27th Colloquium on Automata, Languages, and Programming* (ICALP 2000), LNCS v. 1853, p. 902 – 913.

[94] Makarychev K., Makarychev Yu., Chain independence and common information. *IEEE Transactions on Information Theory*, v. 58 (2012), no. 8, p. 5279 – 5286.
См. также: *Conditionally independent random variables*, arXiv:cs/0510029.

[95] Makarychev K., Makarychev Yu., Romashchenko A., Vereshchagin N., A new class of non-Shannon-type inequalities for entropies, *Communications in Information and Systems,* v. 2 (2002), no. 2, p. 147-166.

[96] Манин Ю. И. *Вычислимое и невычислимое*, М.: Советское радио, 1980, 128 с.

[97] Martin-Löf P., The definition of random sequences, *Information and Control* (now *Information and Computation*), v. 9 (1966), p. 602 – 619.

[98] Martin-Löf P., Algorithmen und zufällige Folgen, Vier Vorträge von Per Martin-Löf (Stockholm) gehalten am Mathematischen Institut der Universität Erlangen – Nürnberg, Als Manuskript vervielfältigt, Erlangen, 1966, 61 pp, см. также www.probabilityandfinance.com/misc/erlangen.pdf.

[99] Martin-Löf P., Complexity oscillations in infinite binary sequences, *Z. Wahrscheinlichkeitstheorie verw. Geb.*, v. 19 (1971), p. 225 – 230.

[100] Mayordomo E., A Kolmogorov complexity characterization of constructive Hausdorff dimension, *Information Processing Letters*, v. 84 (2002), no. 1, p. 1 – 3.

[101] Merkle W. The Kolmogorov – Loveland stochastic sequences are not closed under selecting subsequences, *Journal of Symbolic Logic*, v. 68 (2003), p. 1362 – 1376.
Предварительная версия: *Proc. 29th Colloquium on Automata, Languages, and Programming* (ICALP 2002), LNCS v. 2380, p. 390 – 400.

[102] Merkle W., The complexity of stochastic sequences, *Proc. 18th IEEE Conference on Computational Complexity* (CCC 2003), p. 230 – 235.

[103] Miller J., Every 2-random real is Kolmogorov random, *Journal of Symbolic Logic*, v. 69 (2004), no. 3, p. 907 – 913.

[104] Miller J., Contrasting plain and prefix-free Kolmogorov complexity, (preprint from 2006), www.math.wisc.edu/~jmiller/Notes/contrasting.pdf.

[105] Miller J., Yu L., On initial segment complexity and degrees of randomness, *Transactions of the American Mathematical Society*, v. 360 (2008), no. 6, p. 3193 – 3210.

[106] Miller J., Two notes on subshifts, *Proceedings of the American Mathematical Society*, v. 140 (2012), no. 5, p. 1617–1622. См. также www.math.wisc.edu/~jmiller/Papers/subshifts.pdf

[107] von Mises R., Grundlagen der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift*, Bd. 5 (1919), S. 52 – 99. Перепечатано в книге: Selected Papers of Richard von Mises. Volume Two. Probability and Statistics, General. American Mathematical Society, 1964. p. 57 – 106.

[108] von Mises R., *Wahrscheinlichkeit, Statistik und Wahrheit*, Wien: Springer-Verlag, 1928, 189 p.
Русский перевод: Мизес Р., *Вероятность и статистика*, М. – Л.: Госиздат, 1930.

[109] von Mises R., On the foundations of probability and statistics, *Annals of Mathematical Statistics*, v. 12 (1941), p. 191 – 205. Перепечатано в книге: Selected Papers of Richard von Mises. Volume Two. Probability and Statistics, General. American Mathematical Society, 1964. p. 340 – 355.

[110] von Mises R., Doob J. L., Discussion of papers on probability theory. *Annals of Mathematical Statistics*, v. 12 (1941), p. 215 – 217. (Перепечатано в книге: Selected Papers of Richard von Mises. Volume Two. Probability and Statistics, General. American Mathematical Society, 1964, p. 356 – 359.)

[111] Moser R.A., A constructive proof of the Lovasz Local Lemma. *Proc. 41st ACM Symp. on Theory of Computing* (STOC 2009), p. 343-350, см. также `arXiv:0810.4812`.

[112] Moser R.A., Tardos G., A constructive proof of the general Lovasz Local Lemma, *Journal of the ACM*, v.57 (2010), no. 2, p. 11.1 – 11.15, см. также `arXiv:0903.0544`.

[113] Мучник Ан. А. Об основных структурах дескриптивной теории алгоритмов. *Доклады АН СССР*, 1985, т. 285, №2, с. 280 – 283.
Английский перевод: Muchnik An. A., On the basic structures of the descriptive theory of algorithms, *Soviet Math. Dokl.*, v. 32 (1985), no. 3, p. 671 – 674.

[114] Мучник Ан. А., Нижние пределы частот в вычислимых последовательностях и релятивизованная априорная вероятность, *Теория вероятностей и её применения*, т. 32 (1987), №3, с. 563 – 565.
Английский перевод: Muchnik An. A., Lower limits on frequences in computable sequences and relativized a priori probability, *SIAM Theory Probab. Appl.*, v. 32 (1987) p. 513 – 514.

[115] Muchnik An. A., On common information, *Theoretical Computer Science*, v. 207, no. 2 (1998), p. 319 – 328.

[116] Muchnik An. A. Conditional complexity and codes, *Theoretical Computer Science*, v. 271 (2002), no. 1 – 2, p. 97 – 109.
Предварительная версия: Muchnik A., Semenov A., Multi-conditional descriptions and codes in Kolmogorov complexity, ECCC TR00-015 (2000).

[117] Muchnik An. A., Mezhirov I., Shen A., Vereshchagin N. K., *Game interpretation of Kolmogorov complexity*, 2010, `arXiv:1003.4712`.

[118] Muchnik An. A., Positselsky S. E., Kolmogorov entropy in the context of computability theory, *Theoretical Computer Science*, v. 271 (2002), no. 1 – 2, p. 15 – 35.

[119] Мучник Ан. А., Ромащенко А. Е., Устойчивость колмогоровских свойств при релятивизации. *Проблемы передачи информации*, т. 46, вып. 1 (2010), с. 42 – 67.
Английский перевод: Muchnik An. A., Romashchenko A., Stability of properties of Kolmogorov complexity under relativization. *Problems of Information Transmission*, v. 46, no. 1 (2010), p. 38 – 61.
Предварительная версия: Muchnik An. A., Romashchenko A., Random oracle does not help extract the mutual information, *Proc. 33rd Symp. on Mathematical Foundations of Computer Science* (MFCS 2008), LNCS v. 5162, p. 527 – 538.

[120] Muchnik An. A., Semenov A. L., Uspensky V. A., Mathematical metaphysics of randomness, *Theoretical Computer Science*, v. 207, no. 2 (1998), p. 263 – 317.

[121] Muchnik An. A., Shen A., Ustinov M., Vereshchagin N. K., Vyugin M., Non-reducible descriptions for conditional Kolmogorov complexity. *Theoretical Computer Science*, v. 384 (2007), no. 1, p. 77 – 86.
Предварительные версии: Muchnik An. A., Shen A., Vereshchagin N. K., Vyugin M., Non-reducible descriptions for conditional Kolmogorov complexity, *Proc. 3rd Conference on Theory and Applications of Models of Computation* (TAMC 2006), LNCS v. 3959, p. 308 – 317 и ECCC TR04-054 (2004).

[122] Muchnik An. and Vereshchagin N., Shannon entropy vs. Kolmogorov complexity, *Proc. 1st Computer Science Symp. in Russia* (CSR 2006), LNCS v. 3967, p. 281 – 291.

[123] Musatov D., Improving the space-bounded version of Muchnik's conditional complexity theorem via "naive" derandomization, *Proc. 6th Computer Science Symp. in Russia* (CSR 2011), LNCS v. 6651, p. 64 – 76.

[124] Musatov D., Space-bounded Kolmogorov extractors, *Proc. 7th Computer Science Symp. in Russia* (CSR 2012), LNCS v. 7353, p. 266 – 277.

[125] Musatov D., Romashchenko A., Shen A., Variations on Muchnik's conditional complexity theorem, *Theory Comput. Syst.*, v. 49 (2011), no.2, p. 227 – 245.
Предварительная версия: *Proc. 4th Computer Science Symp. in Russia* (CSR 2009), LNCS v. 5675, p. 250 – 262.

[126] Niederreiter H., A combinatorial problem for vector spaces over finite fields, *Discrete Mathematics*, v. 96 (1991), no. 3, p. 221 – 228.

[127] Nies A., *Computability and randomness*, Oxford University Press, 2009, 420 pp. ISBN 978-0-19-923076-1

[128] Nies A., Stephan F., Terwijn S., Randomness, relativization and Turing degrees, *Journal of Symbolic Logic*, v. 70 (2005), no. 2, p. 515 – 535.

[129] Razenshteyn I., Common information revisited, 2012, `arXiv:1104.3207`.

[130] Reimann J., *Computability and fractal dimension*, PhD thesis, Ruprecht-Karls Universität Heidelberg, 2004 (URN: `urn:nbn:de:bsz:16-opus-55430`), см. `www.ub.uni-heidelberg.de/archiv/5543`.

[131] Ромащенко А. Е., Пары слов с нематериализуемой взаимной информацией, *Проблемы передачи информации*, т. 36 (2000), вып. 1, с. 3 – 20.
Английский перевод: Romashchenko A., Pairs of words with nonmaterializable mutual information, *Problems of Information Transmission*, v. 36 (2000), no. 1, p. 3 – 20.

[132] Romashchenko A. Extracting the mutual information for a triple of binary strings. *Proc. 18th IEEE Conference on Computational Complexity* (CCC 2003), p. 221 – 235.

[133] Ромащенко А., Румянцев А., Шень А., *Заметки по теории кодирования*, Изд-во МЦНМО, 2011, 80 с.

[134] Romashchenko A., Shen A., Vereshchagin N., Combinatorial interpretaion of Kolmogorov complexity, *Theoretical Computer Science*, v. 271 (2002), no. 1 – 2, p. 111 – 123.
Предварительные версии: *Proc. 15th IEEE Conference on Computational Complexity* (CCC 2000), p. 131 – 137 и ECCC TR00-026 (2000).

[135] Rumyantsev A., Kolmogorov complexity, Lovasz Local Lemma and critical exponents, *Proc. 2nd Computer Science Symp. in Russia* (CSR 2007), LNCS v. 4649, p. 349 – 355. См. также: `arXiv:1009.4995`.

[136] Rumyantsev A., Infinite computable version of Lovasz Local Lemma, 2010, arXiv:1012.0557.

[137] Rumyantsev A., Ushakov M., Forbidden substrings, Kolmogorov complexity and almost periodic sequences, *Proc. 23rd Symp. on Theoretical Aspects of Computer Science* (STACS 2006), LNCS v. 3884, p. 396 – 407. См. также: arXiv:1009.4455.

[138] Schmidt W., On normal numbers, *Pacific Journal of Mathematics*, v. 10 (1960), no. 2, p. 661 – 672.

[139] Schnorr C. P., Eine Bemerkung zum Begriff der zufälligen Folge. *Zeitschrift für Wahrscheinlichkeitstheorie und Verw. Gebiete*, v. 14 (1969), p. 27 – 35.

[140] Schnorr C. P., Über die Definition effektiver Zufallstests. Teil I. *Zeitschrift für Wahrscheinlichkeitstheorie und Verw. Gebiete*, v. 15 (1970), p. 297 – 312.

[141] Schnorr C. P., Über die Definition effektiver Zufallstests. Teil II. *Zeitschrift für Wahrscheinlichkeitstheorie und Verw. Gebiete*, v. 15 (1970), p. 313 – 328.

[142] Schnorr C. P., Klassifikation der Zufallsgesetze nach Komplexität und Ordnung, *Zeitschrift für Wahrscheinlichkeitstheorie und Verw. Gebiete*, v. 16 (1970), p. 1 – 21.

[143] Schnorr C. P., *Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie*, Lecture Notes in Mathematics, v. 218, IV+212 S, Springer, 1971.

[144] Schnorr C. P., A unified approach to the definition of random sequences, *Mathematical Systems Theory* (now *Theory of Computing Systems*), v. 5 (1971), no. 3, p. 246 – 258.

[145] Schnorr C. P, Optimal Gödel numberings. *Proc. IFIP congress 71* (1971), v. 1. p. 56 – 58.

[146] Schnorr C. P., Process complexity and effective random tests, *Journal of Computer and System Sciences*, v. 7 (1973), p. 376–388.
Предварительная версия: *Proc. 4th ACM Symp. on Theory of Computing* (STOC 1972), p. 168 – 176.

[147] Schnorr C. P., Optimal enumerations and optimal Gödel numberings, *Mathematical Systems Theory* (now *Theory of Computing Systems*), v. 8 (1975), no. 2, p. 182 – 191.

[148] Shafer G., Shen A., Vereshchagin N. K., Vovk V., Test martingales, Bayes factors, and p-values, *Statistical Science* v. 26 (2011), no. 1, p. 84 – 101. См. также arXiv:0912.4269.

[149] Shafer G., Vovk V. Probability and finance: it's only a game! New York: Wiley, 2001.

[150] Шень А., Аксиоматическое описание понятия энтропии конечного объекта, *Логика и основания математики. Тезисы VIII Всесоюзной конференции «Логика и методология науки», г. Паланга, 26 – 28 сентября 1982 г.*, Вильнюс, 1982, с. 104 – 105.

[151] Шень А., К логическим основам применения теории вероятностей, *Семиотические аспекты формализации интеллектуальной деятельности, Школа-семинар, г. Телави, 29 октября – 6 ноября 1983 г.*, с. 144 – 146.

[152] Шень А., Алгоритмические варианты понятия энтропии, *Доклады АН СССР*, т. 276 (1984), № 3, с. 563 – 566.
Английский перевод: Shen A., Algoritmic variants of the notion of entropy, *Soviet Math. Dokl.*, v. 29 (1984), no. 3, p. 569 – 573.

[153] Шень А., О соотношениях между различными алгоритмическими определениями случайности, *Доклады АН СССР*, т. 302 (1988), № 3, с. 548 – 552.
Английский перевод: Shen A., On relations between different algorithmic definitions of randomness, *Soviet Math. Dokl.*, v. 38 (1989), no. 2, p. 316 – 319.

[154] Shen A., Algorithmic information theory and Kolmogorov complexity, kecture notes, http://www.it.uu.se/research/publications/reports/2000-034.

[155] Shen A., Multisource algorithmic information theory, *Proc. 3rd Conference on Theory and Applications of Models of Computation* (TAMC 2006), LNCS v. 3959, p. 327 – 338.

[156] Shen A., Algorithmic information theory and foundations of probability, *Proc. 3rd Workshop on Reachability Problems* (2009), LNCS v. 5797, p. 26 – 34.

[157] Shen A. and Vereshchagin N., Logical operations and Kolmogorov complexity, *Theoretical Computer Science,* v. 271 (2002), p. 125 – 129.
Предварительная версия: ECCC TR01-088 (2001).

[158] Sipser M., *Introduction to the theory of computation*, PWS Publishing, 1996.

[159] Slepian D., Wolf J. K., Noiseless coding of correlated information sources, *IEEE Transactions on Information Theory*, v. IT-19 (1973), no. 4, p. 471 – 480.

[160] Solomonoff R. J., A formal theory of inductive inference, part 1, part 2. *Information and Control* (now *Information and Computation*), v. 7 (1964), p. 1 – 22, p. 224 – 254.

[161] Solovay R., *Draft of a paper (or series of papers) on Chaitin's work*, done for the most part during Sept. – Dec. 1974, unpublished (but available to many people in the field).

[162] Solovay R. M., On Random R.E. Sets, в сборнике: A.I. Arruda, N.C.A. da Costa, R. Chaqui (Eds.), Non-classical logics, model theory and computability, North-Holland, Amsterdam, 1977, p. 283 – 307.

[163] Tadaki K., A generalization of Chaitin's halting probability Ω and halting self-similar sets, *Hokkaido mathematical journal*, v. 31 (2002), no. 1, p. 219 – 253. См. также: arXiv:nlin/0212001.

[164] Takahashi H., On a definition of random sequences with respect to conditional probability, *Information and Computation*, v. 206 (2008), no. 12, p. 1375 – 1382.

[165] Takahashi H., Algorithmic randomness and monotone complexity on product space, *Information and Computation*, v. 209 (2011), no. 2, p. 183 – 197, dx.doi.org/10.1016/j.ic.2010.10.003. См. также arXiv.org:0910.5076.

[166] Успенский В. А., Семёнов А. Л., *Алгоритмы: основные идеи и приложения*, М.: Наука, 1987, 288 с.
Английский перевод: Uspensky V., Semenov A., *Algorithms: main ideas and applications*, translated by A. Shen, Kluwer Academic Publishers, 1993, 269 pp. ISBN: 0-7923-2210-X.

[167] Успенский В. А., Семёнов А. Л., Шень А., Может ли (индивидуальная) последовательность нулей и единиц быть случайной? *Успехи математических наук*, т. 45 (1990), вып. 1 (271), с. 105 – 162
Английский перевод: Uspenskii V. A., Semenov A. L., Shen' A. Kh., Can an individual sequence of zeros and ones be random? *Russian Math. Surveys*, v. 45 (1990), no. 1, p. 121 – 189.

[168] Uspensky V. A., Shen A., Relations between varieties of Kolmogorov complexities, *Mathematical Systems Theory* (now *Theory of Computing Systems*), v. 29 (1996), no. 3, p. 271 – 292.

[169] Vereshchagin N. K. Kolmogorov complexity conditional to large integers. *Theoretical Computer Science*, v. 271 (2002), no. 1 – 2, p. 59 – 67.
Предварительная версия: ECCC TR01-086 (2001).

[170] Vereshchagin N., Kolmogorov complexity of enumerating finite sets, *Information Processing Letters*, v. 103 (2007), p. 34 – 39.
Предварительная версия: ECCC TR04-030 (2004).

[171] Vereshchagin N. K., Kolmogorov complexity and games, *Bulletin of the EATCS,* v. 94 (2008), p. 43 – 75.

[172] Vereshchagin N., Minimal sufficient statistic revisited, *Proc. 5th Conference on Computability in Europe* (CiE 2009), LNCS v. 5635, p. 478 – 487.

[173] Верещагин Н. К., Мучник Ан. А., О совместной условной сложности (энтропии), *Труды МИАН,* т. 274 (2011), с. 103 – 118.
Английский перевод: Vereshchagin N. K. and Muchnik An. A, On joint conditional complexity (Entropy), *Proc. of the Steklov Institute of Mathematics*, v. 274 (2011), p. 90-104.
Предварительные версии: Muchnik An. A. and Vereshchagin N. K., Logical operations and Kolmogorov complexity. II. *Proc. 16th IEEE Conference on Computational Complexity* (CCC 2001), p. 256 – 265 и ECCC TR01-089 (2001).

[174] Верещагин Н. К., Шень А., *Лекции по математической логике и теории алгоритмов, Часть 2, Языки и исчисления*, 3-е изд., М.: МЦНМО, 2008, ftp://ftp.mccme.ru/users/shen/logic/firstord.

[175] Верещагин Н. К., Шень А., *Лекции по математической логике и теории алгоритмов, Часть 3, Вычислимые функции,* 3-е изд., М.: МЦНМО, 2008, ftp://ftp.mccme.ru/users/shen/logic/comput.
Английский перевод: Vereshchagin N K. and Shen A., *Computable functions*, American Mathematical Society, Student mathematical library, vol. 19, 2003.

[176] Верещагин Н. К., Скворцов Д. П., Скворцова Е. З., Чернов А. В., Варианты понятия реализуемости для пропозициональных формул, приводящие к логике слабого закона исключённого третьего, *Труды МИАН*, т. 242 (2003), с. 77 – 97.
Английский перевод: Chernov A., Skvortsov D., Skvortsova E., and Vereshchagin N., Variants of realizability for propositional formulas and the logic of the weak law of excluded middle. *Proc. of Steklov Institute of Mathematics,* v. 242 (2003), p. 67 – 85. Предварительная английская версия: *Proc. 16th Workshop on Computer Science Logic* (CSL 2002), LNCS v. 2471, p. 74 – 88.

[177] Vereshchagin N. K, and Vitányi P. M. B, Kolmogorov's structure functions with an application to the foundations of model selection, *IEEE Transactions on Information Theory*, v. 50 (2004), no. 12, p. 3265 – 3290.
Предварительная версия: *Proc. 43th IEEE Symp. on Foundations of Computer Science* (FOCS 2002), p. 751 – 760.

[178] Vereshchagin N. K, and Vitányi P. M. B, Rate distortion and denoising of individual data using Kolmogorov complexity, *IEEE Transactions on Information Theory*, v. 56 (2010), no. 7, p. 3438 – 3454, см. также arXiv:cs/0411014 (2004).
выложено в arXiv:cs/0411014 (2004).

[179] Vereshchagin N. and Vyugin M., Independent minimum length programs to translate between given strings, *Theoretical Computer Science,* v. 271 (2002), p. 131 – 143.
Предварительная версия: ECCC TR00-035 (2000).

[180] Ville J., *Étude critique de la notion de collectif*, Gauthier-Villars, 1939. (Monographies des probabilités. Calcul des probabilites et ses applications. Publiée sous la direction de M. Émile Borel. Fascicule III.)

[181] Вовк В. Г., Об одном критерии случайности, *Доклады АН СССР*, т. 294 (1987), № 6, с. 1298 – 1302.
Английский перевод: Vovk V.G., On a randomness criterion, *Soviet Mathematics Doklady*, v. 35 (1987), no. 3, p. 656 – 660. См. также: www.vovk.net/criterion.pdf

[182] Вовк В. Г., Закон повторного логарифма для случайных по Колмогорову, или хаотических, последовательностей. *Теория вероятностей и её применения*, т. 32 (1987), № 3, с. 456 – 468. Английский перевод: Vovk V.G., The law of the iterated logarithm for random Kolmogorov, or chaotic, sequences. *Theory of Probability and its Applications*, v. 32 (1987), no. 3, p. 413 – 425.

[183] Vovk V., V'yugin V.V., On the empirical validity of Bayesian method, *Journal of the Royal Statistical Society*, B, v. 55 (1993), p. 253 – 266.

[184] Вьюгин В. В., Алгоритмическая энтропия (сложность) конечных объектов и её применение к определению случайности и количества информации, *Семиотика и информатика,* вып. 16 (1981), с. 14 – 43, М: ВИНИТИ.
Английский перевод: V'yugin V. V., Algorithmic entropy (complexity) of finite objects and its application to defining randomness and amount of information. *Selecta Mathematica formerly Sovietica*, v. 13 (1994), no. 4, p. 357–389.

[185] V'yugin V.V., Ergodic theorems for individual random sequences, *Theoretical Computer Science*, v. 207 (1998), no. 2, p. 343 – 361.

[186] V'yugin V.V., Non-stochastic infinite and finite sequences, *Theoretical Computer Science*, v. 207 (1998), no. 2, p. 363 – 382.

[187] Wald A., Sur la notion de collectif dans le calcul des probabilités (On the notion of collective in probability theory), présentée par M. Émile Borel. *Comptes rendus*, v. 202, p. 180 – 183 (séance du 20 janvier 1936).

[188] Wald A., Die Wiederspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung, *Ergebnisse eines matematischen Kolloquiums*, v. 8 (1937), p. 38 – 72. Перепечатано в: Menger K., *Ergebnisse eines Mathematischen Kolloquiums*, Springer: Wien, New York, 1998.

[189] Wall D. D., *Normal Numbers.* Ph. D thesis, University of California, Berkeley CA, 1949.

[190] Wigderson, A., Randomness extractors — applications and constructions. *Proc. 29th Conference on Foundations of Software Technology and Theoretical Computer Science* (FSTTCS 2009), DOI 10.4230/LIPIcs.FSTTCS.2009.2340,
drops.dagstuhl.de/opus/volltexte/2009/2340/.

[191] Yeung, R. W., *A First course in information theory*, Kluwer Academic / Plenum Publishers, 2002.

[192] Заславский И. Д., Цейтин Г. С., О сингулярных покрытиях и связанных с ними свойствах конструктивных функций, *Труды МИАН,* т. 67 (1962), с. 458 – 502.

[193] Zhang Z., Yeung R.W., A non-Shannon-type conditional information inequality, *IEEE Transactions on Information Theory*, v. 43 (1997), p. 1982 – 1986.

[194] Zhang Z., Yeung R.W., On characterization of entropy function via information inequalities, *IEEE Transactions on Information Theory*, v. 44 (1998), p. 1440 – 1450.

[195] Zimand M., Guest Column: Possibilities and Impossibilities in Kolmogorov complexity extraction. *SIGACT News*, December 2010.

[196] Звонкин А. К., Левин Л. А., Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов. *Успехи математических наук*, т. 25 (1970), вып. 6 (156), с. 85 – 127.

Английский перевод: Zvonkin A K., Levin L. A., The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, v. 25 (1970), no. 6, p. 83 – 124.

[197] Miller J.S., The K-Degrees, Low for K Degrees, and Weakly Low for K Sets. *Notre Dame Journal of Formal Logic*, v. 50, no. 4 (2009), p. 381–391.

[198] Bauwens B., *Plain and prefix complexity characterizations of* 2*-randomness: simple proofs*, arXiv:1310.5230 (2013).

Предметный указатель

Указатель имён

Оглавление