Maximum Entropy,
Information Without Probability and Complex Fractals

Fundamental Theories of Physics

An International Book Series on The Fundamental Theories of Physics: Their Clarification, Development and Application

Editor:
ALWYN VAN DER MERWE, *University of Denver, U.S.A.*

Editorial Advisory Board:
JAMES T. CUSHING, *University of Notre Dame, U.S.A.*
GIANCARLO GHIRARDI, *University of Trieste, Italy*
LAWRENCE P. HORWITZ, *Tel-Aviv University, Israel*
BRIAN D. JOSEPHSON, *University of Cambridge, U.K.*
CLIVE KILMISTER, *University of London, U.K.*
PEKKA J. LAHTI, *University of Turku, Finland*
ASHER PERES, *Israel Institute of Technology, Israel*
EDUARD PRUGOVECKI, *University of Toronto, Canada*
TONY SUDBURY, *University of York, U.K.*
HANS-JÜRGEN TREDER, *Zentralinstitut für Astrophysik der Akademie der Wissenschaften, Germany*

Volume 112

Maximum Entropy, Information Without Probability and Complex Fractals

Classical and Quantum Approach

by

Guy Jumarie

Department of Mathematics,
University of Quebec at Montreal,
Montreal, Quebec, Canada

SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-90-481-5467-8 ISBN 978-94-015-9496-7 (eBook)
DOI 10.1007/978-94-015-9496-7

Printed on acid-free paper

All Rights Reserved
© 2000 Springer Science+Business Media Dordrecht
Originally published by Kluwer Academic Publishers in 2000
Softcover reprint of the hardcover 1st edition 2000

No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the copyright owner.

So all my best is dressing all words new,
Spending again what is already spent.

Shakespeare (Sonnet LXXVI)

All that is straight tells lies, the dwarf murmurred scornfully. All truths are bent, and time itself is a circle.

Nietzsche (Zarathustra (III,2))

Table of contents

Preface

Every thought is a throw of dice.
Stéphane Mallarmé

This book is the last one of a trilogy which reports a part of our research work over nearly thirty years (we discard our non-conventional results in automatic control theory and applications on the one hand, and fuzzy sets on the other), and its main key words are Information Theory, Entropy, Maximum Entropy Principle, Linguistics, Thermodynamics, Quantum Mechanics, Fractals, Fractional Brownian Motion, Stochastic Differential Equations of Order n, Stochastic Optimal Control, Computer Vision. Our obsession has been always the same: Shannon's information theory should play a basic role in the foundations of sciences, but subject to the condition that it be suitably generalized to allow us to deal with problems which are not necessarily related to communication engineering. With this objective in mind, two questions are of utmost importance:

(i) How can we introduce meaning or significance of information in Shannon's information theory?

(ii) How can we define and/or measure the amount of information involved in a form or a pattern without using a probabilistic scheme?

It is obligatory to find suitable answers to these problems if we want to apply Shannon's theory to science with some chance of success. For instance, its use in biology has been very disappointing, for the very reason that the meaning of information is there of basic importance, and is not involved in this approach.

Starting from concerns which originated from systems engineering, our first book *Subjectivity, Information, Systems. Introduction to a Theory of Relavistic Cybernetics* (Gordon and Breach, 1986) laid the foundations of a cybernetic theory which involves explicitly the point of view of the observer. A system S does not have an absolute definition, and is defined only with respect to an observer R, say (S/R).

But this monograph was only a general model of system structures (something like the theory of categories in mathematics), and we had to construct a quantitative framework to obtain more insight in this relativistic feature of systems. This has been

done in our second monograph *Relative Information. Theories and Applications* (Springer Verlag, 1990) in which, by using considerations related to the linguistics of natural languages, for the first time in the literature, we stated a principle of informational invariance or of invariance of information, which provided a model of entropy involving explicitly both syntax and semantics.

The present book displays the results which we obtained in trying to define information without probability, and which,incidentally, led us to quantum mechanics and fractals. Indeed, starting from Shannon's theory we defined the entropy of non random functions and the quantum entropy of non-probabilistic matrices, and then we came across some connections between these models and fractals. Surprisingly, as in a photography laboratory, the finishing revealed the hidden presence of the maximum entropy principle in each film. The relation between the entropy of non-random functions and fractals drew our attention to the modelling of fractional Brownian motion, and by using the complex roots of the unity, we found a new model of complex-valued fractional Brownian motion of order n defined by using a random walk in the complex plane. The point of importance is that one of the assumptions at the basis of this approach is quite similar to Heisenberg's principle in quantum mechanics, with prospects that the reader can guess.

This books arrives at two main conclusions..

The first is that *the so called maximum entropy principle can be used as a unified approach to so many questions, such as the definition of the Shannon entropy of random variables, the definition of the entropy of non-random functions, the definition of the quantum entropy of non-probabilistic matrices, the determination of the solution of the Fokker Planck Kolmogorov equation, the definition of complex-valued fractional Brownian motion of order n, therefore some prospects in the thermodynamics of systems involving fractals.*

The second conclusion is that *the presence of fractals in science should be much more important than suspected at first glance, and so from the simplicity and universality of the standard conditions (quantization in magnitude and quantization in angle) which are required for their generation.* Fractals have been identified in an incredible number of areas (the French Breton coast, mountains, Brownian motion, lung structure, the lunar geography, the solar spectrum, the brain, the entire vascular system, hadrons, asteroids, Navier Stokes equations, the stock exchange, ...) *and our claim is that we should expect to find them elsewhere repeatedly. In addition, fractal dimension is directly related to the entropy of non-random functions, therefore a new framework for the thermodynamics of fractal structures.*

The prequisites which iare necessary for reading and understanding this book is very elementary. The main definitions of information theory, a basic course in stochastic processes, the basics of square matrices and differential equations. Basically its mathematical framework is what is usually scornfully referred to as engineering mathematics, having in mind that we shall have time later to put all the material in the context of abstract mathematical spaces (i.e. Hilbert space, the Radon integral, filters, Wiener measure, Borel measure, cylindrical measure in infinite dimensional spaces, the probabilistic triple (Ω,μ,σ), martingales, and the like). Here, we want to attract as large a readearship as possible, and thus the mathematical framework is summarized as far as possible.

The main contributions of the book are (i) a theory of information of non-random functions and its applications to random functions, which provides a new concept of the random path entropy of stochastic processes; (ii) a theory of quantum information for non-probabilistic square matrices; and (iii) a theory of complex-valued fractional Brownian motion defined via the limit of a random walk in the complex plane. With many applications such as, for instance, computer vision, fractals in human systems, thermodynamics of order n, Fokker-Planck equation of fractional order, stochastic optimal control of order n.

This book has been written while the author, who is a French citizen from the West Indies or Caribbean Islands (Guadeloupe) has been teaching in a Canadian university. All help and support came from outside Canada (mainly the USA, France, Germany, England and UNESCO). This book was typed on a friendly twelve year old Macintosh IIsi computer given by a friend instead of throwing it out in his garbage.

I am very pleased to express my gratitude to Kluwer and to Professor Van der Merwe for their welcome cooperation.

Montréal, September 1999 *Guy Jumarie*

Chapter 1

Introduction

Systems are not in Nature, but in Man's mind
Claude Bernard

1.1. Cybernetics, control, systems

Strictly speaking, "cybernetics" deals with the art of managing (or of governing), you manage something, say a "system", and very earlier the researches in cybernetics have been splitted into two parts: namely automatic control theory and applications on the one hand, and systems science on the other hand. But whilst the purpose of automatic control is well defined and its results were quite convincing, we cannot say that it is the same for systems science. As a matter of fact, to the best of our understanding, the content of this research area has never been exactly defined, or at least, there is no common agreement about what it should be. In such a manner that, very earlier, this term turned to be somethink like a bag where one could put anything such as fuzzy sets, information, artificial intelligence, computer science, graph theory, control, problems of organization and structure, and the like.

At the beginning, the purpose of "systems science" was quite well defined. One had to develop a general theory of structures, in order to unite it with a suitable theory of control later. But by this way, the various theories so constructed were too general, too formal to be of some use in practical real problems. Most papers on this topics were verbose papers only.

In a first approach, loosely speaking, a system is a given subset of a given physical universe. In a second approach, see for instance Ackoff et al [1.1], a system can be thought of as a set of interrelated elements of a given universe, each of which being related directly or indirectly to every other element, and furthermore with no subset (or part) being unrelated to any other subset. Strictly speaking, one would like to build up a unified theory which would encompass such various components as biological systems, urban systems, societal systems, economical systems, engineering systems, and so on. But in order to achieve this goal, we have firstly to identify the common features of these different systems (if there is any), to be able to construct the theory in terms of the corresponding variables.

G. Jumarie, *Maximum Entropy, Information Without Probability and Complex Fractals*
© Springer Science+Business Media Dordrecht 2000

Very earlier I have been led to assume that this Ariane's thread would be *information*, and I suggested the following axioms.

(A1) A system S alone is meaningless, and it makes sense only with respect to its environment $\overline{S}$ with which it exchanges information.

(A2) A system $(S, \overline{S})$ is defined only with respect to a given observer R, say *(S/R)* and $(\overline{S}/R)$.

(A3) A system is characterized by the amounts of uncertainty $H_i(S/R)$ and $H_0(S/R)$ that the observer R has about its internal structure and the internal structure of its environment, respectively. And the knowledge of R about $(S, \overline{S})$ will evolve in such a manner that the following equation holds,

$$H_i^2(S/R) - \gamma^2 H_0^2(S/R) = constant, \tag{1.1}$$

where γ is a positive constant which depends upon the measurement units.

Cave canem. The assumtion (A3) has not been chosen for convenience only, by merely using an obvious analogy, but rather it has been suggested by a careful analysis of real systems. Firstly, simple remarks of a physical nature show that, for most real systems, one has the inequality

$$H_i(S/R) \geq \gamma H_o(S/R). \tag{1.2}$$

Furthermore, our uncertainty $H_i(S/R) + \gamma H_0(S/R)$ about the system increases when our information $H_i(S/R) - \gamma H_0(S/R)$ about this same system decreases, therefore the modelling

$$\left(H_i(S/R) + \gamma H_0(S/R)\right)\left(H_i(S/R) - \gamma H_0(S/R)\right) = constant. \tag{1.3}$$

Of course, we could, for instance, select the model

$$H_i^2(S/R) + \gamma^2 H_0^2(S/R) = constant, \tag{1.4}$$

but this equation corresponds to unlearning systems whilst the preceding one (1.1) describes learning systems. For further details see [1.4].

1.2. Relative information

It was evident that Shannon information theory could not be successfully used in the analysis of general systems for the very reason that it does not account for the meaning, the significance, of information.

Basically, Shannon's theory deals with encoding. We have at hand two alphabets $A = \{\alpha\}$ and $B = \{\beta\}$ of which the generic letters are respectively denoted by α and β, with the additional condition $|A| > |B|$ (clearly A contains more letters than B); and we want to use the letters of B to encode those of A. As an illustrative example, the classical problem is to encode the latin alphabet $(a,b,c,..)$ with the computer alphabet which contains only the symbols "1" and "0". Shannon's theory analyzes some properties of this problem, and mainly provides the exact value of the minimum number of letters of B which are necessary for encoding each letter of A.

Obviously the exchange of information between a system and its environment can be thought of as an encoding process, but in the general case a system does not read symbols only as an automaton, but rather refers to linguistic information, that is to say, information involving both syntax and semantics. A system can use a given amount of information when and only when this information has some meaning in it. In other words, a source of information is available, and the system can use it or not, depending upon the relative significance of this information.

To some extent, the meaning of information is already taken into account by the probability function involved in the theory, but this is only partly true. Indeed, consider the following sentence: "The long strive of the workers undermined the economy of the country". In terms of probability distribution of the letters, this sentence is quite likely, but from a practical standpoint it is confusing and meaningless. Nevertheless, a reporter who is (already) acquainted with the general topic of the message will correct it by himself (to read "the long strike of the workers..."), whilst another reader will not be able to recover the initial message as it was issued. This prior knowledge (or lack of knowledge) results in some subjectivity in the observer, and an information theory for general systems should take account of this feature in order to deal with subjective information.

Starting from remarks suggested by the linguistics of natural languages, we have been led to describe a language which conveys a message by means of a syntactic space and a semantic space, and the coupling effects between these two spaces has resulted in a

modelling of subjective entropy which, incidentally, is expressed by equations of the special relativity. For further details see Chapter 2.

1.3. Complexity, organization, entropy without probability

Another problem which drew the attention of researchers is the analysis of the various relations which might exist between complexity, organization, and self-organization of systems, all being terms, which, once again, are not yet clearly defined. Systems composed of many subsystems can produce macroscopic spatial, temporal or functional structures with increasing structure complexities, that is to say, with structures which require an increasing number of parameters in order to be fully described. This can be found in fields as different as cosmology, linguistics, communication networks, sociology, fluid mechanics, paleontology, atmospheric science, and so on.

Problems of complexity and of self-organization have usually been addressed by using the concept of informational entropy. Entropy is introduced as a statistical concept applicable to an ensemble, but not to the individual members or events comprising them. This attraction of the entropic approach came from the evident analogy with thermodynamics, but using entropy exhibits some flaws.

(i) It is not possible to obain measure of macroscopic entropy by taking the ensemble average of some microscopic variables (exactly the opposite of what happens with temperature);

(ii) entropy is related to probability together with uncertainty, and consequently requires the randomization of the problem under consideration.

Strictly speaking, the first remark above is not wrong, but nevertheless it is possible to derive at least approximate composition laws for entropy in order to partially circumvent this drawback [1.5]. As a matter of fact, a system can be characterized either at the microscopic, mesoscopic, or macroscopic level, and one of the problems of systems theory is to determine how the microscopic model can provide data on the mesoscopic one, and in the same way, to determine the kind of relations which might exist between the mesoscopic model, and, the macroscopic approach. Nevertheless, when one observes a system from outside, such as a black box, mainly (or perhaps only) the macroscopic approach is meaningful, in such a manner that entropy is quite relevant.

Entropy is used in organization and self-organization problems mainly via Jaynes maximum entropy principle [1.2][1.3]. If we want to extend this approach we necessarily need a generalized model of entropy defined without probability. For instance, a problem

of interest is the following one: How can we measure the amount of information involved in a non-random mapping, or a pattern, or a form?

With such a generalization we shall have at hand entropy of random data on the one hand and entropy of non-random data on the other hand, and one can expect to tackle self-organization problems with more efficient tools.

By using the maximum conditional entropy principle we obtained a model for this entropy of non-random functions, and we shall display it in the following. Loosely speaking, the main feature of the model can be summarized as follows: the density of the amount of uncertainty involved in a one-dimensional time function $x(t)$ is $\ln|dx(t)/dt|$. Or again, the dynamical equation of information associated with the dynamics

$$\Delta x(t) = f(x)\Delta t \tag{1.5}$$

is

$$U(\Delta x) = \ln|f(x)| + U(\Delta t) \tag{1.6}$$

where $U(\Delta x) := \ln|\Delta x|$ is the amount of uncertainty involved in Δx.

1.4. Information and fractals

Our definition of the entropy of non-random functions applies to continuously differentiable functions only, and we have extended it easily to functions which are differentiable almost anywhere. In quite a natural way we have been led to ask what happens for functions which are continous everywhere, but are nowhere differentiable, that is to say, for fractal functions.

A one-dimensional function of time $x(t)$ is referred to as a fractional (or fractal) function if it satisfies the condition

$$|\Delta x| \propto (\Delta t)^{\alpha} \quad , \quad \alpha \in (0,1), \tag{1.7}$$

and, acccording to Mandelbrot [1.6], a fractional stochastic process $x(t)$ is characterized by the property

$$Var\{\Delta x\} = (\Delta t)^{\alpha} \quad , \quad \alpha \in (0,1) \tag{1.8}$$

Mandelbrot considered fractional Brownian motion in the form of integral of Gaussian white noises related to the theory of fractional derivatives; but this model defines

long range dependence processes. And thus it describes a process which is fundamentally different from classical Brownian motion.

Thus, one may ask whether one could not derive another model which would have independent increments. This would allow us to expand a theory parallel to that of classical Brownian motion by using Levy's ideas, and mainly to develop a new theory of stochastic differential order of order n, which would be the direct generalization of the theory of Itô's stochastic differential equations.

In order to achieve this objective, we need a stochastic process which satisfies the conditions

$$E\left\{(\Delta x)^j\right\} = 0, \quad j = 1,2,..,n-1, \tag{1.9}$$

$$E\left\{|\Delta x|^n\right\} \propto \Delta t. \tag{1.10}$$

and, fortunately, it is provided by the complex roots of unity.

1.5. How we shall proceed?

In the following we shall display the results which we obtained in the investigation of all these problems.

For the sake of completeness and also for the convenience of the reader, in the second chapter we shall bear in mind the essentials of information theory. In the third chapter, we shall see how, by using the maximum *conditional* entropy principle, one can meaningfully define the informational entropy of a non-random function. This entropy will be defined for non-random functions and for non-random distributed functions. In the fourth chapter, we shall use this concept to derive a new concept of path entropy for stochastic processes and distributed stochastic processes. We shall refer to the entropy of a trajectory sample of the process, which is then a random entropy, in such a manner that the stochastic process will be characterized by a family of random entropies.

In Chapter 5, by using the rationale of Chapter 3, we shall obtain a model for the entropy of non-probabilistics square matrices, that is to say matrices which do not necessarily have a meaning in terms of density matrices, as in quantum mechanics. And in order to emphasize the unified approach so provided by the maximum entropy principle, we show that, to some extent, the definition of Shannon entropy itself can be obtained as a consequence of this principle.

The Chapters 6 and 7 deal with complex-valued Brownian motion of fractional order. We firstly construct this process by means of a Gaussian white noise rotating on the finite grid defined by the complex roots of unity, and then we show how it can be

obtained as the limit of a random walk in the complex plane, a random walk which is also defined by the complex roots of unity. And then we examine some of the consequences of this new model: Itô's stochastic differential equation of order n, Itô's lemma of order n, the Kramer Moyal expansion, the Feynman Kac formula, the relation with Heisenberg's principle in quantum mechanics. The point of importance is that the assumptions which are made to construct this complex-valued fractional Brownian motion are very close to Heisenberg's principle.

Chapter 8 examines how one can use these results to construct a thermodynamics in the complex plane. We put in evidence a relation between the maximum entropy principle and the complex-valued Brownian motion of order n, and one exhibits the meaning of the latter in terms of collisions of more than two particles.

In the Chapter 9, we shall apply path entropy to (deterministic and stochastic) dynamical systems. We shall put in evidence an equivalence between the heat equation of order n and the fractional Fokker Planck equation of order $1/n$, and we shall so arrive in quite a natural way at a model of the fractional Fokker Planck equation.

Chapter 10, which can be thought of as a conclusion, outlines some prospects of applications of these theories: the stability of stochastic systems subject to white noises of order n, fractional time series in the complex plane, computer vision, the solution of the master equation, optimal control of systems subject to complex-valued fractional white noises. And lastly we show that, to some extent, the presence of fractional models in dynamical systems is not too surprising, since they define systems which evolve with loss of information.

1.6. How to read the book

The book can be considered to comprising two parts which are respectively Entropy of Non-Random Data (Part I, Chapters 2,3,4) and Complex-Valued Fractional Brownian Motion (Part II, Chapters 5-10). There is some interference between them, but they can be considered separately and independently from each other. So various reading schemes are possible, depending upon the interests of the reader.

(i) scheme (1,2,3,4,5,6,7,8,9,10): the ideal scheme to follows the author ideas.

(ii) scheme (1,2,3) for the reader who is interested in information of non-random functions only.

(iii) scheme (1,2,4): the reader who is interested in the quantum entropy of non-probabilistic square matrices.

(iv) scheme (1,5,6,7,8,9,10): the reader who is interested in complex-valued fractional Brownian motion only and has time to consider the applications.

(v) scheme (1,5,6,7,8,9): the reader who is interested in complex-valued fractional Brownian motion only and has no time to consider the applications.

(vi) scheme (1,5): the reader in haste who is interested in the definitions and the properties of complex-valued fractional Bronian motion, only.

(vii) scheme (1,5,6): complex-valued fractional Brownian motion and applications to stochastic calculus of order n.

(viii) scheme (1,5,7): the reader who is interested in informational thermodynamics in the complex plane.

(ix) scheme (1,2,5,8): reader who is interested on the relations between entropy of non-random functions and complex fractals.

Warning to the reader. As in computer programming, in all the book the variables are locally defined only, and their meaning may change from one chapter to another.

References

1.1. ACKOFF, R.L. and EMERY, F.E.; *On Purposeful Systems,* Tavistock, London, 1972
1.2. JAYNES, E.T.; Information theory and statistical mechanics, *Physical Review*, Vol 106, p 620-630, 1957
1.3. JAYNES, E.T.; Information theory and statistical mechanics, *Physical Review*, Vol 108, pp171-190, 1957
1.4 JUMARIE, G.; *Subjectivity, Information, Systems. Introduction to a Theory of Relativistic* Cybernetics, Gordon and Breach, London, New York, 1986
1.5. JUMARIE, G.; Composition laws for entropy and temperature in tree-like graphs, *Math. Comput. Modelling,* Vol 18, No 5, pp 107-118, 1993
1.6. MANDELBROT B.B.; *The Fractal Geometry of Nature*, W.H. Freeman, New York, 1977, reprint, Plenum Publishing, New York, 1992
1.7 WEINER, N.; *Cybernetics*, 2nd edition, John Wiley, New York, 1961

Chapter 2

Summary of Information Theory

As far as the laws of mathematics refer to reality they are not certain, and as far as they are certain they do not refer to reality.

Einstein

2.1. Introduction

In the present chapter, for the convenience of the reader, we shall briefly call to mind the essentials of (Shannon) information theory. The three sections which follow deal with Hartley entropy, Shannon entropy, Renyi entropy, and the so called structural entropy of order s, for discrete probability distributions. One then defines informational divergence (or relative entropy or cross-entropy), again in the special case of discrete probabilities. All these concepts are formally extended to continuous probability densities. There is a gap between the entropy of discrete probabilities and the entropy of continuous probabilities, in the sense that the latter cannot be considered as the limiting value of the former (when the discretising span tends to zero). One shows how the concept of total entropy of discrete random variables, which we introduced recently, can circumvent this difficulty and unify the two theories. Next, we comment on the maximum entropy principle. And finally, for the sake of completeness, we outline the so called theory of relative information which we proposed a few years ago.

2.2. Hartley entropy

Consider a random experiment α which provides the random issues A_1, $A_2,...,A_m$ with the respective probabilities $p(A_1) =: p_1$, $p(A_2) =: p_2$, ..., $p(A_m) =: p_m$, $p_1 + p_2 + ... + p_m = 1$ (the symbol =: means that the right side is defined by the left side whilst := means that the left side is defined by the right). We shall shorten the notation to the form $\alpha := (p_1,p_2,...,p_m)$. Assume that we are going to carry out this experiment α. Can we guess in advance which outcome A_i will be so obtained? In information theory the expression *entropy of* α refers to the (or a) quantitative measure which characterises the degree of difficulty of this prediction. In other words, it defines the amount of uncertainty involved in α.

G. Jumarie, *Maximum Entropy, Information Without Probability and Complex Fractals*
© Springer Science+Business Media Dordrecht 2000

DEFINITION 2.1 Following Hartley [2.2], the entropy $H_H(\alpha)$ of the random experiment α is defined by the expression

$$H_H(\alpha) := \ln m , \tag{2.1}$$

where the natural logarithm is chosen for convenience only, and is not compulsory at all.■

This expression can be obtained as a result of the following properties which a suitable measure of uncertainty should satisfy:

(A1) The entropy is a function $f(m)$ of m only.

(A2) $f(m)$ is an increasing function of m.

(A3) $f(m)$ should satisfy the equation $f(m.n) = f(m) + f(n)$.

2.3. Shannon entropy

2.3.1. Approach via axiomatic derivation

DEFINITION 2.2. Following Shannon [2.13],[2.14], the entropy $H(\alpha)$ of the random experiment α is defined by the expression

$$H(\alpha) := -\sum_{i=1}^{m} p_i \ln p_i \cdot \blacksquare \tag{2.2}$$

This expression can be obtained by using the following set of desiderata that a measure of information should likely satisfy.

(B1) $H(\alpha)$ is a function $U(p_1,p_2,...,p_m)$ of the probability distribution.

(B2) When $p_1 = p_2 =...= p_m = 1/m$, U is an increasing function of m.

(B3) Assume that a given outcome can be considered as a sequence of the two events A_{m-1} and A_m; then the complete uncertainty U should be the weighted combination of the partial uncertainties involved in the distributions $(p_1,p_2,...,p_{m-1}+p_m)$ and $(p_{m-1}/(p_{m-1}+p_m)\ ,\ p_m/(p_{m-1}+p_m))$ respectively. For instance, in the case of three events, one should have the equality

$$U(p_1,p_2,p_3) = U(p_1,p_2+p_3)+(p_2+p_3)U\left(\frac{p_2}{p_2+p_3},\frac{p_3}{p_2+p_3}\right) \tag{2.3}$$

2.3.2. Approach via the law of large numbers

The expression of the Shannon entropy can be also obtained as the result of a counting procedure combined with the law of large numbers. To this end, we shall consider the following problem.

PROBLEM. Assume that α is a generic symbol which randomly takes the values $A_1,A_2,...,A_m$ with the respective probabilities $p_1,p_2,...,p_m$. We want to encode the sequences $\alpha_1\alpha_2...\alpha_N$ of N such symbols, say, the N-length sequences, assuming that they are mutually independent (from the stochastic standpoint); and to this end we need to determine the number M of all possible sequences which are most likely to occur.

As an example, α can be a bit which takes on the values 0 and 1 with the respective probabilities p_1 and p_2, and we want to determine the number M of likely N-length sequences.

Solution. (i) First of all, we shall assume that each one of these N-length sequences occurs with the same probability $1/M$.

(ii) According to the law of large numbers, for N large enough, the N-sequences which will occur with maximum likelihood are those in which A_i appears $N_i = p_iN$ times for every i, in such a manner that, as result, one can write the equality

$$\frac{1}{M} \cong p_1^{N_1} p_2^{N_2} ... p_m^{N_m}, \tag{2.4}$$

$$\cong p_1^{Np_1} p_2^{Np_2} ... p_m^{Np_m}, \tag{2.5}$$

therefore we obtain the estimate

$$M \cong e^{NH(\alpha)}, \tag{2.6}$$

which explicitly involves the entropy $H(\alpha)$.

It may be of interest to recall the following basic properties.

(i) The (Shannon) entropy satisfies the inequality

$$H(\alpha) \geq 0, \tag{2.7}$$

and the equality holds when and only when the issue of α does not involve uncertainty, what amounts to say that α is a non random event.

(ii) $H(\alpha)$ achieves its maximum value lnm when and only when $p_1 = p_2 =...= p_m = 1/m$.

(iii) When α and β are two mutually independent random experiments, then the amount of uncertainty $H(\alpha,\beta)$ involved in the pair (α,β) satisfies the equality

$$H(\alpha,\beta) = H(\alpha)+H(\beta) \ . \tag{2.8}$$

2.3.3. Conditional entropy

When the experiments α and β above are not mutually independent the equation (2.8) may be re-written in terms of the conditional entropy. To this end, we proceed as follows.

(i) Assume again that α yields the issues $A_1, A_2,...,A_m$ with the probabilities $p(A_i) = p_i$, $i = 1,2,...,m$.

(ii) Assume that for each A_i, $i = 1,2,...,m$, β provides the results B_{ji}, $j = 1,...,n$ with the respective probabilities $p(B_{ji})$ which define the entropies

$$H(\beta|A_i) = -\sum_{j=1}^{n} p(B_{ji}) \ln p(B_{ji}), \quad i = 1,2,...,m \ . \tag{2.9}$$

(iii) DEFINITION 2.3. Define the *conditional entropy* $H(\beta|\alpha)$ *of* β *given* α by the expression

$$H(\beta|\alpha) := \sum_{i=1}^{m} p(A_i) H(\beta|A_i) \ . \ \blacksquare \tag{2.10}$$

(iv) Then the entropy $H(\alpha,\beta)$ of the pair (α,β) satisfies the equality

$$H(\alpha,\beta) = H(\alpha)+H(\beta|\alpha) \ . \tag{2.11}$$

2.4. Renyi entropy and structural entropy

Shannon entropy has been extended in an incredible number of so-called generalized entropies, and we shall herein only mention the Renyi entropy and the structural entropy which can be of valuable help in some applications.

DEFINITION 2.4. The *Renyi entropy of order s* of the random experiment α is defined by the expression [2.11],[2.12]

$$H_{R,s}(\alpha) := \frac{1}{1-s}\ln\sum_{i=1}^{m} p_i^s, \quad s \geq 0, \quad s \neq 1, \tag{2.12}$$

with the convention

$$0^s = 1 \ , \ s \in \Re. \ \blacksquare \tag{2.13}$$

$H_{R,s}(\alpha)$ achieves its maximum value $ln\ m$ when $p_1 = p_2 = \ldots = p_m = 1/m$; and moreover one has the following inequalities, i.e.

$$0 \leq H_{R,s}(\alpha) \lhd H(\alpha), \qquad s \rhd 1, \tag{2.14}$$

$$0 \leq H(\alpha) \lhd H_{R,s}(\alpha), \qquad 0 \lhd s \lhd 1, \tag{2.15}$$

in such a manner that if we consider Shannon emtropy $H(\alpha)$ as being the absolute (physical) amount of uncertainty involved in α then $H_{R,s}(\alpha)$ could be used as a measure of uncertainty in the presence of subjectivity, that is to say with coupling effects between symbols and meanings [2.8].

In addition, one has the following relations,

$$\lim H_{R,s}(\alpha) = H(\alpha) \quad as \quad s \to 0, \tag{2.16}$$

and

$$H_{R,s}(\alpha,\beta) = H_{R,s}(\alpha) + H_{R,s}(\beta), \tag{2.17}$$

provided that α and β are mutually independent.

DEFINITION 2.5. Following Havrda and Charvat [2,3] the *structural entropy of order s* of the random experiment α is defined by the expresion

$$H_{C,s}(\alpha) := \frac{\sum_{i=1}^{m} p_i^s - 1}{e^{1-s} - 1}, \quad s \geq 0, \quad s \neq 1. \blacksquare \tag{2.18}$$

The constant e is selected here mainly to be consistent with the natural logarithm involved in Shannon entropy; but if instead, $H(\alpha)$ is defined by a logarithm to the base a, then we shall subsitute a^{1-s} for e^{1-s} in equation (2.18).

$H_{C,s}(\alpha)$ is positive or zero, and satisfies the condition

$$\lim \quad H_{C,s}(\alpha) \quad = \quad H(\alpha) \quad as \quad s \to 1. \tag{2.19}$$

But here, in contrast to the Shannon entropy and the Renyi entropy, one has

$$H_{C,s}(\alpha,\beta) \quad \neq \quad H_{C,s}(\alpha)+H_{C,s}(\beta), \tag{2.20}$$

even when α and β are mutually independent. More explicitly, the joint structural entropy can be written in the form

$$H_{C,s}(\alpha,\beta) \quad = \quad H_{C,s}(\alpha)+\sum_{i=1}^{m} p_i^s H_{C,s}\left(\beta|A_i\right), \tag{2.21}$$

where $H_{C,s}(\beta/A_i)$ is the conditional structural entropy of β given A_i, i.e.

$$H_{C,s}\left(\beta|A_i\right) \quad := \quad \frac{\sum_{j=1}^{n} p^s\left(B_j|A_i\right)-1}{e^{1-s}-1}, \tag{2.22}$$

with $p(B_j|A_i)$ denoting the conditional probability of the issue B_j of β given A_i.

One can verify that

$$\leq \quad H_{Cs}(\beta), \quad s \geq 1 \tag{2.23}$$

$$\sum_{i=1}^{m} p_i^s H_{C,s}(\beta|A_i)$$

$$\geq \quad H_{C,s}(\beta), \quad s \leq 1 \quad . \tag{2.24}$$

in such a manner that one has the inequalities

$$H_{C,s}(\alpha,\beta) \quad \leq \quad H_{C,s}(\alpha)+H_{C,s}(\beta), \quad s \geq 1 \quad , \tag{2.25}$$

$$H_{C,s}(\alpha,\beta) \quad \geq \quad H_{C,s}(\alpha)+H_{C,s}(\beta), \quad s \leq 1 \quad . \tag{2.26}$$

which qualify the structural entropy as a valuable candidate to deal with information in the presence of subjectivity.

2.5. Informational divergence

DEFINITION 2.6. *Kullback divergence.* Let $\{p_i\}$ and $\{q_i\}$, $i = 1,2,\ldots,m$ denote two complete probability distribution (i.e. $\Sigma p_i = \Sigma q_i = 1$). Their *cross-entropy* or *relative entropy* or *divergence* $H(q,p)$ is defined by the expression (Kullback [2.16])

$$H(q,p) := \sum_{i=1}^{m} q_i \ln \frac{q_i}{p_i}. \quad \blacksquare \tag{2.27}$$

$H(q,p)$ satisfies the condition $H(q,p) \geq 0$, and the equality holds when and only when $p_i = q_i$ for all i.

On a practical (physical) standpoint $H(q,p)$ measures the deviation between the two probability distributions in the information-theoretic framework. Strictly speaking, from a theoretical standpoint, one could select any distance to measure this deviation, for instance

$$d(q,p) := \left[\sum_{i=1}^{m} (q_i - p_i)^2\right]^{1/2}, \tag{2.28}$$

but $H(q,p)$ has meanings in terms of information that $d(q,p)$ has not.

DEFINITION 2.7. *Renyi divergence.* The parallel of $H(q,p)$ in terms of Renyi entropy is defined by the expression [2.18]

$$H_{R,s}(q,p) := -\frac{1}{1-s} \ln \sum_{i=1}^{m} q_i \left(\frac{q_i}{p_i}\right)^{s-1} \quad \blacksquare \tag{2.29}$$

$H(q,p)$ is consistent with Shannon entropy, $H_{R,s}(q,p)$ is consistent with Renyi entropy and in addition one has the limit

$$\lim \quad H_{R,s}(q,p) = H(q,p) \quad as \quad s \to 1 \ . \tag{2.30}$$

2.6. Mutual information

The problem of defining mutual information can be stated as follows. Consider the two random experiments α and β of sub-section 2.2.3. The pair (α,β) yields the outcome (A_i,B_j) with the probability r_{ij}, $i = 1,2,...,m$, $j = 1,2,...,n$. When α and β are independent one has the equality $r_{ij} = p_i q_j$. In contrast, when $r_{ij} \neq p_i q_j$ then α and β are not mutually independent, and in terms of information we shall say that α (resp. β) involves information about β (resp. α). How can we measure this amount of information?

DEFINITION 2.8. *Shannon mutual information or transinformation.* Following Shannon, the amount of mutual information $I(\alpha,\beta)$ provided by α about β is defined by the expression

$$I(\alpha,\beta) = H(\beta) - H(\beta|\alpha) . \quad \blacksquare \tag{2.31}$$

As a result of the equalities

$$H(\alpha,\beta) = H(\alpha) + H(\beta|\alpha), \tag{2.32}$$
$$= H(\beta) + H(\alpha|\beta) . \tag{2.33}$$

the amount of information provided by α about β is equal to the amount of information involved in β about α, $I(\alpha,\beta) = I(\beta,\alpha)$, therefore the expression of mutual information which is used instead of information (alone).

Important remark. Many authors refer to $H(\beta)$ as a measure of information, but this is a bit confusing. Stricly speaking, by its definition, the entropy $H(\beta)$ is firstly a measure of uncertainty, but according to (2.31), it is also the maximum value of the mutual information $I(\alpha,\beta)$, therefore one is used to saying that $H(\beta)$ is the maximum value of the amount of information involved in β.

DEFINITION 2.9. *Renyi mutual information.* A simple calculation allows us to re-write $I(\beta,\alpha)$ in the form

$$I(\beta,\alpha) = \sum_{i,j=1,1}^{m,n} r_{ij} \ln \frac{r_{ij}}{p_i q_j} , \tag{2.34}$$

which clearly is the Kullback divergence

$$I(\beta,\alpha) = H(r,qp) . \tag{2.35}$$

This remark suggests using the Renyi divergence $H_{R,s}(r,qp)$ to define the Renyi mutual information in the form

$$I_{R,s}(\beta,\alpha) := -\frac{1}{1-s}\ln\sum_{i,j=1,1}^{m,n} r_{ij}\left(\frac{r_{ij}}{p_i q_j}\right)^{s-1}, \tag{2.36}$$

which provides the equality

$$I_{R,s}(\beta,\alpha) = I_{R,s}(\alpha,\beta).\blacksquare \tag{2.37}$$

On the definition of structural mutual information. Havrda and Charvat did not define the concept of mutual information associated with their structural entropy, and this is probably because they had not the corresponding (structural) divergence. We herein suggest two possible alternative which could be of some help in the applications.

(i) As a consequence of the equation (2.21) we generalize the (Shannon) conditional entropy $H(\beta|\alpha)$ in the form

$$H_{C,s}(\beta|\alpha) := \frac{\sum_{i=1}^{m} p_i^s H(\beta|A_i)}{\sum_{i=1}^{s} p_i^s}, \tag{2.38}$$

therefore the first model of structural mutual information

$$I_{C,s}(\beta,\alpha) := H_{C,s}(\beta) - H_{C,s}(\beta|\alpha)\,. \tag{2.39}$$

(ii) Another approach is to re-write the Shannon mutual information in the form

$$I(\beta,\alpha) = H(\alpha)+H(\beta)-H(\alpha,\beta),$$

and to generalize this expression to obtain

$$I_{C,s}(\alpha,\beta) := H_{C,s}(\alpha)+H_{C,s}(\beta)-H_{C,s}(\alpha,\beta). \tag{2.40}$$

2.7. Information and continuous probabilities

All the preceding formulae can be formally duplicated to deal with probability density functions, and the corresponding results can be summarized as follows.

Let $X \in \Re$ denote a continuous random variable with the probability density $p(x)$; its Shannon entropy is

$$H(X) := -\int_{\Re} p(x) \ln p(x) dx, \tag{2.41}$$

its Renyi entropy of order s is

$$H_{R,s}(X) := \frac{1}{1-s} \ln \int_{\Re} p^s(x) dx, \quad s \geq 0, \quad s \neq 1 , \tag{2.42}$$

and its structural entropy of order s is

$$H_{C,s}(X) := \frac{\int_{\Re} p^s(x) dx - 1}{e^{1-s} - 1}, \quad s \geq 0, \quad s \neq 1 \quad . \tag{2.43}$$

Given two probability densities $p(x)$ and $q(x)$ defined on $\Re$, their Kullback cross-entropy and their Renyi cross-entropy are respectively given by the expressions

$$H(q,p) := \int_{\Re} q(x) \ln \frac{q(x)}{p(x)} dx, \tag{2.44}$$

and

$$H_{R,s}(q,p) = -\frac{1}{1-s} \ln \int_{\Re} q(x) \left[\frac{q(x)}{p(x)} \right]^{s-1} dx, \quad s \geq 0, \quad s \neq 1 \quad . \tag{2.45}$$

Next, consider the random variables $X \in \Re$ and $Y \in \Re$ with the respective densities $p(x)$ and $q(y)$ and the joint density $r(x,y)$. The conditional entropy $H(Y|X)$ is defined by the expression

$$H(Y|X) := \int_{\Re} p(x) H(Y|x) dx, \tag{2.46}$$

with

$$H(Y|x) := -\int_{\Re} q_x(y) \ln q_x(y) dy, \tag{2.47}$$

where $q_X(y)$ is the conditional probability density of Y given that $X = x$.

With these notations, the Shannon mutual information $I(Y,X)$ and the Renyi mutual $I_{R,s}(X,Y)$ are respectively

$$I(Y,X) := H(Y) - H(Y|X), \tag{2.48}$$

$$= \int_{\Re^2} r(x,y) \ln \frac{r(x,y)}{p(x)q(y)}\, dxdy \quad , \tag{2.49}$$

and

$$I_{R,s}(Y,X) := -\frac{1}{1-s} \ln \int_{\Re^2} r(x,y) \left[\frac{r(x,y)}{p(x)q(y)} \right]^{s-1} dxdy \tag{2.50}$$

In the present reminder, for the sake of brevity, we have formally generalized the definitions related to discrete probabilities, but it is important to point out that the entropy of continuous probabilities is not the limiting form of the entropy of discrete probabilities when the discretizing span (of approximation) tends to zero, and it is exactly this remark which led us to introduce the following concept of *total entropy.*

2.8. Total entropy of discrete random variables

2.8.1. Total Shannon entropy

Preliminary notations. Let $X \in \Re$ denote a discrete random variable which takes on the values $x_1, x_2, \ldots, x_m$ with the respective probabilities $p_1, p_2, \ldots, p_m$. Define the interval lengths $h_i = 1, \ldots, m$ by the expressions

$$h_1 := x_2 - x_1, \tag{2.51}$$

$$h_i := \frac{x_i + x_{i+1}}{2} - \frac{x_i + x_{i-1}}{2}, \quad i = 2, \ldots, m-1 \quad , \tag{2.52}$$

$$h_m := x_m - x_{m-1} \quad . \tag{2.53}$$

With these notations, we have the following

DEFINITION 2.10. The total Shannon entropy $H_e(X)$ of the discrete random variable above is defined by the expression [2.12],[2.14]

$$H_e(X) := -\sum_{i=1}^{m} p_i \ln \frac{p_i}{h_i}. \tag{2.54}$$

which yields

$$\lim \quad H_e(X) \quad = \quad H(X) \quad as \quad \max_i h_i \downarrow 0. \blacksquare \tag{2.55}$$

Theoretical derivation . The intuitive meaning of equation (2.54) is that the abscissa x_i of the point M_i is not exactly known, and as a result, we shall assume that M_i is running on the interval L_i with the uncertainty $ln\ h_i$.

On a theoretical standpoint, $H_e(X)$ can be obtained as a consequence of the following set of desiderata [2.14].

(C1) $H_e(X)$ is a function $\chi[(p_1,h_1),(p_2,h_2),...,(p_m,h_m)]$, of which the value should not be modified by any permutation on the set $\{(p_1,h_1),(p_2,h_2),...,(p_m,h_m)\}$.

(C2) $\chi(.)$ is continuous with respect to p_i and h_i, for every i, except possibly at $h_i = 0$.

(C3) $\chi(.)$ is an increasing function of h_i for every i.

(C4) Let $\phi(p_1,p_2,...p_m)$ denote the Shannon entropy $H(X)$ of X; then the following condition is satisfied,

$$\chi\left[(p_1,1),(p_2,1),...,(p_m,1)\right] \quad = \quad \phi(p_1,p_2,...,p_m). \tag{2.56}$$

(C5) Let $\{(q_1,h_1'),(q_2,h_2'),...,(q_m,h_m')\}$ denote the parameters of another random variable Y which is independent of X, one then has the equality

$$\chi\left[(p_1q_1,h_1h_1'),...,(p_iq_j,h_ih_j'),...,(p_mq_n,h_mh_n'\right]$$

$$= \quad \chi\left[(p_1,h_1),...,(p_m,h_m)\right] + \chi\left[(q_1,h_1'),...,(q_n,h_n')\right] . \tag{2.57}$$

Alternative set of desiderata. An alternative set of axioms which can be used to obtain the mathematical expression of $H_e(X)$ is the following one.

(D1), (D2), (D3) similar to (C1), (C2), (C3) respectively.

(D4) The following functional equation is satisfied, that is

$$\chi\left[(p_1,h_1),...,(p_m,h_m)\right] \quad = \quad \chi\left[(p_1+p_2,1),(p_3,h_3),...,(p_m,h_m)\right] +$$

$$+(p_1+p_2)\chi\left[\left(\frac{p_1}{p_1+p_2},h_1\right),\left(\frac{p_2}{p_1+p_2},h_2\right)\right]. \tag{2.58}$$

Application. This total entropy has not been introduced just to make mathematics, but rather it is relevant whenever we want to take account of the spatial representation of the random variable, and, for instance, it will be of special interest in pattern recognition.

2.8.2. Total Renyi entropy

Here, for the sake of simplicity, we shall consider a uniform interval length defined by the expression

$$h := \lim \frac{x_{j+k}-x_j}{k} \quad as \quad k \uparrow +\infty. \tag{2.59}$$

DEFINITION 2.11. The total Renyi entropy of the discrete random variable X above, is defined by the expression

$$H^e_{R,s}(X) := \frac{1}{1-s}\ln\sum_{i=1}^{m}\left(\frac{p_i}{h}\right)^s h, \quad s \geq 0, \quad s \neq 1, \tag{2.60}$$

which yields

$$\lim H^e_{R,s}(X) = H_{R,s}(X) \quad as \quad h \downarrow 0.\blacksquare \tag{2.61}$$

Theoretical derivation. This definition can be obtained as a result of the following set of desiderata.

(E1) $H^e_{R,s}(X)$ is a function $\theta(p_1,p_2,...,p_m;h)$ of which the value should not be modified by any permutation on the set $\{p_1,p_2,...,p_m\}$.

(E2) $\theta(.)$ is continuous with respect to $p_1,p_2,...,p_m$ and h, except possibly at h = 0.

(E3) $\theta(.)$ is an increasing function of h.

(E4) The following relation is satisfied, that is

$$\theta(p_1q_1,...,p_iq_j,...,p_mq_n;hh') = \theta(p_1,p_2,...,p_m;h) + \\ +\theta(q_1,q_2,...,q_n;h') \quad . \tag{2.62}$$

Here again, this total Renyi entropy will be relevant whenever we shall have to take account of the spatial representation of the random variable.

2.9. The maximum entropy principle

2.9.1. Statement of the principle

PROBLEM. Suppose we know that a system has a set of possible states x_i, $i = 1,2,...,m$,with unknown probabilities $p(x_i)$, and we learn that the distribution p satisfies some constraints, for instance given values of some expectations $\Sigma_i p(x_i)f(x_i)$, or the range of variation of these values. Suppose that we need to define a distribution p^* which is, in some sense, the best estimate of p given these constraints. How should we proceed to determine p^*?

MAXIMUM ENTROPY PRINCIPLE. According to the maximum entropy principle (first proposed by Jaynes [2.4,2.5]), of all the family of distributions which satisfy these constraints we should select that one with the largest (Shannon) entropy.■

This principle has been successfully applied to an incredible variety of fields such as statistical mechanics and thermodynamics, statistics, reliability estimation, traffic network, queuing theory and computer system modelling, pattern recognition, image processing, system simulation, production line decision making, computer memory reference patterns, system modularity, group behaviour, stock market analysis, general probabilistic problem solving and spectral analysis (see for instance [2.21] for an extensive bibliography). More recently, Haken [2.1] took it as the basic principle of a macroscopic approach to complex systems, and we used it to solve the Fokker-Planck equation [2.13] and the conditional probability density partial differential equation which occurs in nonlinear filtering of Markov processes [2.15].

Possibly the most popular (and somewhat surprizing) result which is provided by this principle is the following. Assume that all we know about a probability density $p(x)$ is its mean μ and its standard deviation σ; then as the best estimate $p^*(x)$ of $p(x)$, we should select the density of the Gaussian (or normal) law.

2.9.2. On the validity of the maximum entropy principle

Physical arguments. As so stated by Jaynes, the "maxent principle" is truly a principle, and all we can expect is to have sound arguments to support it. One of them is

its striking similarity with the second law of thermodynamics, in accordance with the thermodynamic entropy S of an isolated system is a non-decreasing function of time (with the convention that any amount of heat energy received by the system is positive). As a matter of fact, S will first increase, and then will achieve a constant value which corresponds to the equilibrium state of the system.

Assume now that we refer to the Boltzmann equation

$$S = kH \tag{2.63}$$

(where k is the so-called Boltzmann constant) to identify S with H, in which case H turns out to be the informational entropy associated with the structural definition of the system; then, according to the second thermodynamic principle, H is non-decreasing.

The relation with Jaynes maximum entropy principle is then transparent. If we assume that the sought probability distribution p together with the given mathematical constraints define an isolated system, then it is quite reasonable to consider more especially the probability distribution which characterizes the equilibrium position of this system.

Mathematical arguments. Recently Shore and Johnson [2.21] proposed an axiomatic derivation of the maximum entropy principle, which assumes that reasonable methods of inductive inference should lead to consistent results when there are different ways of taking the same information into account. They *proved* that this principle is quite correct in the following sense: maximizing any function but entropy will lead to inconsistency unless this function and the entropy have identical maxima. Loosely speaking, their axioms read as follows:

(i) Uniqueness. The result should be unique.

(ii) Invariance. The choice of co-ordinate system should not matter.

(iii)System independence. It should not matter whether one accounts for independent information about independent systems separately in terms of different densities or together in terms of joint density.

(iv) Subset independence. It should not matter whether one treats an independent subset of system states in terms of a separate conditional density or in terms of the full system density.

Nevertheless, Uffink [2.22],[2.23] made an excellent criticism of this paper, and he showed that a possible way of circumventing the problems so involved in this approach could be to apply the maximum entropy principle to the Renyi entropy. But then if we make s tends to the unity in the result so obtained.....

In the next section we shall describe the main elements of a theory of relative information which we proposed a few years ago to take account of subjectivity in communication.

2.10. Relative information

2.10.1. Natural language and communication

Any natural language can be considered as defined by a set $\Omega := \{\alpha\}$ of symbols (one can call them words to fix our thoughts) with a generic element denoted by α, and a set $\Omega' := \{a\}$ of objects, or meanings, of which a generic element is a.The internal structure of Ω is governed by the syntax of this natural language whilst the structure of Ω' defines the semantics of the latter. There is a correspondence between the elements of Ω and those of Ω' which defines a function in the mathematical sense of this term. This function might be a one-to-one mapping, but in most cases it is a multi-valued mapping: a word may have several meanings. The relation between word and meaning is referred to as a *lexem*: in linguistics: loosely speaking, a lexem is a word taken with a given meaning. Formally, a natural language is a mapping of the syntactic space Ω on the set of subsets of the semantic space Ω'; and with this formal definition, it is possible to define the equivalence of natural languages by using a commutative diagram similar to the basic diagram of the category theory.

2.10.2. Subjectivity

We shall consider, as being a result of subjectivity, the capability that an observer has of interpreting, via reference to a prior internal model, the message which he is receiving; and to recover its content, in such a manner that the meaning which he ascribes it is not the observed meaning as received, but rather what the observer thinks it should be.

In this way, a human being can correct a message which involves transmission errors, but it can also ascribe a wrong sense to a message which is however correctly transmitted. Nevertheless, according to common experience, positive subjectivity (that is to say that one which corrects correctly) works in such a manner that uncertainty decreases when information increases, and conversely, uncertainty increases when information decreases. This property is the basis of our model.

2.10.3. Equations of the model

In the following the symbol *H(.)* refers to Shannon entropy.

DEFINITION 2.12 We shall say that the amount of information involved in the natural language $L:\Omega \to \Omega'$, is defined by $H(a)+H(\alpha)$. ■

DEFINITION 2.13 In a like manner, we shall assume that the amount of information involved in the same natural language $L:\Omega \to \Omega'$ is equal to $H(a) - H(\alpha)$. ■

For natural language one generally has $H(a) \geq H(\alpha)$.

In the presence of subjectivity the observer R does not measure the actual entropies $H(a)$ and $H(\alpha)$ but rather the *subjective entropies* $H(a|R)$ and $H(\alpha|R)$,which depend explicitly upon him. In order to write the variation equation of these subjective entropies, we introduce the quantities

$$\sigma^2(L) := H^2(a) - H^2(\alpha), \tag{2.64}$$

and

$$\sigma^2(L|R) := H^2(a|R) - H^2(\alpha|R), \tag{2.65}$$

which are suggested by the identity

$$x^2 - y^2 = (x-y)(x+y). \tag{2.66}$$

and we propose the following principle.

PRINCIPLE OF OBSERVATION WITH INFORMATIONAL INVARIANCE. Assume that a message, expressed by means of the natural language $L:\Omega \to \Omega'$, *is received and read by an observer R with some subjectivity. Then, in a large number of cases, this subjectivity will work in such a manner that uncertainty and information vary in opposite ways, which is defined by the invariance condition*

$$\sigma^2(L) = \sigma^2(L|R). \blacksquare \tag{2.67}$$

As a result, there exists a constant $u_\alpha(R)$ which depends upon both α and R, such that

$$H(\alpha|R) = \rho(u_\alpha)\left[H(\alpha) + u_\alpha H(a)\right], \tag{2.68}$$

$$H(a|R) = \rho(u_\alpha)\left[H(a) + u_\alpha H(\alpha)\right], \tag{2.69}$$

$$\rho(u_\alpha) := \left[1 - u_\alpha^2(R)\right]^{-1/2}. \tag{2.70}$$

For further information on this theory see the Refs [2.6] to[2.12] and[2.14]

2.10.4. On the validity of the model

In order to define the domain of validity of the model it is important to have in mind that the equations (2.68) and (2.69) are straightforward consequences of the invariance equation (2.67).

Indeed, it can be taken for granted without restriction that informaion and uncertainty vary in opposite ways, and it is exactly the modelling (2.67) of this property which provides the expressions (2.68) and (2.69) of the subjective entropies. Our assumption is that this discrimination between information and uncertainty is owed mainly to the *prior internal model* which the observer has about the message, which is referred to as his subjectivity, and which works necessarily on both syntax and semantics, that is to say on the natural language itself as a whole.

As a result, in order to describe this feedback between syntax and semantics we need a concept of global information and a concept of global uncertainty related to the natural language itself, hence the definitions 2.12 and 2.13, which are very simple and direct, and perhaps are the most natural.

It is clear that any other model which might be selected for describing the variation of the pair information, uncertainty would yield new equations different from the Lorentz equations, but then this would necessarily imply new assumtions which should be stated clearly.

The generalization of any theory in natural sciences should, of course, contain known results as a special case, but in addition it should explain and discover new phenomena to be observed later. Well obviously, here in our problem of modelling subjectivity, it is not easy to comply with these requirements, but nevertheless, the general purpose of the approach, loosely speaking, should be the same.

(i) When u = 0 equations (2.65) and (2.66) turn into be identities.

(ii) Assume that

$$H(a) = \rho' H(\alpha), \quad \rho' \triangleright 1, \tag{2.71}$$

then equations (2.68) and (2.69) yield respectively

$$H(\alpha|R) = \rho(1+u_\alpha \rho')H(\alpha), \tag{2.72}$$

$$H(a|R) = \rho\left(1+\frac{u_\alpha}{\rho'}\right)H(a). \tag{2.73}$$

In other words $H(\alpha|R)$ and $H(a|R)$ can be meaningfully identified with Renyi entropy. Or again, we herein have a framework which would allow us to determine the value of parameter s in the definition of Renyi entropy.

Let us rewrite the equation (2.68) in the form

$$-\sum_i p_i \ln q_i \;=\; -\rho \sum_i p_i \ln p_i - \rho u_\alpha \sum_i p_i H(a)\cdot \tag{2.74}$$

We then have the equality

$$\ln q_i \;=\; \rho \ln p_i + \rho u_\alpha H(a), \tag{2.75}$$

which provides

$$q_i \;=\; p_i^\rho \exp\{\rho u_\alpha H(a)\}. \tag{2.76}$$

We obtained this expression in 1987 [2.11]. At that time we referred to $\{q_i\}$ as a fuzzy-subjective probability distribution, and we suggested to considering it as an alternative to the modelling of the possibility distribution $\{q_i\}$. Later, the same transformation was re-discovered by another well known author in systems science and fuzzy sets, who used a so-called "information preserving transformation".

It is clear that the "information preserving transformation" and the "observation with informational invariance" are not different.

2.11. Concluding remarks

As we mentionned in the introductory section of this chapter, our main purpose has been only to recall some elements of information theory, and more especially the background which is required for the following.

The reader who is fully acquainted with the topic could be a bit surprised that we have found room for the structural entropy which, on the surface, is not so useful in the literature. The reason is that, owing to its properties, this entropy could be of valuable help in dealing with information in the presence of subjectivity.

References

2.1. HAKEN, H.; *Information and Self-Organization.* Springer Verlag, Berlin, 1988
2.2. HARTLEY, R.V.; Transmission of information. *Bell Systems Tech. J.,* Vol 7, pp 535-563, 1928
2.3. HAVRDA, J. and CHARVAT, F.; Quantification method of classification processes. Concept of structural entropy. *Kybernetica*, Vol 3, pp 30-35, 1957
2.4. JAYNES, E.T.; Information theory and statistical mechanics I. *Phys. Rev.*, Vol 106, pp 620-630, 1957
2.5. JAYNES, E.T.; Information theory and statistical mechanics II. *Phys. Rev.*, Vol 108, pp 171-190, 1957
2.6 JUMARIE, G.; Theory of relative information and telecommunication. Prospects (in french). *Ann. Telecom.*, (Fr.), Vol 33, Nos 1-2, pp 13-27, 1978
2.7. JUMARIE, G.; Theory of relative information II. Shannon information, Renyi entropy, fuzzy sets (in french). *Ann. Telecom.*, (Fr.), Vol 34. Nos 9-10, pp 491-507, 1979
2.8. JUMARIE, G.; Theory of relative information III. On the meaning and the use of of Renyi entropy in encoding problems (in french). *Ann. Telecom.* , (Fr.), Vol 34, Nos 11-12, pp 521-530, 1979
2.9. JUMARIE, G.; Theory of relative information IV. On the introduction of subjective factors in communication problems (in french). *Ann. Telecom.*, (Fr.), Vol 35, Nos 7-8, pp 281-296, 1980
2.10. JUMARIE, G.; Theory of relative information V. Transfer of information, transfer of uncertainty, information balance (in french). *Ann. Telecom.*, (Fr.), Vol 37, Nos 5-6, pp 201-214, 1982
2.11. JUMARIE, G.; New concepts of fuzzy-subjective probabilities and their applications to soft-making decision processes, *Proceedings of the Second IFSA (International Fuzzy Systems Association) Congress*, pp 568-571, Tokyo, 1987
2.12. JUMARIE, G.; A Minkowskian theory of observation. Application to uncertainty and fuzziness. *Fuzzy Sets and Systems*, Vol 24, No 2, pp 231-254, 1987
2.13. JUMARIE, G.; Solution of the Fokker-Planck equation by using the maximum entropy principle. *J. Math. Physics*, Vol 31, No 10, pp 2389-2392, 1990
2.14. JUMARIE, G.; *Relative Information. Theories and Applications.* Springer Verlag, Berlin, 1990
2.15. JUMARIE, G.; Solution of the dynamical equation of the conditional probability of Markov processes using the maximum entropy principle. *Syst. Analysis, Modelling, Simulation*, Vol 10, No 3, pp 169-186, 1993
2.16. KULLBACK, S.; *Information Theory and Statistics.* Wiley, New York, 1959
2.17. RENYI, A.; On measures of entropy and information. *Proc. 4th Berkeley Symp. Math. Stat. Probability*, Vol 1, pp 547-561, University of California Press, Berkeley, 1961
2.18. RENYI, A.; *Probability Theory*, North Holland, Amsterdam, 1970
2.19. SHANNON, C.E.; A mathematical theory of communication I. *Bell System Tech. J.*, Vol 27, pp 379-423, 1948.
2.20. SHANNON, C.E.; A mathematical theory of communication II. *Bell System Tech. J.*, Vol 27, pp 623-656, 1948.
2.21. SHORE, J.E. and JOHNSON, R.W.; Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, *IEEE Trans. Inf. Theory*, Vol IT-26, pp 26-37, 1980
2.22. UFFINK, J.; Can the maximum entropy principle be explained by a consistency principle? *Studies in the History and the Philosophy of Modern Physics*, Vol 26, pp 223-261, 1995
2.23. UFFINK, J.; The constraint rule of the maximum entropy principle. *Studies in the History and the Philosophy of Modern Physics*, Vol 27, pp 47-79, 1996

Chapter 3

Path Entropies of Non-Random Functions

Material remains and form vanishes.
Pierre de Ronsard (Elégies)

3.1. Introduction

The present chapter is of paramount importance, and is so for two reasons. On the one hand it displays one of the main contributions of the book, and on the other hand, the argument which is used herein to obtain this result will be useful later when we try to derive a model of the quantum entropy of non-probabilistic matrices.

How can we measure the amount of information involved in a non-random or deterministic function? In order to answer this question we shall generalize Shannon information theory, and we shall show that the entropies introduced in the second chapter have their counterparts in terms of the information of deterministic functions. The point of importance is that the expressions so obtained are not new definitions made for convenience only, but rather are direct consequences of a rationale which is already transparent in Shannon theory itself. In the next section we shall carefully explain the latter, and then, starting from this generalizing procedure, we shall obtain various expressions for entropy of non-random continuous functions. We shall apply these definitions to non-random piecewise continuous functions, and finally we shall define a measure of the informational divergence between npn-random functions.

The reader who is interested mainly in the substance of the chapter can read carefully the next section 3.2, and then directly refer to the definitions of these entropies, omitting proofs and comments.

3.2. Path entropies of non-random functions. An outline.

Let X and X' denote two random variables with the respective (Shannon) entropies $H(X)$ and $H(X')$, and with the joint entropy $H(X,X')$. This joint entropy is some functional of their joint probability density to be made precise later, e.g., Shannon or Renyi entropy. In any case we have the relation

G. Jumarie, *Maximum Entropy, Information Without Probability and Complex Fractals*
© Springer Science+Business Media Dordrecht 2000

$$H(X,X') \leq H(X)+H(X'), \tag{3.1}$$

where the equality holds when, and only when X and X' are mutually independent. We may therefore think of $H(X')$ as

$$H(X') = \max_{p(.)} \left[H(X,X')-H(X)\right], \tag{3.2}$$

with $p(x)$ denoting the probability density of X, subject to the condition

$$H(X) = h, \tag{3.3}$$

where h is a constant. One may think of (3.2) as a means of defining $H(X')$ and it actually does not depend on the constant h in (3.3), since one has the equality

$$\max H(X,X') = H(X)+H(X') . \tag{3.4}$$

Here X and X' play symmetrical roles; as a result $H(X')$ can be defined exactly like $H(X)$, and using (3.2) to define it appears to be unnecessarily complicated.

However, we shall consider instances where X and X' do not play such symmetrical roles. We assume that X is still a random variable of which the entropy $H(X)$ is defined in the usual way. We assume further that we have some plausible definitions of $H(X,X')$ but none for $H(X')$. Then (3.2) will be used to *define* $H(X')$. Unlike the case of X' being a random variable, however, the result will generally depend upon the constant h in (3.3).

We shall firstly apply this method to Shannon entropy of continuous random variable, namely

$$H(X) := -\int_{\Re} p(x)\ln p(x)dx , \tag{3.5}$$

and then to Renyi entropy of order s,

$$H_{R,s}(X) := \frac{1}{1-s}\ln\int_{\Re} p^s(x)dx, \quad s \rhd 0, \quad s \neq 1. \tag{3.6}$$

The relation (3.2) will be used to define the entropy of some deterministic (or non random) continuously differentiable functions $f(.)$. Assuming that $Y = f(X)$ we shall interpret $H(Y)$ as the joint entropy $H(X,f(.))$ of X and the function $f(.)$, and use (3.2) to define the entropy of $f(.)$.

In the following we shall generalize to the n-dimensional case. Therefore, we consider a function $f : \Re^n \to \Re^n$, $x \to f(x)$, but no change in notation will result. The variables now denote n-dimensional vectors while $f'(x)$ is the Jacobian determinant of $f(.)$.

We shall furthermore assume that we consider only some domain Ω of $\Re^n$ to which X belongs. If X is defined outside of Ω, then we shall normalize its probability density according to

$$p(x) = \frac{p_a(x)}{\int_\Omega p_a(x)dx}, \tag{3.7}$$

where $p_a(x)$ is the original probability density of X. The dependence on Ω of the entropy just defined will be emphasized by the notation $H(f(.);\Omega)$.

Notice that we take the entropy of *continuous* random variable as a point of departure, instead of the familiar discrete variable. We shall see later how one can derive the entropy of a discrete function from that of a continuous one.

The maximization involved in the equation (3.2) will be carried out by using the conventional Lagrange multipliers method. The constraints are then the condition (3.3) and the additional condition

$$\int_\Omega p(x)dx = 1. \tag{3.8}$$

3.3. Shannon path entropy of continuous functions

PROPOSITION 3.1. *Let* $f : \Re^n \to \Re^n$, $x \to f(x)$, *denote a continuously differentiable function of which the Jacobian determinant is denoted by* $f'(x)$; *then a measure* $H_c(f(.);\Omega)$ *of the amount of uncertainty it involves on the domain* Ω, *and which furthermore is fully consistent with the Shannon entropy of a random variable, is the S-path entropy (S holds for Shannon) of order* $c \in \Re$, *defined by the expression* [3.7,3.9]

$$H_c(f(.);\Omega) := \frac{\int_\Omega |f'(x)|^c \ln|f'(x)|dx}{\int_\Omega |f'(x)|^c dx} . \blacksquare \tag{3.9}$$

Proof. (i) Let $X \in \Omega \subset \Re^n$ and $X' \in \Re$ be two random vectors, and let $p(x)$ denote the probability density of X. The Shannon entropy of the pair (X,X') satisfies the equation

$$H(X,X') = H(X) + H(X'|X) , \tag{3.10}$$

where $H(X'|X)$ is the conditional entropy of X' given X, that is

$$H(X'|X) := \int_\Omega p(x) H(X'|X = x)dx . \tag{3.11}$$

(ii) Assume that the random n-vectors X and Y are dependent via the equation $Y = f(X)$, where $f(.)$ is a continuously differentiable function; then according to a basic result of information theory one has the equality

$$H(Y) = H(X) + \int_\Omega p(x) \ln|f'(x)|dx . \tag{3.12}$$

(iii) Analogously with equation (3.10) we are thus led to write the entropic equation

$$\begin{aligned} H(Y) &=: H(X, f(.);\Omega) \\ &= H(X) + H(f(.);\Omega|X) , \end{aligned} \tag{3.13}$$

and to define the conditional entropy of $f(.)$ on Ω given X, by the expression

$$H(f(.);\Omega|X) := \int_\Omega p(x) \ln|f'(x)|dx . \tag{3.14}$$

(iv) This being so, it is well known that the equation ((3.10) yields

$$H(X,X') = H(X) + H(X'), \tag{3.15}$$

when and only when X and X' are mutually independent, therefore one has the inequality

$$H(X'|X) \leq H(X').$$

As a result, we can write the equality

$$H(X') = \max_{p(x)} H(X'|X), \tag{3.16}$$

subject to the condition

$$H(X) = constant \tag{3.17}$$

(v) Analogously with this remark, we shall define the entropy of the function $f(.)$ on Ω by the relation

$$H(f(.);\Omega) := \max_{p(.)} H(f(.);\Omega|X),$$

given the constraint (3.17). We should express that the quantity

$$\int_{\Omega} p(x)\left[\ln|f'(x)| + \lambda + \mu \ln p(x)\right]dx ,$$

where the Lagrange multipliers λ and μ are associated with the constraint (3.8) and (3.17) respectively, is extremal. We then have the necessary optimality condition

$$\ln|f'(x)| + \lambda + \mu[\ln p(x) + 1] = 0,$$

hence

$$p(x) = |f'(x)|^{-1/\mu} \exp[-(1 + (\lambda/\mu))].$$

The constraint (3.8) obviously yields

$$p(x) = \frac{|f'(x)|^{-1/\mu}}{\int_{\Omega} |f'(x)|^{-1/\mu} dx}, \tag{3.18}$$

In order to determine the constant μ, we shall refer to the constraint (3.17), clearly

$$\int_{\Omega} p(x)\ln p(x)dx \;=\; -h,$$

or

$$\ln K(\mu)+\frac{\int_{\Omega}|f'(x)|^{-1/\mu}\ln|f'(x)|dx}{\mu K(\mu)} \;=\; h \tag{3.19}$$

where $K(\mu)$ is equal to

$$K(\mu) \;:=\; \int_{\Omega}|f'(x)|^{-1/\mu}dx. \tag{3.20}$$

The constant μ then appears as an implicit function of h. Letting $-1/\mu = c$ and substituting the expression (3.18) for $p(x)$ into (3.14) provides the result.

DEFINITION 3.1. We shall refer to $H(f(.);\Omega) := H_1(f(.);\Omega)$, $c = 1$, as to the Shannon path entropy or S-path entropy of the function $f(.)$ on the domain Ω.■

We use the term of path entropy instead of entropy to emphasize the property that, to some extent, the entropy so obtained can be thought of as a measure of the uncertainty involved in the graph $(x,f(x))$ generated by the function on the domain considered. In other words, we have here a first approach to entropy of forms.

Remark that when $f(x)$ is the (cumulative) distribution function $F(x)$ of the one dimensional random variable $(X,p(x))$ (clearly $F(x)=\int_{-\infty}^{x} p(\xi)d\xi$), then we have the equality

$$H_1(F(.);\Re) \;=\; -H(X), \tag{3.21}$$

which shows that entropy and path entropy have opposite signs. We shall come back on this property later (see subsection 3.8.3).

3.4. Renyi path entropy of continuous functions

PROPOSITION 3.2. *Consider the non-random* $\Re^n \to \Re^n$ *function of proposition 3.1, equation (3.9); then a measure* $H_{R,s}(f(.);\Omega,u)$ *of the amount of uncertainty it involves on the domain* Ω*, and which furthermore is fully consistent with Renyi entropy of the random variables, is the generalized R-path entropy (R holds for Renyi) of order* $s \geq 0$*, defined by the expression*

$$H_{R,s}(f(.);\Omega,u)$$

$$:= \frac{1}{1-s} \ln \frac{\int_\Omega \left(u+|f'(x)|^{1-s}\right)^{\frac{s}{1-s}} |f'(x)|^{1-s} dx}{\int_\Omega \left(u+|f'(x)|^{1-s}\right)^{\frac{s}{1-s}} dx} \tag{3.22}$$

where u denotes a non negative real valued parameter: u > 0 ■

Proof. (i) First of all, we bear in mind that the Renyi entropy of $Y = f(X)$ is given by the expression

$$H_{R,s}(Y) = \frac{1}{1-s} \ln \int_\Omega p^s(x)|f'(x)|dx \ . \tag{3.23}$$

(ii) This being so, we apply the method discussed in Section 3.2 to the Renyi entropy as defined by the equation (3.6). The quantity to be maximized now reads

$$H_{R,s}(f(.);\Omega) := \max_{p(.)} \frac{1}{1-s} \ln \frac{\int_\Omega p^s(x)|f'(x)|^{1-s} dx}{\int_\Omega p^s(x)dx}, \tag{3.24}$$

subject to the constraint (3.8) together with

$$\int_\Omega p^s(x)dx = constant$$

We then maximize the Lagrangian

$$\int_\Omega \left[p^s(x)|f'(x)|^{1-s} + u\,p^s(x) + v\,p(x)\right]dx \ ,$$

where u and v are two constants, hence

$$p^{s-1}(x) = \frac{-v}{s\left(u+|f'(x)|^{1-s}\right)} .$$

These constants are such that the right side term above is positive, therefore

$$p(x) = \left[-\frac{s\left(u+|f'(x)|^{1-s}\right)}{v}\right]^{\frac{1}{1-s}} . \tag{3.25}$$

If we restrict ourselves to $v < 0$ and assume that $u \geq 0$, then substituting (3.25) for $p(x)$ into (3.24) yields the result.

Notice that $H_{R,s}(f(.);\Omega,v)$ can be rewritten in the form

$$H_{R,s}f(f(.);\Omega,u)$$

$$= -\frac{1}{1-s}\ln \frac{\int_{\Omega}\left(1+u|f'(x)|^{s-1}\right)^{\frac{s}{1-s}}|f'(x)|^{s}dx}{\int_{\Omega}\left(1+u|f'(x)|^{s-1}\right)^{\frac{s}{1-s}}|f'(x)|dx} \tag{3.26}$$

Useful Special Case. Assume that $u = 0$, then the equation (3.22) yields the expression

$$H_{R,s}(f(.);\Omega) = -\frac{1}{1-s}\ln \frac{\int_{\Omega}|f'(x)|^{s}dx}{\int_{\Omega}|f'(x)|dx} , \tag{3.27}$$

which will be referred to as the *Renyi path entropy of order s of f(.) on Ω.*

3.5. Structural path entropy of continuous functions

PROPOSITION 3.3. *Consider the non-random* $\Re^n \to \Re^n$ *function of Proposition 3.1, equation 3.9. A measure* $H_{C,s}(f(.);\Omega,u)$ *of the amount of uncertainty it involves on the domain* Ω, *which furthermore is fully consistent with the structural entropy of random variables, is the generalized structural path entropy (or C-path entropy) of order* $s \geq 0$ *defined as*

$$H_{C,s}(f(.);\Omega,u)$$

$$:= \frac{1}{e^{1-s}-1} \frac{\int_{\Omega}\left(u+|f'(x)|^{1-s}\right)^{\frac{s}{1-s}}\left(|f'(x)|^{1-s}-1\right)dx}{\int_{\Omega}\left(u+|f'(x)|^{1-s}\right)^{\frac{s}{1-s}}dx} \quad , \quad u \geq 0 \cdot \blacksquare \tag{3.28}$$

Proof. (i) We refer to the equation (2.19) which defines conditional structural entropy, and we notice that when $X \in \Re^n$ and $Y \in \Re^n$ are two mutually independent continuous random variables we have the equality

$$H_{C,s}(X,Y) = H_{Cs}(X) + H_{C,s}(Y)\int_{\Re^n} p^s(x)dx, \tag{3.29}$$

where $p(x)$ is the probability density of X.

When X and Y are not independent, we may have

$$H_{C,s}(X,Y) \leq H_{Cs}(X) + H_{C,s}(Y)\int_{\Re^n} p^s(x)dx, \tag{3.30}$$

,

or on the contrary

$$H_{C,s}(X,Y) \geq H_{Cs}(X) + H_{C,s}(Y)\int_{\Re^n} p^s(x)dx,, \tag{3.31}$$

and as a result, one may think of $H_{C,s}(Y)$ as

$$H_{C,s}(Y) := optimum \frac{H_{C,s}(X,Y) - H_{C,s}(X)}{\int_{\Re^n} p^s(x)dx}, \tag{3.32}$$

subject to the condition

$$H_{C,s}(X) = constant. \tag{3.33}$$

(ii) Next, assume that X and Y are related by the equation $Y = f(X)$, where $f(X)$ is a continuously differentiable function; then a simple transformation of variables yields

$$H_{C,s}(Y) = \frac{\int_\Omega p^s(x)|f'(x)|^{1-s}dx - 1}{e^{1-s}-1}. \tag{3.34}$$

(iii) Analogously with the equation (2.19), we shall write

$$\begin{aligned} H_{C,s}(Y) &=: H_{C,s}(X, f(.);\Omega), \\ &= H_{C,s}(X) + \int_\Omega p^s(x) H_{C,s}\big(f(.);\Omega|x\big)dx, \end{aligned} \tag{3.35}$$

and we are so led to define $H_{C,s}(f(.);\Omega)$ by the expression

$$H_{C,s}(f(.);\Omega) := \textit{optimum} \frac{\int_\Omega p^s(x) H_{C,s}\big(f(.);\Omega|x\big)dx}{\int_\Omega p^s(x)dx}, \tag{3.36}$$

subject to the condition (3.33).

(iv) The quantity to be optimized now reads

$$H_{C,s}(f(.);\Omega) = \underset{p(.)}{\textit{optimum}} \frac{H_{C,s}(Y) - H_{C,s}(X)}{\int_\Omega p^s(x)dx}, \tag{3.37}$$

$$= \underset{p(.)}{\textit{optimum}} \frac{\int_\Omega \big(|f'(x)|^{1-s} - 1\big)p^s(x)dx}{\int_\Omega p^s(x)dx}, \tag{3.38}$$

subject to the constraint (3.8) together with (3.33). We then optimize the Lagrangian

$$\int_\Omega \Big[\big(|f'(x)|^{1-s} - 1\big)p^s(x) + \mu\, p^s(x) + \lambda\, p(x)\Big]dx \quad ,$$

where λ and μ denote the multipliers, hence

$$p^{s-1}(x) = \frac{-\lambda}{s(\mu - 1 + |f'(x)|^{1-s})}.$$

These constants should be such that $p^{s-1}(x)$ is positive, and thus we have

$$p(x) = \left[-\frac{s}{\lambda}\big(\mu - 1 + |f'(x)|^{1-s}\big)\right]^{\frac{1}{1-s}}. \tag{3.39}$$

If we restrict ourselves to negative values for λ, then substituting (3.39) into (3.38) yields the result, with $u := \mu - 1$.

3.6. On the practical meaning of these entropies

(i) One can show that

$$\lim H_c(f(.);\Omega) = \ln|f'(x)|_M \quad as \quad c \uparrow +\infty \tag{3.40}$$

and

$$\lim H_c(f(.);\Omega) = \ln|f'(x)|_m \quad as \quad c \downarrow -\infty\ , \tag{3.41}$$

where $|f'(x)|_M$ and $|f'(x)|_m$ are respectively the maximum value and the minimum value of $|f'(x)|$ on Ω.

(ii) It is by now recognized that when a human being is observing a (human) face, for instance, his eyes examine successively each point of this pattern $f(x)$ by using a scanning distribution frequency $p(x)$, and, furthermore, it appears that the cortex scans the salient features of the picture, such that angles, for instance, with higher density (see for instance [3.12]), and as a result, according to the equation (3.9), we can assume that this scanning distribution is

$$p_{f,c}(x) := \frac{|f'(x)|^c}{\int_\Omega |f'(x)|^c dx}\ . \tag{3.42}$$

The value $c = 0$ defines a uniform scanning distribution, while large values of c characterize scanning processes concentrated on the sharp features of the picture.

For a given c, $p_{f,c}(x)$ thus appears as the scanning distribution which provides the maximum amount of information about $f(.)$.

(iii) The Shannon entropy of an incomplete probability distribution on Ω, that is to say such that $\int_\Omega p(x)dx \lhd 1$ can be mathematically obtained in the form (see for instance [3.1])

$$H(X) := \frac{-\int_\Omega p(x)\ln p(x)dx}{\int_\Omega p(x)dx}\ , \tag{3.43}$$

and it appears that the expression of $H(f(.);\Omega)$ is quite consistent with this result.

(iv) Notice that $H_C(f(.);\Omega)$ depends upon $|f'(x)|$ only (it does not refer to $f(x)$ itself), and this result merely pictures the well known claim according to which information originates from variation of form (see for instance [3.6] and [3.14]). Well obviously the same remark applies to the Renyi entropy $H_{R,s}(f(.);\Omega)$.

(v) The value of c is defined by the value of the constant in the constraint (3.17), and $c = 0$ corresponds to the maximum entropy of X, that is to say, to the uniform distribution if we assume that Ω is bounded.

(vi) We bear in mind that Renyi entropy can be of valuable help in some practical instances. Firstly, sometimes, depending upon the problem, it may be more practical than the Shannon entropy, in such a manner that it could be interesting to use it, and then to let s tends to the unity in the result so obtained. Secondly, it may be selected as a measure of information in the presence of faulty observation. Indeed, consider the measure $Y = X + W$, where X and W are two independent Gaussian variables $N(\mu,\sigma_X^2)$ and $N(0,\sigma_W^2)$, X is the observed variable, and W is a disturbing noise. It is well known that

$$H(Y) = H(X)+\frac{1}{2}\ln\left(1+\frac{\sigma_W^2}{\sigma_X^2}\right). \tag{3.44}$$

On the other hand, when s is close to the unity one has the approximation

$$H_{R,s}(X) = H(X)-\frac{1}{2}(s-1)+o((s-1)^2), \tag{3.45}$$

where $o(.)$ denotes Landau's symbol. As a result, if we make the identification $H_{R,s}(X) \equiv H(Y)$, we obtain the identification

$$s \cong 1-\ln\left(1+\frac{\sigma_W^2}{\sigma_X^2}\right) . \tag{3.46}$$

(vii) After a careful examination the reader will be convinced that the above entropies are different from the Kolmogorov entropy of maps [3.11] which is a measure of information instead of uncertainty, from the Liapunov exponent (see for instance [3.3] or [3.13]) and from the entropy of curves as defined recently by Dupain, Kamae and Mendes France [3.2] (see chapter 9 for further details).

(viii) Notice that, for bounded Ω, the Renyi entropy of order zero is defined by the expression

$$H_{R,0}(f(.);\Omega,u) = \ln \frac{\int_\Omega |f'(x)|dx}{\int_\Omega dx}, \tag{3.47}$$

which explicitly involves the total variation $\int_\Omega |f'(x)|dx$ of $f(.)$ on Ω.

3.7. Path entropy and encoding problem

In the case when $f(x)$ is the (cumulative) distribution function $F(x)$ of the random variable X, we have the equality (3.21) which we re-write below for the convenience of the reader,

$$H_1(F(.);\Re) = -H(X).$$

This relation might look like surprising on the surface, but as a matter of fact, it is already transparent in one of the encoding techniques suggested by Shannon himself.

The basic idea of this method is as follows. We write the letters x_i, $i = 1,2,...,N$, of the alphabet under consideration in the order of decreasing probabilities, $p_1 \geq p_2 \geq ... \geq p_N$(we bear in mind that $pr(X=x_i)=p_i$), and we consider the sums $P_1, P_2, ..., P_N$ defined by the expressions

$$P_1 = 0, \tag{3.48}$$

$$P_i = p_1 + p_2 + ... + p_N \quad , \quad i = 1,2,...,N. \tag{3.49}$$

One then has the multiple inequalities $P_1 \lhd P_2 \lhd ... \lhd P_N$, and one can consider the N numbers $\{P_1, P_2, ..., P_N\}$ as being an original alphabet which is in a one-to-one correspondence with the sequence of the initial alphabet $\{x_i\}$. We then have to encode the P-alphabet, that is to say , we have to associate a sequence of bits with each number P_i, in such a manner that no mistake be possible in the corresponding identification procedure. In other words, encoding the sequence $\{x_i\}$ amounts to encode the sequence $\{P_i\}$ and the amount of information so involved should be the same, at least in absolute value.

The fact that $H_1(F(.);\Re)$ is negative should not be disturbing at all and this is merely due to the fact that one has the inequality $P_i \lhd 1$, for every i, in such a manner that P can be writen in the form

$$P = a_1 2^{-1} + a_2 2^{-2} + ... + a_k 2^{-k} + ... \tag{3.50}$$

3.8. On the relations between these entropies

3.8.1. Shannon path entropy and Renyi path entropy

It is well known that for a random variable X $H_{R,s}(X)$ converges to $H(X)$ when s tends to the unity, and, at first glance, one should expect to have a similar property with the entropy of non random functions. In this way, we can state:

PROPOSITION 3.4. *The following property holds, that is*

$$\lim H_{R,s}(f(.);\Omega,u) = H_{1/(1+u)}(f(.);\Omega) \quad as \quad s \to 1 \,.\blacksquare \tag{3.51}$$

Proof. We refer to the expression (3.26) of the Renyi entropy of non-random functions which we re-write in the form

$$H_{R,s}(f(.);\Omega,u) = -\frac{1}{1-s} \ln \frac{\int_\Omega N(f')dx}{\int_\Omega D(f')dx}. \tag{3.52}$$

(i) When s is close to the unity, we have the following Taylor approximations,

$$|f'|^{s-1} = 1 + (s-1)\ln|f'| + o((s-1)^2), \tag{3.53}$$
$$|f'|^{s} = |f'|\left(1 + (s-1)\ln|f'|\right) + o((s-1)^2). \tag{3.54}$$

(ii) The equality (3.53) provides

$$1 + u|f'|^{s-1} \cong (1+u)\left(1 + \frac{(s-1)u}{1+u}\ln|f'|\right),$$

hence

$$\left(1 + u|f'|^{s-1}\right)^{\frac{s}{s-1}} \cong (1+u)^{\frac{s}{1-s}} \exp\left\{\frac{s}{1-s}\ln\left(1 + \frac{(s-1)u}{1+u}\ln|f'|\right)\right\}$$

$$\cong (1+u)^{\frac{s}{1-s}} \exp\left\{-\frac{u}{1+u}\ln|f'|\right\}$$

$$\cong (1+u)^{\frac{s}{1-s}}|f'|^{-\frac{u}{1+u}}. \tag{3.55}$$

(iii) Combining (3.54) and (3.55) yields

$$N(f') \cong (1+u)^{\frac{s}{1-s}}|f'|^{\frac{1}{1+u}}(1+(s-1)\ln|f'|),$$

while (3.55) provides

$$D(f') \cong (1+u)^{\frac{s}{1-s}}|f'|^{\frac{1}{1+u}}.$$

(iv) We finally obtain the approximation

$$H_{R,s}(f(.);\Omega,u)$$

$$\cong -\frac{1}{1-s}\ln\left[1+(s-1)\frac{\int_\Omega |f'|^{\frac{1}{1+u}}\ln|f'|dx}{\int_\Omega |f'|^{\frac{1}{1+u}}dx}\right], \tag{3.56}$$

whence the result.

Comments. (i) In the special case of $u = 0$ the relation (3.51) provides

$$\lim H_{R,s}(f(.);\Omega) = H(f(.);\Omega) \quad as \quad s \to 1 \tag{3.57}$$

which is exactly the counterpart of (2.14).

(ii) The relation (3.51) suggests the identification $c = 1/(1+u)$, or

$$u := (1-c)/c \tag{3.58}$$

therefore

$$H_{R,s}(f(.);\Omega,u) \equiv H_{R,s}\left(f(.);\Omega,\frac{1-c}{c}\right). \tag{3.59}$$

(iii) According to (3.59) the Renyi entropy which is associated with a uniform scanning distribution (i.e. $c = 0$) is defined by $u = \infty$, that is to say, in the special case when Ω is bounded one has

$$H_{R,s}(f(.);\Omega,\infty) = \frac{1}{1-s}\ln\frac{\int_{\Omega}|f'(x)|^{1-s}dx}{\int_{\Omega}dx}. \tag{3.60}$$

(iv) As an application of this expression (3.60) assume that $f(x)$ is the (cumulative) distribution function $F(x)$ (i.e., $F'(x) = p(x)$) defined on the finite interval $x \in [a,b] \subset \Re$, one has the entropy

$$H_{R,s}(F(.);[a,b],\infty) = \frac{1}{1-s}\ln\left[\frac{1}{b-a}\int_a^b|p(x)|^{1-s}dx\right], \tag{3.61}$$

and the opposite of this expression can be considered as a new measure of the amount of uncertainty defined by the associated random variable X.

3.8.2. *Some properties of structural entropy of non-random functions*

(i) The relation between structural entropy and Shannon entropy of non random functions is illustrated by the following result:

PROPOSITION 3.5. *One has the following equality, that is*:

$$\lim H_{C,s}(f(.);\Omega,u) = H_{1/(1+u)}(f(.);\Omega) \quad as \quad s \to 1 \ .\blacksquare \tag{3.62}$$

Proof. The proof is similar to the proof of Proposition 3.4, and starts with the following Taylor's approximation,

$$\left(1+u^{-1}|f'|^{1-s}\right)^{\frac{s}{1-s}} = \left(\frac{1+u}{u}\right)^{\frac{s}{1-s}}|f'|^{\frac{1}{1+u}} + o((s-1)^2),$$

$$|f'|^{1-s} - 1 = 1 + (1-s)\ln|f'| + o((s-1)^2),$$

$$e^{1-s} - 1 = 1 - s + o((s-1)^2).$$

(ii) *A special useful case.* As a consequence of the equation (3.62) we shall ascribe a special interest to the value $u = 0$ in the expression of $H_{C,s}(f(.);\Omega, u)$ and we shall refer to the structural entropy of order s in the form

$$H_{C,s}(f(.);\Omega,u) = \frac{1}{e^{1-s}-1}\frac{\int_\Omega |f'(x)|^s\left(|f'(x)|^{1-s}-1\right)dx}{\int_\Omega |f'(x)|^s dx}, \tag{3.63}$$

$$= \frac{1}{e^{1-s}-1}\left(\frac{\int_\Omega |f'(x)|dx}{\int_\Omega |f'(x)|^s dx}-1\right). \tag{3.64}$$

(iii) Relation with Renyi entropy of non-random functions.

The relation between Renyi entropy and structural entropy of random variables is quite clear by their definitions and one can write directly

$$H_{R,s}(f(.);\Omega,u) = \frac{1}{1-s}\ln\left[\left(e^{1-s}-1\right)H_{C,s}+1\right]. \tag{3.65}$$

The equations (3.27) and (3.64) provide a similar relation for path entropies of non random functions, and we have

$$H_{R,s}(f(.);\Omega) = \frac{1}{1-s}\ln\left[\left(e^{1-s}-1\right)H_{C,s}(f(.);\Omega)+1\right]. \tag{3.66}$$

3.8.3. Entropies of non-random functions and entropies of random variables

Assume that $f(x)$ is the cumulative distribution $F(x)$ of the probability density $p(x)$; then one has the equality

$$H(F(.);\Re) = -H(X). \tag{3.67}$$

In other words, $H(X)$ and $H(F(.);\Re)$ vary in opposite ways. One of them is maximum when the other achieves its minimum value; and it is quite right, so there is no inconsistency at all. For instance if $p(x)$ is the uniform distribution then $H(X)$ is a maximum, but the uncertainty involved in $F(x)$, which is then a straightline, is minimum!

The same remark holds for Renyi entropy via the equation

$$H_{R,s}(f(.);\Re) = -H_{R,s}(X) \quad . \tag{3.68}$$

For the structural entropy one has the relation

$$H_{C,s}(F(.);\Re) = \frac{1}{e^{1-s}-1}\left[\frac{1}{\int_{\Re} p^{s}(x)dx}-1\right] \tag{3.69}$$

which shows that here again these two entropies vary in opposite ways.

3.9. Informational divergence between non-random functions

3.9.1. Cross-entropic variance of probability densities

Assume that we want to compare two probability densities $p(x)$ and $q(x)$ (from the informational standpoint) on the domain $\Omega \subset \Re^n$ in such a manner that we have

$$\int_{\Omega} p(x)dx \lhd 1 \qquad \textit{and} \qquad \int_{\Omega} q(x) \lhd 1.$$

We cannot utilize the integral (2,41) which defines Kullback cross-entropy $H(q,p)$ since then the corresponding expression may have negative values, and thus we need an alternative. A possible approach is the following:

DEFINITION 3.2. A measure of the informational divergence between the probability densities $p(x)$ and $q(x)$ on the domain Ω, which, furthermore, can be thought of as an extension of Kullback cross-entropy, is provided by the *cross-entropic variance* or the *relative entropic variance*

$$\sigma_H^2(q,p;\Omega) := \frac{\int_{\Omega} q(x)\ln^2[q(x)/p(x)]dx}{\int_{\Omega} q(x)dx} - \left[\frac{\int_{\Omega} q(x)\ln[q(x)/p(x)]dx}{\int_{\Omega} q(x)dx}\right]^2 .\blacksquare \tag{3.70}$$

According to Schwarz inequality one has $\sigma_H^2(q,p;\Omega) \geq 0$, and the equality holds when and only when there exists a positive constant k such that

$$q(x) = kp(x). \tag{3.71}$$

The value of k is defined by the equation

$$\int_{\Omega} q(x)dx \;=\; k\int_{\Omega} p(x)dx \;. \tag{3.72}$$

In the special case when $\Omega \equiv \Re^n$, k has necessarily the unit value in such a manner that $\sigma_H^2(q,p;\Omega)$ and $H(q,p)$ are more or less equivalent.

Remarks and commemts. (i) According to the equation (3.71), the cross-entropic variance can be thought of as a measure of similarity on the domain Ω..

(ii) The cross-entropic variance is related to the so-called Fisher information as follows.

Consider the probability density $p(x,\theta)$, where θ is an external parameter, $x \in \Re^n, \theta \in \Theta \subset \Re$, and define the set $E_\theta := \{x \in \Re^n; p(x,\theta) \triangleright 0\}$. The amount of Fischer information involved in $p(x,\theta)$ about the value of θ is the quantity

$$I_F(p,\theta) \;:=\; \int_{E_\theta}\left[\frac{\partial}{\partial\theta}\ln p(x,\theta)\right]^2 p(x,\theta)dx. \tag{3.73}$$

We then have the relation

$$\sigma_H^2\big(p(.,\theta), p(.,\theta+\Delta\theta;E_\theta)\big) \;=\; I_F(p,\theta)(\Delta\theta)^2. \tag{3.74}$$

Indeed on making the substitution $q(x) \leftarrow p(x,\theta)$, $p(x) \leftarrow p(x,\theta+\Delta\theta)$ in the equation (3.67) yields

$$\sigma_H^2\big(p(.,\theta), p(.,\theta+\Delta\theta);E_\theta\big) \;=\; \int_{E_\theta} p(x,\theta)\left[\frac{\partial}{\partial\theta}\ln p(x,\theta)\right]^2 dx$$

$$-(\Delta\theta)^2\left[\int_{\Omega} p'_x(x,\theta)dx\right]^2, \tag{3.75}$$

and according to the equality

$$\int_{E_\theta} p(x,\theta)dx \;=\; 1$$

one has necessarily

$$\int_{E_\theta} p'_\theta(x,\theta)dx \;=\; 0.$$

3.9.2. *Relation between Kullback cross-entropy and entropy of non-random functions*

PROPOSITION 3.6. *The relative entropies $H(q,p)$ and $H_{R,s}(q,p)$ of random variables can be thought of as direct consequences of Shannon path entropy and Renyi path entropy of non-random functions.* ■

Proof. (i) Let $F(x)$ and $G(x)$ denote two $\Re \to \Re$ functions which are strictly increasing and continuously differentiable. The function $G(F)$ is then meaningfully defined, it is continuously differentiable with respect to F, and one has $dF > 0$ and $dG > 0$; and according to the equations (3.9) and (3.27) which define, respectively, S-entropy and R-entropy of non random functions, one can meaningfully consider the entropies

$$H(G(F(.));\Omega) \;:=\; \frac{\int_\Omega \frac{dG}{dF}\ln\left(\frac{dG}{dF}\right)dF}{\int_\Omega \left(\frac{dG}{dF}\right)dF}, \tag{3.76}$$

and

$$H_{R,s}(G(F(.));\Omega) \;:=\; -\frac{1}{1-s}\ln\frac{\int_\Omega \left(\frac{dG}{dF}\right)^s dF}{\int_\Omega \left(\frac{dG}{dF}\right)dF}. \tag{3.77}$$

(ii) Assume that $F(x)$ and $G(x)$ are the (cumulative) distributions associated with the densities $p(x)$ and $q(x)$ respectively, clearly $F'(x) = p(x)$ and $G'(x) = q(x)$, then on substituting these values into (3.76) and (3.77) we obtain the Kullback cross-entropy

$$H(G(F(.));\Re) \;=\; H(q,p),$$

and the Renyi cross-entropy

$$H_{R,s}(G(F(.));\Re) \;=\; H_{R,s}(q,p).$$

The interest of this result is two-fold. Firstly, it enlightens the consistency between entropy of non-random functions and Shannon information theory (which is not surprising at all since we started from the latter to derive the former), in other words it supports the arguments which allowed us to define the entropies of non-random functions; and secondly, it will suggest new meanings for the entropies of non-random functions, which will be of help later.

3.9.3. Structural cross-entropy of probability densities

We can now use the results of the preceding sections to derive the expression of the cross-entropy defined in terms of structural entropy. Indeed, we have,

PROPOSITION 3.7. *A definition of cross-entropy of random variable, which is fully consistent with structural entropy, is provided by the expression*

$$H_{C,s}(q,p) = \frac{1}{e^{1-s}-1}\left[\frac{\int_{\Re} q(x)dx}{\int_{\Re} q(x)\left(\frac{q(x)}{p(x)}\right)^{s-1}dx}-1\right]. \blacksquare \tag{3.78}$$

Proof. (i) We consider the cumulative distributions $F(x)$ and $G(x)$ and we refer to the expression (2.40) of the structural entropy to write the structuraal entropy of $G(F)$ as

$$H_{C,s}(G(F(.));\Re) = \frac{1}{e^{1-s}-1}\left[\frac{\int_{\Re}\left(\frac{dG}{dF}\right)dF}{\int_{\Re}\left(\frac{dG}{dF}\right)^{s}dF}-1\right]. \tag{3.79}$$

(ii) We now write that $dF = pdx$ and $dG = qdx$ to obtain the result.
It is easy to verify that $H_{C,s}(q,p)$ is a divergence. In other words, one has the inequality

$$H_{C,s}(q,p) \geq 0 \ ,$$

where the equality holds when and only when $p = q$.

Indeed, it is sufficient to show that the following conditions are satisfied,

$$\int_{\Re} q(x)\left(\frac{q(x)}{p(x)}\right)^{s-1} dx \quad \begin{cases} \lhd\ 1, & s \lhd 1 \qquad (3.80) \\ , & \\ \rhd\ 1, & s \rhd 1 \quad . \qquad (3.81) \end{cases}$$

To this end we assume that $q(.)$ is given, and we consider the optimization

$$\underset{p(.)}{optimum} \int_{\Re} q(x)\left(\frac{q(x)}{p(x)}\right)^{s-1} dx \ , \qquad (3.82)$$

subject to the condition

$$\int_{\Re} p(x)dx \ = \ 1, \qquad (3.83)$$

and we examine whether the optimum so obtained is a maximum or on the contrary a minimum.

The Lagrangian of the problem is

$$L \ := \ \int_{\Re}\left[q(x)\left(\frac{q(x)}{p(x)}\right)^{s-1} + \lambda\ p(x)\right]dx, \qquad (3.84)$$

where λ denotes the Lagrange parameter; and a simple calculation shows that it is optimized when

$$p(x) \ = \ \left(\frac{s-1}{\lambda}\right)^{s} q(x). \qquad (3.85)$$

It follows that one has necessarily that

$$\lambda \ = \ s-1, \qquad (3.86)$$

therefore the equality $p(x) = q(x)$.

In addition, the second derivative of the Lagrangian is

$$\frac{\partial^2 L}{\partial p^2} = s(s-1)\int_{\Re}\left(\frac{q(x)}{p(x)}\right)^s \frac{dx}{p(x)}, \tag{3.87}$$

and as a result, L is maximized when $s < 1$, and is minimized when $s > 1$. In other words the conditions (3.80) and (3.81) are satisfied.

3.10. Path entropies of non-random multivariate functions

3.10.1. Bi-variate function f(x,y)

PROPOSITION 3.8. *A measure of the amount of uncertainty involved in the twice continuously differentiable* $\Re^2 \to \Re$ *function f(x,y) on the domain* Ω*, which moreover, is fully consistent with the Shannon entropy of random variables, is its S-path entropy of order* $c \in \Re$ *defined by the expression*

$$H_c(f(.,.);\Omega) := \frac{\int_\Omega \left|f''_{xy}(x,y)\right|^c \ln\left|f''_{xy}(x,y)\right| dxdy}{\int_\Omega \left|f''_{xy}(x,y)\right|^c dxdy}, \tag{3.88}$$

with $f''_{xy} := \partial^2 f / \partial x \partial y$. *In a like manner, its Renyi path entropy of order s on* Ω *is*

$$H_{R,s}(f(.,.);\Omega) := -\frac{1}{1-s}\ln \frac{\int_\Omega \left|f''_{xy}(x,y)\right|^s dxdy}{\int_\Omega \left|f''_{xy}(x,y)\right| dxdy}, \tag{3.89}$$

and its structural path entropy of order s on Ω *is*

$$H_{C,s}(f(.,.);\Omega) := \frac{1}{e^{1-s}-1}\left[\frac{\int_\Omega \left|f''_{xy}(x,y)\right| dxdy}{\int_\Omega \left|f''_{xy}(x,y)\right|^s dxdy} - 1\right]. \blacksquare \tag{3.90}$$

Proof. First approach. We start from the following remark. The Renyi entropy $H_{R,c}(X)$ of the random variable $X \in \Omega \subset \Re$ satisfies the relation

$$(1-c)\frac{d}{dc}H_{R,c}(X) - H_{R,c}(X) = H_c(F(.);\Omega), \tag{3.91}$$

where $H_C(F(.);\Omega)$ is the S-entropy of the (cumulative) distribution function $F(x)$.

In a like manner, for a random pair $(X,Y) \in \Omega \subset \Re^2$ with the probability density $r(x,y)$, one has the relation

$$(1-c)\frac{d}{dc}H_{R,c}(X,Y) - H_{R,c}(X,Y) = \frac{\int_\Omega r^c(x,y)\ln r(x,y)dxdy}{\int_\Omega r^c(x,y)dxdy}$$

$$= H_c(F(.,.);\Omega), \tag{3.92}$$

in which $F(x,y)$ is the (cumulative) distribution of (X,Y) defined by the equation

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = r(x,y).$$

As a result, we are led to consider the right side term of this equation as measuring the amount of uncertainty involved in $F(.,.)$.

It follows that the density of uncertainty which characterizes the function $f(x,y)$ is $\ln\left|f''_{xy}(x,y)\right|$, therefore the expression (3.88).

Second approach. Another inductive approach to obtain the expression (3.88) is the following one.

(i) Assume that $\Omega = \Omega_1 \times \Omega_2$ and $f(x,y) = g(x)h(y)$, $x \in \Omega_1$, $y \in \Omega_2$; then a suitable definition of $H_c(f(.,.);\Omega_1 \times \Omega_2)$ should satisfy the condition

$$H_c(g(.)h(.);\Omega_1 \times \Omega_2) = H_c(g(.);\Omega_1) + H_c(h(.);\Omega_2). \tag{3.93}$$

(ii) This equation can be re-written in the form

$$H_c(g(.)h(.);\Omega_1 \times \Omega_2) = \frac{\int_\Omega |g'(x)h'(y)|^c \ln|g'(x)h'(y)|dxdy}{\int_\Omega |g'(x)h'(y)|^c dxdy}. \tag{3.94}$$

(iii) We remark that

$$f''_{xy}(x,y) = g'(x)h'(y),$$

and we are so led to generalize (3.94) in the form (3.88).

(iv) Similar arguments apply to the Renyi entropy.

(v) The expression (3.90) is obtained by using the equation which relates Renyi entropy and structural entropy, clearly

$$H_{C,s}(f(.);\Omega) = \frac{\exp\left((1-s)H_{R,s}(f(.);\Omega)\right)-1}{e^{1-s}-1} . \tag{3.95}$$

The extension to n-variate functions is quite direct. For instance, when n=3, the S-entropy reads

$$H_c f(.,.,.);\Omega) = \frac{\int_\Omega \left|f'''_{xyz}(x,y,z)\right|^c \ln\left|f'''_{xyz}(x,y,z)\right| dxdydz}{\int_\Omega \left|f'''_{xyz}(x,y,z)\right|^c dxdydz}. \tag{3.96}$$

3.10.2. Application to other generalizations

According to the preceding subsection the density of uncertainty involved in the bivariate function $f(x,y)$ is $\ln\left|f''_{xy}(x,y)\right|$. It follows that all the preceding results related to $f(x)$ can be duplicated by merely substituting $\left|f''_{xy}(x,y)\right|$ for $|f'(x)|$ in the formulae.

For instance, if $f(x,y,z)$, $z \in D$, is a distributed function indexed by z, then, according to equation (3.96), its S-path entropy is

$$H_{c,c}(f_z(.);\Omega \times D) = \frac{\int_D \int_\Omega \left|f''_{xy}(x,y,z)\right|^c \ln\left|f''_{xy}(x,y,z)\right| dxdydz}{\int_D \int_\Omega \left|f''_{xy}(x,y,z)\right|^c dxdydz} . \tag{3.97}$$

3.11. Path entropies of non-random piecewise continuous functions

3.11.1. Stairwise functions

Preliminary remark on the entropy of random variables. As we mentionned it in the second chapter (the background on information theory), section 2.6, continuous entropy is not the limiting form of discrete entropy when the discretizing span of approximation tends to zero, and in what follows, we shall comment on this point.

Consider the entropy

$$H(X) = -\int_{\Re} p(x) \ln p(x) dx$$

of the random variable $X \in \Re$.

Using $p(x)$ we define the discrete probability distribution $\{p_i\}$ as follows. We select a (small) constant interval length ε, which then defines the family of intervals $\{[x_i, x_{i+1}] \equiv [x_i, x_i + \varepsilon]\}$, and we set

$$p_i := \int_{x_i}^{x_{i+1}} p(x)dx \quad , \tag{3.98}$$

$$= F(x_{i+1}) - F(x_i) \quad ,$$

$$=: \Delta F(x_i) \quad , \tag{3.99}$$

$$= p(x_i)\varepsilon \quad . \tag{3.100}$$

where $F(x)$ is the (cumulative) distribution of X.

This probability distribution $\{p_i\}$ defines a random variable X_ε of which the Shannon entropy is

$$H(X_\varepsilon) = -\sum_i \Delta F(x_i) \ln \Delta F(x_i), \tag{3.101}$$

$$= -\ln \varepsilon - \varepsilon \sum_i p(x_i) \ln p(x_i). \tag{3.102}$$

$H(X_\varepsilon)$ does not converge to $H(X)$ when ε tends to zero, and this feature, of course, questions the nature of the approximation of the latter by the former. Nevertheless, when $F(x)$ is a stairwise function, that is to say when it is the distribution function of a disrete random variable, the approximation (3.101) *formally* provides the corresponding suitable definition of entropy.

In the following we shall assume that the same remark applies to the entropy of non random functions.

Uncertainty involved in stairwise functions. We can now state the following result:

PROPOSITION 3.9. *Let* $Y_d(x)$ *denote the stairwise function*

$$Y_d(x) := \sum_{i=1}^{n} \delta_i Y(x - x_i) \quad , \quad x \in \Re, \tag{3.103}$$

where δ_i, $i = 1,2,...,n$, *denotes constant steps and* $Y(x-x_i)$ *is the Heaviside function. Then as a consequence of the definitions of entropies of continuously differentiable functions its Shannon path entropy of order s is* [3.10]

$$H_c\left(Y_d(.);[m_1,m_2]\right) = \frac{\sum_{i=m_1}^{m_2}|\delta_i|^c \ln|\delta_i|}{\sum_{i=m_1}^{m_2}|\delta_i|^c}, \tag{3.104}$$

its Renyi path entropy of order s is

$$H_{R,s}\left(Y_d(.);[m_1,m_2]\right) = -\frac{1}{1-s}\ln\frac{\sum_{i=m_1}^{m_2}|\delta_i|^s}{\sum_{i=m_1}^{m_2}|\delta_i|}, \tag{3.105}$$

and its structural path entropy of order s is

$$H_{C,s}\left(Y_d(.);[m_1m_2]\right) = \frac{1}{e^{1-s}-1}\left[\frac{\sum_{i=m_1}^{m_2}|\delta_i|}{\sum_{i=m_1}^{m_2}|\delta_i|^s}-1\right].\blacksquare \tag{3.106}$$

Proof. (i) We first notice that the pairs *(f(x),f'(x))* and *(F(x),p(x))* play parallel roles in the theory, and they give exactly the same result, except that the *S*-entropy of random variables and the *S*-entropy of non random functions have opposite signs.

(ii) This remark suggests the approximation (3.101), therefore the result.

3.11.2. Piecewise continuous functions

PROPOSITION 3.10. *Let* $f:\Re\to\Re, x\to f(x)$ *denote a non-random function which is continuously differentiable anywhere but on the finite set* $\{x_1,x_2,...,x_n\}$ *where it exhibits jumps of discontinuity in the form*

$$\delta_i := f(x_i+0)-f(x_i-0);$$

then, as a consequence of the definitions of entropies of continuously differentiable functions, its S-entropy of order c is

$$H_c\left(f(.);[a,b]\right) = \frac{\int_a^b|f'(x)|^c\ln|f'(x)|dx+\sum_{i=1}^n|\delta_i|^c\ln|\delta_i|}{\int_a^b|f'(x)|^c dx+\sum_{i=1}^n|\delta_i|^c}, \tag{3.107}$$

its R-entropy of order s is

$$H_{R,s}(f(.);[a,b]) = -\frac{1}{1-s}\ln\frac{\int_a^b|f'(x)|^s dx+\sum_{i=1}^n|\delta_i|^s}{\int_a^b|f'(x)|dx+\sum_{i=1}^n|\delta_i|}, \tag{3.108}$$

and its structural entropy of order s is

$$H_{C,s}(f(.);[a,b]) = \frac{1}{e^{1-s}-1}\left[\frac{\int_a^b|f'(x)|dx+\sum_{i=1}^n|\delta_i|}{\int_a^b|f'(x)|^s dx+\sum_{i=1}^n|\delta_i|^s}-1\right]. \blacksquare \tag{3.109}$$

Proof. The distribution of the uncertainty involved in $f(.)$ appears to be defined by the density $\ln|f'(x)|$ on the one hand, and the discrete distribution $\{\ln|\delta_i|\}$ on the other hand. We are then led to define the S-entropy of $f(.)$ on the interval $[a,b]$ by the expression

$$H_c(f(.);[a,b]) := \max_{p,\{q_i\}}\left(\int_a^b p(x)\ln|f'(x)|dx+\sum_{i=1}^n q_i\ln|\delta_i|\right), \tag{3.110}$$

subject to the condition

$$-\int_a^b p(x)\ln p(x)dx = constant, \tag{3.111}$$

$$-\sum_{i=1}^n q_i\ln q_i = constant, \tag{3.112}$$

$$\int_a^b p(x)dx = 1, \tag{3.113}$$

$$\sum_{i=1}^n q_i = 1, \tag{3.114}$$

therefore the expression (3.107)

3.12. Path entropies of functions defined in the complex plane

3.12.1. Analytic functions

Our purpose in the following is to show how one can extend, in a very simple manner, the above definitions to functions which are defined in the complex plane, and to this end we shall firstly consider analytic functions. We can state:

PROPOSITION 3.11. *Let* $f : C \to C, \xi \to f(\xi)$ *denote a non-random analytic (or holomorphic) function. As a direct consequence of the definition of S-entropy of non-random functions in* $\Re^2$, *its S-entropy of order c is defined by the expression*

$$H_c(f(.);\Omega) = \frac{\int_\Omega |f'(\xi)|^{2c} \ln |f'(\xi)|^2 d\xi}{\int_\Omega |f'(\xi)|^{2c} d\xi}, \quad c \in \Re. \blacksquare \tag{3.115}$$

Proof. We write $f(\xi)$ in the form $f(\xi) = u(x,y)+iv(x,y)$ and we identify the *S*-entropy of $f(.)$ with the entropy $H_c(\phi(.);\Omega)$ of the $\Re^2 \to \Re^2$ function $\phi := (u, v)$. This entropy involves the Jacobian determinant

$$|\phi'| = |u'_x v'_y - u'_y v'_x|,$$

but since $f(\xi)$ is analytic, one has the equalities $u'_x = v'_y$ and $u'_y = -v'_x$, which provides $|\phi'| = |f'|^2$ therefore the result.

R-entropy and C-entropy can be obtained in exactly the same way.

3.12.2. Real-valued functions defined in the complex plane

PROPOSITION 3.12. *Let* f: $C \to \Re$, $\xi \to f(\xi,\xi^*)$, *where* ξ^**is the conjugate of* ξ, *denote a real valued function defined in the complex plane. Define*

$$D^2 f(\xi,\xi^*) := \frac{\partial^2 f}{\partial \xi^2} - \frac{\partial^2 f}{\partial \xi^{*2}} \tag{3.116}$$

Then, as a direct consequence of the definition of S-entropy of $\Re^2 \to \Re$ *non-random functions, its S-entropy of order c on the domain* Ω *is defined by the expression*

$$H_c(f(.,.);\Omega) = \frac{\int_\Omega |D^2 f(\xi,\xi^*)|^c \ln |D^2 f(\xi,\xi^*)| d\xi d\xi^*}{\int_\Omega |D^2 f(\xi,\xi^*)|^c d\xi d\xi^*}, \quad c \in \Re. \blacksquare \tag{3.117}$$

Proof. Using the transformation

$$\begin{aligned} \xi &= x+iy \\ \xi^* &= x-iy \end{aligned} \tag{3.118}$$

we have the identities

$$\begin{aligned} f(\xi,\xi^*) &= g(x,y), \\ g''_{xy}(x,y) &= i\,D^2 f(\xi,\xi^*), \end{aligned}$$

and

$$d\xi d\xi^* = 2dxdy, \tag{3.119}$$

therefore the result.

3.13. Concluding remarks

The present modelling of information of non-random functions has been obtained by combining two principles which are general enough to be of some help for further extensions.

(i) First, by using well known equations related to the uncertainty $U(X)$ of continuous random variables (i.e., Shannon entropy or another definition of entropy) we split $U(f(X))$ in the form

$$U(f(X)) =: U(X)+U(F(.)|X), \tag{3.120}$$

and then,

(ii) we noticed that entropy is nothing else but the maximum value of conditional entropy.

The point of importance is that the equation (3.12) applies to continuous variables only (there is conservation of uncertainty for discrete variables) and this is one of the reasons why we argued that the definition of entropy of continuous variables is possibly more accurate than its parallel for discrete variables [3.8]. Nevertheless, if we describe discrete variables in terms of total entropy (see Chapter 2) then equation (3.12) holds.

Maximizing conditional entropy looks like applying Jaynes' maximum entropy principle [3.4,3.5], but it is a semblance only. The maximum conditional entropy

principle is a direct consequence of the basic equations of information theory, and instead we should rather consider that it is the maximum entropy principle which is an extension of the maximum conditional entropy principle.

References

3.1. ACZEL.J. and DAROCZY. Z.; *On Measures of Information and their Characterizations*, Academic Press, New York, 1975

3.2. DUPAIN, Y., KAMAE, T. and MENDES FRANCE, M.; Can one measure the temperature of a Curve? *Arch. Rational Mechanics*, Vol 94, pp 155-163, 1986

3.3 HAKEN, H.; *Advanced Synergetics,* Springer Verlag, Berlin, 1983

3.4 JAYNES, E.T.; Information theory ans statistical mechanics I, *Phys. Rev.,* Vol 106, pp 620-630, 1957

3.5 JAYNES, E.T.; Information theory and statistical mechanics II, *Phys. Rev.*, Vol 108, pp 171-190, 1957

3.6 JUMARIE, G.; *Subjectivity, Information, Systems. Introduction to a Theory of Relativistic Cybernetics*, Gordon and Breach, London, 1986

3.7 JUMARIE, G.; New concepts in information theory, in *Quantum Physics and Chaos, Proceedings the Adriatico Conference* (International Centre of Theoretical Physics), Vol RS 9, pp 68-73, 1987

3.8 JUMARIE, G.; *Relative Information, Theories and Applications*, Springer Verlag, Berlin, 1990

3.9 JUMARIE, G.; Further results in the information theory of deterministic functions and its applications to pattern recognition, *Annals of Telecommunications*, Vol 45, Nos 1-2, pp 66-88, 1990

3.10 JUMARIE, G.; A new information theoretic approach to the entropy of non random discrete maps. Relation to fractal dimension and temperature of curves, *Chaos, Solitons and Fractals*, Vol 8, No 6, pp 953-970, 1997

3.11. KOLMOGOROV, A.N.; *Dokl.. Acad. Nauk USSR,* Vol 98, pp 527-532, 1959

3.13. NICOLIS, J.S., MEYER-KRESS, G., and HAUBS, G.; Non-uniform chaotic dynamics with implications, *Zeit.für Naturforschung*, Vol 38a, pp 1157-1169, 1993

3.13 SCHUSTER, H.G.; *Deterministic Chaos*, Physik Verlag, Weinheim, 1984

3.14 THOM, R.; *Stabilié Structurelle et Morphogéneese*, Benjamin, Reading, 1972

Chapter 4

Path Entropies of Random Functions and of Non-Random Distributed Functions

A throw of dice will never abolish chance.
Stéphane Mallarmé

In the present chapter, we shall show how one can extend the preceding concept of entropy of non-random functions in order to obtain a new approach to the entropy of stochastic processes on the one hand, and a possible modelling for the entropy of non-random distributed functions, on the other. We shall consider a stochastic process as a random function, which will allow us to define the random entropy of a path sample of the process. To some extent the rationale of this approach is similar to the point of view which provides the definition of the so called almost sure convergence. Then we shall apply the results so obtained to distributed functions by using a randomization technique.

4.1. Path entropies of random functions

4.1.1. Observation modes of family of functions

As we shall understand later, the amount of uncertainty involved in a family of functions $f(x,\beta)$, where $\beta \in \Re^n$ denotes an external random parameter, will depend upon how these functions, or rather the graphs of these functions, are observed; and in this way we shall consider the following modes of observation.

Serial mode of observation. In this observation process the observer first selects a given value β_0 of β, and then he examines the complete graph (or curve) so generated by the map $f(.,\beta_o)$ on Ω; then he selects another value β_1 for β, he observes $f(.,\beta_1)$, and so on. In this approach, the observer apprehends a collection of *S*-path entropies (or *R*-path entropies or *C*-path entropies), and their combination, in some sense to define, will determine the amount of uncertainty so involved in this family of functions.

Parallel mode of observation. In this model the observer selects a given value x_0 for x, and considers the amount of uncertainty involved in the random variable $f(x_o,\beta)$; then he selects another value x_1; and so on. Here the observer firstly considers the mean value (with respect to β) of the density of uncertainty $\ln|f'(x,\beta)|$, and then

G. Jumarie, *Maximum Entropy, Information Without Probability and Complex Fractals*
© Springer Science+Business Media Dordrecht 2000

combines them to measure the uncertainty involved in the family of functions indexed by β.

4.1.2. Shannon path entropies of random functions

PROPOSITION 4.1 *Assume that the* $\Re^n \to \Re^n$ *function* $f(x,\beta)$ *depends explicitly upon a random parameter* $B \in D \subset \Re^m$ *independent of* x, *which takes on the value* β *with the probability density* $q(\beta)$; *then in the serial mode of observation, its S-path entropy of order c conditional to B is*

$$H_c^{Sr}(f_B(.);\Omega \times D|B) = \int_D q(\beta) H_c(f(.,\beta);\Omega) d\beta, \tag{4.1}$$

and in the parallel mode of observation, this conditional entropy is

$$H_c^{\Pr}(f_B(.);\Omega \times D|B) =$$

$$= \frac{\int_\Omega \exp\left(c\, E_B\left\{\ln |f_x'(x,B)|\right\}\right) E_B\left\{\ln |f_x'(x,B)|\right\} dx}{\int_\Omega \exp\left(c\, E_B\left\{\ln |f_x'(x,B)|\right\}\right) dx}, \tag{4.2}$$

with

$$E_B\left\{\ln |f_x'(x,B)|\right\} := \int_D q(\beta) \ln |f_x'(x,\beta)| d\beta. \quad \blacksquare \tag{4.3}$$

Proof. (i) For a given β the graph of $f(x,\beta)$ is well defined, and the transformation $Y_\beta = f(X,\beta)$ yields the conditional entropy

$$H(Y|\beta) = H(X) + \int_\Omega p(x) \ln |f_x'(x,\beta)| dx. \tag{4.4}$$

(ii) In the serial mode of observation, according to equation (4.4), for a given β, the uncertainty seized by the observer is $H_c(f(.,\beta)$, hence the equation (4.1).

(iii) In order to obtain the equation associated with the parallel mode of observation, we multiply both sides of (4.4) by $q(\beta)$ and we integrate over D to have (do not forget that B is independent of$X$)

$$H(Y|B) = H(X) + \int_D \int_\Omega q(\beta) p(x) \ln |f_x'(x,\beta)| dx d\beta, \tag{4.5}$$

$$= H(X) + \int_{\Omega} p(x) E_B\left\{\ln\left|f'_x(x,B)\right|\right\} dx \quad . \tag{4.6}$$

We now define the entropy

$$H_c^{\Pr}(f_B(.);\Omega\times D) \;:=\; \max_{p(.)} \int_{\Omega} p(x) E_B\left\{\ln\left|f'_x(x,\beta|\right\} dx \quad , \tag{4.7}$$

subject to the condition ($H(X)$ = constant), and we use the Lagrange multipiers to obtain (4.2).

It is important to notice that, as so defined, these entropies are conditional on B, since they explicitly depend upon the density $q(\beta)$.

4.1.3. Renyi path entropies of random functions

PROPOSITION 4.2. *Consider the* $\Re^n \to \Re^n$ *random function* $f(x,\beta)$ *of Proposition 4.1. equation (4.1), then in the serial mode of observation, its Renyi path entropy of order* s, $s \geq 0$, $s \neq 1$, *conditional to B, is*

$$H_{R,s}^{Sr}\left(f_B(.);\Omega\times D,u|B\right) \;=\; \int_D q(\beta) H_{R,s}(f(.,\beta);\Omega,u)\, d\beta, \tag{4.8}$$

and, in the parallel mode of observation, its R-path entropy of order s, which corresponds to $u = 0$, *is*

$$H_{R,s}^{\Pr}\left(f_B(.);\Omega\times D|B\right) \;=\; -\frac{1}{1-s}\ln\frac{\int_{\Omega} E_s^s\left\{\left|f'_x(x,B)\right|\right\} dx}{\int_{\Omega} E_s\left\{\left|f'_x(x,B)\right|\right\} dx}, \tag{4.9}$$

with the notation

$$E_s\left\{\left|f'_x(x,B)\right|\right\} \;:=\; \left[\frac{\int_D q^s(\beta)\left|f'_x(x,\beta)\right|^{1-s} d\beta}{\int_D q^s(\beta) d\beta}\right]^{\frac{1}{1-s}} . \blacksquare \tag{4.10}$$

Proof. (i) In the serial mode of observation, for a given β, the observer measures $H_{R,s}(f(.,\beta;\Omega,u)$, hence the expression (4.8), by assuming that u is constant.

(ii) We now consider the parallel mode of observation, and we make the transformation $(X,B) \rightarrow (Y,B) \equiv f(X,B),B)$ where X is a random vector with the probability density $p(x)$; then according to the transformation (3.23), one can write

$$H_{R,s}(Y) = \frac{1}{1-s} \ln \int_{\Omega}\int_{D} q^{s}(\beta)p^{s}(x)\left|f'_{x}(x,\beta)\right|^{1-s} d\beta dx, \tag{4.11}$$

and similarly to the maximization (3.24) which defines the Renyi path entropy of non-random functions, we shall define the Renyi path entropy $H_{R,s}(f_B(.))$ of random functions by the expression

$$H_{R,s}(f_B(.);\Omega \times D)$$

$$:= \max_{p(.)} \frac{1}{1-s} \ln \frac{\int_{\Omega}\int_{D} q^{s}(\beta)p^{s}(x)\left|f'_{x}(x,\beta)\right|^{1-s} d\beta dx}{\int_{\Omega}\int_{D} q^{s}(\beta)p^{s}(x) d\beta dx}, \tag{4.12}$$

subject to the condition

$$H_{R,s}(X,B) := \frac{1}{1-s} \ln \int_{\Omega}\int_{D} q^{s}(\beta)p^{s}(x)dx \ .$$
$$= constant \ .$$

(iii) In the parallel mode of observation, the equation (4.12) provides

$$H_{R,s}(f_B(.);\Omega \times D)$$

$$= \max_{p(x)} \frac{1}{1-s} \ln \frac{\int_{\Omega} p^{s}(x)\mathrm{E}_{s}^{1-s}\left\{\left|f'_{x}(x,B)\right|\right\}dx}{\int_{\Omega} p^{s}(x)dx}, \tag{4.13}$$

with the condition

$$\int_{\Omega} p^{s}(x)dx = constant \ .$$

We then optimize the Lagrangian

$$L := \int_{\Omega}\left[p^{s}(x)\left(E_{s}^{1-s}\left\{\left|f_{x}'(x,B)\right|\right\}+\lambda\right)+\mu p(x)\right]dx,$$

where λ and μ are the multipliers, and we so obtain the special form of the optimal $p(x)$, corresponding to $u = 0$ in the expression of $H_{R,s}(f(.);\Omega,u)$, i.e.,

$$p(x) = \frac{E_{s}\left\{\left|f_{x}'(x,B)\right|\right\}}{\int_{\Omega}E_{s}\left\{\left|f_{x}'(x,B)\right|\right\}dx},$$

therefore the equation (4.9).

This complete the proof.

One has the following result:

PROPOSITION 4.3. *The following equalities hold:*

$$\lim_{s\to 1} H_{R,s}^{Sr}(f_{B}(.);\Omega\times D,u|B) = H_{1/(1+u)}^{Sr}(f_{B}(.);\Omega\times D|B), \tag{4.14}$$

and

$$\lim_{s\to 1} H_{R,s}^{\Pr}(f_{B}(.);\Omega\times D|B) = H_{1}^{\Pr}(f_{B}(.);\Omega\times D|B). \quad \blacksquare \tag{4.15}$$

Proof. The proof of (4.14) is straightforward and is a direct consequence of Proposition (3.4).

The proof of (4.15) reads as follows. (i) According to (4.10), one has the equivalence

$$\begin{aligned} E_{s}\left\{\left|f'\right|\right\} &\cong \left(\int_{D}q(\beta)\left|f'\right|^{1-s}d\beta\right)^{\frac{1}{1-s}} \\ &\cong \left(\int_{D}q(\beta)\left[1+(1-s)\ln\left|f'\right|\right]d\beta\right)^{\frac{1}{1-s}}, \\ &\cong \left[1+(1-s)\int_{D}q(\beta)\ln\left|f'\right|d\beta\right]^{\frac{1}{1-s}} \\ &\cong \exp\left(E_{B}\left\{\ln\left|f'\right|\right\}\right). \end{aligned} \tag{4.16}$$

(ii) By using (4.16) one can write

$$\begin{aligned} E_{s}^{s}\left\{\left|f'\right|\right\} &\cong \exp\left(sE_{B}\left\{\ln\left|f'\right|\right\}\right) \\ &\cong \exp\left(E_{B}\left\{\ln\left|f'\right|\right\}\right)\left[1+(s-1)\exp\left(E_{B}\left\{\ln\left|f'\right|\right\}\right)\right]. \end{aligned} \tag{4.17}$$

(iii) Substituting (4.16) and (4.17) into (4.9) yields the equality

$$H^{\mathrm{Pr}}_{C,s}(f_B(.);\Omega\times D|B)$$

$$\cong\ -\frac{1}{1-s}\ln\left[1+(s-1)\frac{\int_\Omega \exp\left(E_B\{\ln|f'|\}\right)E_B\{\ln|f'|\}dx}{\int_\Omega \exp\left(E_B\{\ln|f'|\}\right)dx}\right],$$

therefore the result.

4.1.4. Structural path entropies of random functions

PROPOSITION 4.4. *Consider the* $\Re^n \to \Re^n$ *random function* $f(x,\beta)$ *of Proposition (4.1), then in the serial mode of observation its structural path entropy of order* $s \geq 0$ *conditional to* B *is*

$$H^{Sr}_{C,s}(f_B(.);\Omega\times D,u|B) \ =\ \int_D q(\beta)H_{C,s}(f(.,\beta);\Omega,u)d\beta\ , \tag{4.18}$$

and in the parallel mode of observation, the structural path entropy corresponding to $u = 0$ *is*

$$H^{\mathrm{Pr}}_{C,s}(f_B(.);\Omega\times D,u|B) \ =\ \frac{1}{e^{1-s}-1}\left[\frac{\int_\Omega E_s\{|f'_x(x,B)|\}dx}{\int_\Omega E^s_s\{|f'_x(x,B)|\}dx}-1\right].\blacksquare \tag{4.19}$$

Proof. The proof is parallel to the proofs of proposition 4.1 and 4.2, and we shall only outline it below.

(i) In the serial mode of observation, the result is now obvious.

(ii) In the parallel mode of observation, using again the transformation $(X,B)\to(Y,B)\equiv(f(X,B),B)$, one obtains the equality

$$H_{C,s}(Y) \ =\ \frac{\int_\Omega\int_D q^s(\beta)p^s(x)|f'_x(x,\beta)|^{1-s}dx-1}{e^{1-s}-1}\ , \tag{4.20}$$

which provides the definition of $H_{C,s}(f_B(.);\Omega)$ in the form

$$H_{C,s}(f_B(.);\Omega\times D)$$

$$= \max_{p(.)} \frac{\int_\Omega\int_D \left(|f'_x(x,\beta)|^{1-s}-1\right)p^s(x)q^s(\beta)dxd\beta}{\int_\Omega\int_D p^s(x)q^s(\beta)dxd\beta}, \tag{4.21}$$

subject to the condition

$$H_{C,s}(X,B) := \frac{\int_\Omega\int_D p^s(x)q^s(\beta)dxd\beta - 1}{e^{1-s}-1}$$
$$= constant\ . \tag{4.22}$$

Since B is independent of X, one has also

$$H_{C,s}(f_B(.);\Omega\times D)$$

$$= \max_{p(.)} \frac{\int_\Omega p^s(x)\left[E^{1-s}\left\{|f'_x(x,B)|\right\}-1\right]dx}{\int_\Omega p^s(x)dx}, \tag{4.23}$$

subject to the condition

$$\int_\Omega p^s(x)dx = constant\ . \tag{4.24}$$

The usual optimization via Lagrange multipliers yields equation (4.19).

We shall conclude this section on random functions by the following result:

PROPOSITION 4.5. *The following equalies holds:*

$$\lim_{s\to 1} H^{Sr}_{C,s}(f_B(.);\Omega\times D,u|B) = H^{Sr}_{1/(1+u)}(f_B(.0;\Omega\times D|B), \tag{4.25}$$

and

$$\lim_{s\to 1} H^{\Pr}_{C,s}(f_B(.);\Omega\times D|B) = H^{\Pr}_1(f_B(.);\Omega\times D|B)\cdot \blacksquare \tag{4.26}$$

The proof is now familiar and is left to the reader.

4.2. Application to stochastic processes

4.2.1. Random path entropies of stochastic proceses

As is evident, the above definitions apply directly to stochastic processes, but here we shall reconsider the problem of defining the entropy of random functions from a new point of view (the entropy of a path sample is obtained by using the definition of entropy of non-random functions) which will lead us to introduce a new concept of random path entropy.

DEFINITION 4.1. Let $x(t) \in \Re$ denote a mean square differentiable stochastic process, and consider the graph (or trajectory) defined by a given sample $\{x(t),\ 0 \le \tau \le T\}$ on the interval $[0,T]$. Then three measures of the amount of uncertainty involved in this sample trajectory, and which are fully consistent with the entropy of non-random functions, are the Shannon path entropy

$$H_c(x(.);0,T) = \frac{\int_0^T |\dot{x}(t)|^c \ln |\dot{x}(t)| dt}{\int_0^T |\dot{x}(t)|^c dt}, \tag{4.27}$$

the Renyi path entropy

$$H_{R,s}(x(.);),0,T;u)$$

$$= \frac{1}{1-s} \ln \frac{\int_0^T \left(u + |\dot{x}(t)|^{1-s}\right)^{\frac{1}{1-s}} |\dot{x}(t)|^{1-s} dt}{\int_0^T \left(u + |\dot{x}(t)|^{\frac{1}{1-s}}\right) dt}, \tag{4.28}$$

and the structural path entropy

$$H_{C,s}(x(.);0,T;u)$$

$$= \frac{1}{e^{1-s}-1} \frac{\int_0^T \left(u + |\dot{x}(t)|^{1-s}\right)\left(|\dot{x}(t)|^{1-s} - 1\right) dt}{\int_0^T \left(u + |\dot{x}(t)|^{\frac{s}{1-s}}\right) dt} \ . \ \blacksquare \tag{4.29}$$

Useful approximation. Let us write

$$H_c(x(.);0,T) \quad =: \quad Y_1 / Y_2 \ .$$

Assume that $\{\dot{x}(t_1),\dot{x}(t_2),...,\dot{x}(t_k)\}$ is a set of mutually independent random variables for any $\{t_o,t_1,...,t_k\}$ and any k, and define

$$\mu_1(t,c) \ := \ E\left\{|\dot{x}(t)|^c \ln|\dot{x}(t)|\right\}, \tag{4.30}$$

$$\sigma_1^2(t,c) \ := \ Var\left\{|\dot{x}(t)|^c \ln|\dot{x}(t)|\right\}, \tag{4.31}$$

$$\mu_2(t,c) \ := \ E\left\{|\dot{x}(t)|^c\right\} \ , \tag{4.32}$$

$$\sigma_2^2(t,c) \ := \ Var\left\{|\dot{x}(t)|^c\right\}, \tag{4.33}$$

then, according to the central limit theorem, the numerator Y_1 in the expression of H_c is a Gaussian random variable,

$$Y_1 \ \approx \ N\left(\int_0^T \mu_1(t,c)dt, \int_0^T \sigma_1^2(t,c)\right), \tag{4.34}$$

and in a like manner Y_2 is a Gaussian random variable

$$Y_2 \ \approx \ N\left(\int_0^T \mu_2(t,c)dt, \int_0^T \sigma_2^2(t,c)\right) \cdot \tag{4.35}$$

4.2.2. Average path entropy of order k

In some instances, it may be of some interest to consider the average entropy of order k defined by the expression [4.2]

$$\left\langle H_c^k(x(.);0,T)\right\rangle \ := \ \frac{\int_0^T |\dot{m}_k(t)|^c \ln|\dot{m}_k(t)|dt}{\int_0^T |\dot{m}_k(t)|^c dt} \quad , \quad k \in N_+, \tag{4.36}$$

with

$$\dot{m}_k(t) \;:=\; \left[E\left\{\dot{x}^k(t)\right\}\right]^{1/k}. \tag{4.37}$$

4.3. Path entropies of functions indexed by a distributed parameter

4.3.1 Shannon path entropy of distributed functions

We now consider a family of function $f(x,z)$ indexed by the vector parameter z, which differs from the family of random function $f(x,B)$ above in the sense that z is a deterministic parameter. We have the following result.

PROPOSITION 4.6. *Let* $f:\Omega\times D\to\Re^n,\Omega\subset\Re^n,D\subset\Re^m,(x,z)\to f(x,z)$, *denote a non-random function which is continuous with respect to z and is continuously differentiable w.r.t. x, and the Jacobian determinant of which is denoted by* $f'_x(x,z)$. *Then as a direct consequence of the definition of Shannon entropy of random variables, in the serial mode of observation, its S-path entropy of order* $(b,c(z))$, $b\in\Re$, $c(z)\in\Re$, *on the domain* $\Omega\times D$, *is given by the expression*

$$H_{b,c(z)}(f_z(.);\Omega\times D)$$

$$:=\;\frac{\int_D \exp\left(bH_{c(z)}(f(.,z);\Omega)+H\left(X_f;z,c(z)\right)\right)H_{c(z)}(f(.,z);\Omega)dz}{\int_D \exp\left(bH_{c(z)}(f(.,z);\Omega)+H\left(X_f;z,c(z)\right)\right)dz}, \tag{4.38}$$

where $H(X_f;z,c(z))$ *is the Shannon entropy of the random variable* X_f *with the probability density*

$$p_{f,c(z)}(x,z) \;:=\; \frac{\left|f'_x(x,z)\right|^{c(z)}}{\int_\Omega \left|f'_x(x,z)\right|^{c(z)}dx}, \tag{4.39}$$

and where b together with c(z) denote two parameters which describe how the family of curves so defined by f(.,z) is scanned by the observation process. ■

Proof. (i) We randomize z in the form of the random variable Z with the probability density $\rho(z)$; and we point out that this randomization is quite relevant since then $\rho(z)$ defines the scanning distribution with which the observer (whether it is a human being or a physical device) examines each x-trajectory $f(.,z)$, $z\in D$.

(ii) This being so, according to the relation (3.12) which now depends upon z, the transformation $Y=f(X,z)$ yields the conditional entropy

$$H(Y|z) = H(X|z) + \int_D p(x,z) \ln |f_x'(x,z)| dx \ , \tag{4.40}$$

where $H(X|x)$ is the conditional entropy of X given that $Z = z$, as defined by the conditional probability density

$$p(x,z) := q(f(x,z))|f_x'(x,z)|,$$

with $q(y)$ denoting the probability density of Y.

We now multiply both sides of this equation by $\rho(x)$ and we integrate over D to obtain

$$H(Y|Z) = H(X|Z) + \int_D \int_\Omega \rho(z) p(x,z) \ln |f_x'(x,z)| dxdz \ , \tag{4.41}$$

with

$$H(X|Z) := -\int_D \int_\Omega \rho(z) p(x,z) \ln \ p(x,z) dxdz \ . \tag{4.42}$$

(iii) Next, we consider the pair (Y,Z), and analogously with the equation (3.13) which defines the path entropy of non-random functions we expand the joint entropy $H(Y,Z)$ in the form

$$\begin{aligned} H(Y,Z) &=: H(X,Z,f_z(.);\Omega \times D) \\ &= H(X,Z) + H(f_z(.);\Omega \times D|X,Z) \\ &= H(Z) + H(X|Z) + H(f_z(.);\Omega \times D|X,Z), \end{aligned} \tag{4.43}$$

and we notice that

$$H(Y,Z) = H(Z) + H(Y|Z) \ .$$

Taking (3.13) into account, (4.43) provides the identification

$$H(f_z(.);\Omega \times D|X,Z) := \int_D \int_\Omega \rho(z) p(x,z) \ln |f_x'(x,z)| dxdz \ , \tag{4.44}$$

together with

$$H(f_z(.);\Omega\times D) \;:=\; \max_{\rho,p}\; H(f_z(.);\Omega\times D|X,Z), \tag{4.45}$$

subject to the condition

$$H(Z)+H(X|Z) \;=\; constant\ . \tag{4.46}$$

Next, since we are dealing with the serial mode of observation, according to section 3.2 on path entropies of non-random functions, one has necessarily the equality

$$p(x,z) \;=\; p_{f,c(z)}(x,z)\quad , \tag{4.47}$$

where $p_{f,c(z)}(x,z)$ is given by the expression (4.39); and as a result, equation (4.45) yields

$$H(f_z(.);\Omega\times D) \;=\; \max_{\rho}\int_D \rho(z)\mathrm{H}_{c(z)}(f(.,z);\Omega)dz\quad ,$$

subject to the condition

$$-\int_D \rho(z)\ln\,\rho(z)dz+\int_D \rho(z)H(X_f,z)dz \;=\; const\,ant\ ,$$

with the additional normalizing condition

$$\int_D \rho(z)dz \;=\; 1\ .$$

(v) We then define the Lagrangian

$$L \;=\; \int_D \rho(z)\Big\{H_{c(z)}(f(.,z);\Omega)+\lambda\Big(H(X_f,z)-\ln\rho(z)\Big)+\mu\Big\},$$

to obtain the result (4.38).

A useful special case. Assume that $c(z)$ is a constant c in the equation (4.38); then, by using the explicit form of the entropy $H(X_f)$ as expressed in terms of $p_{f,c}(x,z)$, one obtains the expression

$$H_{b,c}(f_z(.);\Omega\times D) \tag{4.48}$$

$$= \frac{\int_D \exp\big((b-c)H_c(f(.,z);\Omega)\big)\Big[\int_\Omega |f'_x(x,z)|^c \ln|f'_x(x,z)|dx\Big]dz}{\int_D \exp\big((b-c)H_c(f(.,z);\Omega)\big)\Big[\int_\Omega |f'_x(x,z)|^c dx\Big]dz}.$$

Assume further that $b = c$, then (4.48) provides the simplified useful expression

$$H_{c,c}(f_z(.);\Omega\times D) = \frac{\int_D\int_\Omega |f'_x(x,z)|^c \ln|f'_x(x,z)|dxdz}{\int_D\int_\Omega |f'_x(x,z)|^c dxdz}. \tag{4.49}$$

On the surface this expression looks like a direct generalization of the definition of $H_C(f(.);\Omega)$. Nevertheless, it is of importance to point out that it is not merely a formal extension at all, but rather a direct consequence of the basic argument which allowed us to obtain $H_C(f(.);\Omega)$.

Notice also that the assumption $b = c$ is quite acceptable and meaningful. Indeed, on the practical standpoint, and loosely speaking, it merely illustrates that the scanning observation of each trajectory on the one hand, and that of the family of trajectories, on the other, are similar.

4.3.2. Renyi path entropy of distributed functions

PROPOSITION 4.7. *Consider the non random distributed function f(x,z) of Proposition* 4.5, *equation* (4.38); *then, in the serial mode of observation, a measure* $H_{R,s}(f_z(.);\Omega\times D,u)$ *of the amount of uncertainty it involves, which furthermore is fully consistent with the Renyi path entropy of random variables, is defined by the expression*

$$H_{R,s}\big(f_z(.);\Omega\times D,v\big) = -\frac{1}{1-s}\ln\frac{\int_D A(z)dz}{\int_D B(z)dz}, \tag{4.50}$$

with

$$A(z) := \Big[V^{1-s}(f;z,\Omega)+v\exp\big((1-s)H_{R,s}(X_f;z)\big)\Big]^{\frac{s}{1-s}} \times V^{-s}(f;z,\Omega)\int_\Omega |f'_x(x,z)|^s dx, \tag{4.51}$$

and

$$B(z) \;:=\; \left[V^{1-s}(f;z,\Omega)+v\exp\!\left((1-s)H_{R,s}(X_f;z)\right)\right]^{\frac{s}{1-s}} V^{1-s}(f;z,\Omega). \tag{4.52}$$

$v \in \Re$ *is a real valued parameter;* $V(f;z,\Omega)$ *is the total variation of* $f(.,z)$ *on* Ω, *i.e.*

$$V(f;z,\Omega) \;:=\; \int_\Omega \left|f'_x(x,z)\right| dx \;, \tag{4.53}$$

and $H_{R,s}(X;z)$ *is the Renyi entropy of the random variable* X *defined by the probability density* $p_{f,1}(x,z)$, *see equation* (4.28), *clearly*

$$H_{R,s}(X_f;z) \;:=\; \frac{1}{1-s}\ln\frac{\int_\Omega \left|f'_x(x,z)\right|^s dx}{\left[\int_\Omega \left|f'_x(x,z)\right| dx\right]^s} \;. \;\blacksquare \tag{4.54}$$

Proof. (i) Once again, we randomize z in the form Z with the probability density $\rho(z)$.

This being so, we consider the transformation $(X,Z) \to (Y,Z) \equiv (f(X,Z),Z)$, where X denotes a random variable with the conditional probability density $p(x,z)$ given z. According to the transformation equation (3.14) one has

$$H_{R,s}(Y) \;=\; \frac{1}{1-s}\ln\int_D\int_\Omega \rho^s(z)p^s(x,z)\left|f'_x(x,z)\right|^{1-s} dxdz \;, \tag{4.55}$$

and similarly to (4.45) and (4.46), we shall define $H_{R,s}(f_z(.))$ by the expression

$$H_{R,s}(f_z(.);\Omega\times D)$$

$$= \max_{\rho,p}\frac{1}{1-s}\ln\frac{\int_D\int_\Omega \rho^s(z)p^s(x,z)\left|f'_x(x,z)\right|^{1-s} dxdz}{\int_D\int_\Omega \rho^s(z)p^s(x,z)dxdz} \;, \tag{4.56}$$

subject to the condition

$$H_{R,s}(X,Z) \;:=\; \frac{1}{1-s}\ln\int_D\int_\Omega \rho^s(z)p^s(x,z)dxdz \;,$$
$$= \; constant \;. \tag{4.57}$$

(ii) In the serial mode of observation we first select $p(x,z)$, and according to the equation (3.25), in which we make $u = 0$ to simplify it, one has the density

$$p(x,z) \;=\; p_{f,c(z)}(x,z). \tag{4.58}$$

We now substitute $p_{f,c(z)}(x,z)$ for $p(x,z)$ into (4.54), and as a result the problem is then to maximize

$$\max_{\rho} \int_D \rho^s(z) M(z) dz \;, \tag{4.59}$$

subject to the condition

$$\int_D \rho^s(z) N(z) dz \;=\; constant \;, \tag{4.60}$$

with the notation

$$M(z) \;:=\; \int_\Omega p^s_{f,1}(x,z) \left| f'_x(x,z) \right|^{1-s} dx \;, \tag{4.61}$$

$$N(z) \;:=\; \int_\Omega p^s_{f,1}(x,z) dx \;. \tag{4.62}$$

(iii) To this end, we introduce the Lagrangian function

$$L \;:=\; \int_\Omega \left[\rho^s(z)\left(M(z) + vN(z)\right) + w\rho(z) \right] dz \;,$$

where v and w are the multipliers (indeed do not forget the constraint $\int_D \rho(z) dz = 1$) and maximizing L with respect to $\rho(.)$ yields

$$\rho(z) \;=\; \frac{\left[M(z) + vN(z)\right]^{1/(1-s)}}{\int_D \left[M(z) + vN(z)\right]^{1/(1-s)} dz} \;. \tag{4.63}$$

We substitute this result into (4.45) to have

$$H_{R,s}(f_z(.);\Omega\times D,v)$$

$$= -\frac{1}{1-s}\ln\frac{\int_D\left[M(z)+vN(z)\right]^{\frac{s}{1-s}}V^{-s}(f;z,\Omega)\left[\int_\Omega\left|f'_x(x,z)\right|^s dx\right]dz}{\int_D\left[M(z)+vN(z)\right]^{\frac{s}{1-s}}V^{1-s}(f;z,\Omega)dz}. \qquad (4.64)$$

(iv) In order to obtain the result, we simply remark that

$$M(z) = V^{1-s}(f;z,\Omega)$$

and

$$N(z) = \exp\left((1-s)H_{R,s}(X_f;z)\right).$$

A useful special case. If we make $v = 0$ in the equation (4.64), we have the simplified expression

$$H_{R,s}(f_z(.);\Omega\times D) = -\frac{1}{1-s}\ln\frac{\int_D\int_\Omega\left|f'_x(x,z)\right|^s dxdz}{\int_D\int_\Omega\left|f'_x(x,z)\right|dxdz}, \qquad (4.65)$$

which is exactly the parallel of the simplified form (4.49) of the *S*-entropy.

As a last remark we shall point out that the distributed Renyi entropy so obtained is not the most general one, and this is because we have restricted ourselves to the special case defined by equation (4.58). As a matter of fact, the generalized Renyi entropy would be obtained by using the general expression of the optimal $p(x,z)$ which is, according to (3.25)

$$p(x,z;u) = \frac{\left[u+\left|f'_x(x,z)\right|\right]^{s/(1-s)}}{\int_\Omega\left[u+\left|f'_x(x,z)\right|\right]^{s/(1-s)}dx}. \qquad (4.66)$$

In other words, according to the identification $c = 1/(1+u)$, see equation (3.55), the family of Renyi entropies so obtained corresponds to a constant $c(z) = 1$.

The meaning of the parameter v introduced in the equation (4.39) is quite different from that of u, and we shall see later that u plays a role similar to that of b in equation (4.27), whilst v is equivalent to c.

4.3.3. Structural path entropy of distributed functions

PROPOSITION 4.8. *Consider the family of distributed functions f(x,z) of proposition 4.5, equation (4.38); then in the serial mode of observation, a measure* $H_{C,s}(f_z(.);\Omega\times D,v)$ *of the amount of uncertainty it involves, which furthermore is fully consistent with the structural entropy of random variables, is defined by the expression*

$$H_{C,s}\left(f_z(.);\Omega\times D,v\right) \;:=\; \frac{1}{e^{1-s}-1}\left[\frac{\int_D B(z)dz}{\int_D A(z)dz}-1\right], \tag{4.67}$$

where A(z) and B(z) are defined by the equations (4.51) and (4.52). ■

Proof. The proof is parallel to the proof of proposition 4.6, and starts with the definition

$$H_{C,s}(f_z(.);\Omega\times D,v)$$

$$= \max_{\rho,p} \frac{\int_D\int_\Omega\left(\left|f'_x(x,z)\right|^{1-s}-1\right)\rho^s(z)p^s(x,z)dxdz}{\int_D\int_\Omega\rho^s(z)p^s(x,z)dxdz}, \tag{4.68}$$

subject to the condition

$$\begin{aligned} H_{C,s}(X,Z) &= \frac{\int_D\int_\Omega\rho^s(z)p^s(x,z)dxdz-1}{e^{1-s}-1} \\ &= constant. \end{aligned} \tag{4.69}$$

The simplified form which corresponds to $v = 0$ is given by the expression

$$H_{C,s}(f_z(.);\Omega\times D) \;=\; \frac{1}{e^{1-s}-1}\left[\frac{\int_D\int_\Omega\left|f'_x(x,z)\right|dxdz}{\int_D\int_\Omega\left|f'_x(x,z)\right|^s dxdz}-1\right], \tag{4.70}$$

which is the parallel of the reduced *S*-entropy (4.49) and of the reduced *R*-entropy (4.54).

4.3.4. Relations between the path entropies of distributed functions

As expected, one has the following result:

PROPOSITION 4.9. *The following property holds, that is*

$$\lim H_{R,s}(f_z(.);\Omega\times D,v) \;=\; H_{1/(1+v)}(f_z(.);\Omega\times D) \qquad as \quad s\to 1\,.\blacksquare \tag{4.71}$$

Proof. (i) For s close to the unity one has the approximations

$$\left[V^{1-s}(z)+v^{(1-s)H_{R,s}(X_f;\Omega)}\right]^{\frac{s}{1-s}}$$

$$= (1+v)^{\frac{s}{1-s}}[V(z)]^{\frac{s}{1+u}}\exp\left(\frac{su}{1+u}H(X_f;\Omega)\right)+o((s-1)^2), \tag{4.72}$$

and

$$\frac{\int_\Omega |f'_x(x,z)|^s dx}{\int_\Omega |f'_x(x,z)|dx} \;=\; 1+(s-1)H(f_z(.);\Omega)+o((s-1)^2)\,. \tag{4.73}$$

(ii) Substituting these expressions into (4.50), and after some manipulations, one finds that

$$\lim_{s\to 1} H_{R,s}(f_z(.);\Omega\times D,v)$$

$$= \frac{\int_D [V(z)]^{\frac{1}{1+u}}\exp\left(\frac{u}{1+u}H(X_f;z,1)\right)H(f_z(.);\Omega)dz}{\int_D [V(z)]^{\frac{1}{1+u}}\exp\left(\frac{u}{1+u}H(X_f;z,1)\right)dz}$$

$$= \frac{\int_D \exp\left(-\frac{u}{1+u}H(f_z(.);\Omega)\right)V(z)H(f_z(.);\Omega)dz}{\int_D \exp\left(-\frac{u}{1+u}H(f_z(.);\Omega)\right)V(z)dz}\,.$$

(iii) We now remark that here one has $c(z) = c = 1$, and comparing with (4.48), we make the identification

$$b - c = -u/(1+u),$$

hence the result.

With respect to the structural entropy, a Taylor expansion in the vicinity of $s = 1$ yields the following result:

PROPOSITION 4.10. *The following property holds, that is*

$$\lim \quad H_{C,s}(f_z(.);\Omega\times D,v) = H_{1/(1+v),1}(f_z(.);\Omega\times D) \quad as \quad s\to 1 \ \blacksquare \qquad (4.74)$$

Notice that this property can be obtained directly by using the relation between *R*-entropy and *C*-entropy. Indeed, comparing their respective expressions (4.50) and (4.67) yields

$$H_{C,s}(f_z(.);\Omega\times D,v) = \frac{\exp\big((1-s)H_{R,s}(f_z(.);\Omega\times D,v)\big)-1}{e^{1-s}-1}, \qquad (4.75)$$

hence the equivalence

$$H_{C,s}(f_z(.);\Omega\times D,v) = H_{R,s}(f_z(.);\Omega\times D,v)+o((s-1)^2) . \qquad (4.76)$$

4.4. Path entropies of distributed functions via parallel observation

In section 4.2 we considered the model in which the patterns generated by the function $f(x,z)$ are observed in a serial manner only; but, in fact, the results so obtained can be directly utilised to obtain the expression of the entropies defined by the parallel observation mode.

The key idea is the following. At a given x the observer measures the set of values of $f(x,z)$ for the admissible z; in other words he is considering the z function $f_x(z)$ indexed by x ($f_x(z)$ is not to be confused with the derivative $f'_x(x,z)$) in such a manner that he is dealing with a family of distributed functions with x as distributed parameters. It follows that the roles of x and z are reverted, and in order to obtain the new formulae, it

will be sufficient to duplicate the equations of Section 4.2, with suitable alterations to take account of this symmetry.

For instance, the counterparts of the serial entropies defined by the expressions (4.49), (4.65) and (4.70) are respectively

$$H_{c,c}(f_x(.);\Omega\times D) = \frac{\int_\Omega\int_D |f_x'(x,z)|^c \ln|f_x'(x,z)|dxdz}{\int_\Omega\int_D |f_x'(x,z)|^c dxdz}, \tag{4.77}$$

$$H_{R,s}(f_x(.);\Omega\times D) = -\frac{1}{1-s}\ln\frac{\int_D\int_\Omega |f_x'(x,z)|^c dxdz}{\int_D\int_\Omega |f_x'(x,z)|dxdz}, \tag{4.78}$$

$$H_{C,s}(f_x(.);\Omega\times D) = \frac{1}{e^{1-s}-1}\left[\frac{\int_D\int_\Omega |f_x'(x,z)|dxdz}{\int_D\int_\Omega |f_x'(x,z)|^s dxdz}-1\right], \tag{4.79}$$

subject to the condition that $f(x,z)$ is continuously differentiable with respect to z.

4.5. Path entropies of piece-wise continuous distributed functions

As it is evident, one can combine the results of section 3.9 with those of section 4.2 to deal with distributed stairwise functions. For instance, for the function

$$\tilde{Y}(x,z) := \sum_{i=1}^{n}\delta_i(z)Y(x-x_i), \quad x\in\Re, \tag{4.80}$$

the serial S-entropy of order c is

$$H_c^{Sr}\left((\tilde{Y}_x(.);[m_1m_2]\right) = \frac{\int_D\sum_{i=m_1}^{i=m_2}|\delta_i(z)|^c \ln|\delta_i(z)|}{\int_D\sum_{i=m_1}^{i=m_2}|\delta_i(z)|^c dz}, \tag{4.81}$$

and the parallel S-entropy is

$$H_c^{\Pr}\left(\tilde{Y}_x(.);[m_1,m_2]\right) = \frac{\int_D \sum_{i=m_1}^{i=m_2} |\delta_i'(z)|^c \ln|\delta_i'(z)| dz}{\int_D \sum_{i=m_1}^{i=m_2} |\delta_i'(z)|^c dz}. \tag{4.82}$$

In a like manner, the extension to piecewise continuous distributed functions is straifgtforward. We shall denote by $\delta_i(z)$ the discontinuity jump

$$\delta_i(z) = f(x_i+0,z) - f(x_i-0,z) \tag{4.83}$$

and, for instance, the serial S-entropy reads

$$H_c\left(f_z(.);[a,b]\right)$$

$$= \frac{\int_D \left[\int_a^b |f_x'(x,z)|^c \ln|f_x'(x,z)| dx + \sum_{i=1}^{i=n} |\delta_i(z)|^c \ln|\delta_i(z)|\right] dz}{\int_D \left[\int_a^b |f_x'(x,z)|^c dx + \sum_{i=1}^{i=n} |\delta_i|^c\right] dz}. \tag{4.84}$$

4.6. Cross-entropic variance of distributed functions

Assume that we have a field of probabilities defined by the distributed (cumulative) distribution functions $F(x,z)$, $x \in \Re$, $z \in D \subset \Re^m$; then on making the substitution $f(x,z) \leftarrow F(x,z)$, and taking account of the equation (3.21) we are led to define the entropy of order c of X on the domain D by the expression

$$H_{c,c}(X;R\times D) = -\frac{\int_D \int_{\Re} p^c(x,z) \ln p(x,z) dx dz}{\int_D \int_{\Re} p^c(x,z) dx dz}; \tag{4.85}$$

and combining this expression with the equation (3.67) we can now define the *distributed cross-entropic variance* of the two probability densities $p(x,z)$ and $q(x,z)$ by the expression

$$\sigma_H^2(q,p;\Omega,D) := \frac{\int_D\int_\Omega q(x,z)\ln^2[q(x,z)/p(x,z)]dxdz}{\int_D\int_\Omega q(x,z)dxdz} - \left[\frac{\int_D\int_\Omega q(x,z)\ln[q(x,z)/p(x,z)]dxdz}{\int_D\int_\Omega q(x,z)dxdz}\right]^2. \quad (4.86)$$

In a like manner one can meaningfully define a cross-entropic variance between the distributed entropies themselves. Indeed, by making $c = 1$ and $b = 0$ into equation (4.48), we have the *S*-entropy (do not forget that the entropy of a random variable is the opposite of the entropy of a non-random function!)

$$H(X;R\times D) = -\frac{\int_D e^{H(X;z)}H(X;z)dz}{\int_D e^{H(X;z)}dz}, \quad (4.87)$$

with

$$H(X;z) := -\int_{\Re} p(x,z)\ln p(x,z)dx. \quad (4.88)$$

As a result, given another distributed entropy *H(Y;z)* defined by the probability density *q(x,z)*,

$$H(Y;z) := -\int_R q(y,z)\ln q(y,z)dy, \quad (4.89)$$

we shall define the cross-entropic variance between *H(X;z)* and *H(Y;z)* by the relation

$$\sigma_H^2(H(X;z),H(Y;z);D) := \frac{\int_D e^{H(Y;z)}[H(Y;z)-H(X;z)]^2dz}{\int_D e^{H(Y;z)}dz} - \left[\frac{\int_D e^{H(Y;z)}[H(Y;z)-H(X;z)]dz}{\int_D e^{H(Y;z)}dz}\right]^2. \quad (4.90)$$

4.7. Concluding remarks

In the present chapter we have shown how entropy of non random functions provides a new family of entropies for random functions! In this approach the entropy so obtained can be thought of as mean values of entropies defined on random trajectories or paths, and in this way it is basically different from the entropy of dynamical systems considered as Markovian processes (see for instance [4.1,4.3]. Loosely speaking, in the first model we considered the entropy of the mean value of the stochastic trajectory, and in the second we referred to the mean value of the entropy of each trajectory considered as the graph of a deterministic function. We have exactly the same difference of meaning between convergence in probability and almost sure convergence. In the first case one considers the convergence of the variable at each instant, and in the second one considers the convergence of the trajectory as the whole. And this is the reason why we have carefully avoided to use the term stochastic processes of which the entropy is well defined in the literature. Instead, we referred to entropy of random functions.

And then using these results regarding random functions, we have obtained the entropy of distributed functions.

References

4.1 JUMARIE, G.; *Relative Information, Theories and Applications,* Springer Verlag, Berlin, 1990
4.2 JUMARIE, G.; A family of geometrical entropies for stochastic processes in path space, *Physica A*, Vol 184, pp 499-522, 1992
4.3 JUMARIE, G.; Extension of Gatlin's informational divergence to Markovian stochastic processes, *J. Biological Physics*, Vol 23, No 2, pp 101-110, 1997

Chapter 5

Quantum Entropies of Non-Probabilistic Square Matrices

We are seeking different causes.
Nietzsche (Zarathustra(III,2))

5.1. Introduction

In physics we deal with systems of which the dimensions may be macroscopic, mesoscopic, or microscopic. It follows that if we want to analyse them from the information-theoretic standpoint we shall need an appropriate framework which takes account of this level of definition. But whilst Shannon's theory is quite suitable in the first case and can be used in the second, it cannot be applied to microscopic analysis. Other approaches are necessary, and one of them is quantum information.

The same dichotomy will hold for information of non-random data, so that it seems of interest to look for an extension of quantum information theory which would define the amount of uncertainty involved in square matrices which do not necessarily describe phenomena of random nature, and which are here referred to as non-probabilistic matrices.

In the present chapter we shall see that this goal is, to a large extent, achievable; and more specifically, we shall obtain some meaningful concepts of entropy for non-probabilistic operators, that is to say operators which are defined irrespective of any probabilistic framework.

The reader need not be familiar with the basic notions of quantum mechanics, since we shall expand our approach without explicitly referring to the latter. Nevertheless, for the sake of completeness we shall begin with a reminder of von Neumann's quantum entropy.

This being so, one of our purposes is to arrive at a unified approach to Shannon information and to quantum information via the maximum entropy principle, and to this end, firstly, we shall show how one can obtain the explicit expression of Shannon entropy of random variables by using this principle. Then we shall get various definitions for the entropy of non-probabilistic matrices, therefore we shall expand the theory.

G. Jumarie, *Maximum Entropy, Information Without Probability and Complex Fractals*
© Springer Science+Business Media Dordrecht 2000

5.2. Background of quantum mechanical entropy

5.2.1. Basic principles of quantum mechanics

Loosely speaking, the basic principles of quantum mechanics can be summarized as follows.

Principle of quantification. To any observable physical variable $\underline{A}$, is associated a linear Hermitian operator A. When we measure the value of this variable in a real observation process, all we can obtain is one of the (real) eigenvalues of A.■

Principle of spectral decomposition. Assume that the observable is associated with the operator A of which the eigenvalues and the eigenfunctions are α_i and ϕ_i respectively, $i = 1,2,3,...,$ for a discrete spectrum. Assume, moreover, that the physical system under consideration is described by a (wave) function $\psi(q,t)$. Then one can expand the latter in the form

$$\psi(q,t) = \sum_i c_i(t)\phi_i(q),$$

and the probability that a measurement ascribes the value α_i to $\underline{A}$ is $|c_i|^2 = c_i^* c_i$, where the star denotes the complex conjugate.

In addition, $|\psi(q,t)|^2$ is a probability density, in other words one has the equality

$$\langle \psi | \psi \rangle = 1, \tag{5.1}$$

where the symbol $\langle (.)|(.) \rangle$ holds for the inner product. The mean value of the operator A is defined by the expression

$$E\{A\} := \langle \psi | A\psi \rangle . \tag{5.2}$$

Assume (to simplify) that A has no multiple eigenvalues, then one has the equality

$$\sum_{i=1}^{\infty} |c_i|^2 = \langle \psi | \psi \rangle = 1, \tag{5.3}$$

which is meaningful in terms of probability.

5.2.2. Matrix representation of operators

Given a linear operator A and a complete orthonormal system of functions $\{u_n\}$, A can be described by the matrix (A_{ij}) where A_{ij} is defined by the expression

$$A_{ij} = \langle u_i | A u_j \rangle \ . \tag{5.4}$$

Indeed, assume that A is applied to a state function $\psi(q,t)$ in such a manner that we have to calculate $A\psi(q,t)$. To this end we shall expand $\psi(q,t)$ in the form

$$\psi(q,t) = \sum_j c_j(t) u_j(q) \ ,$$

therefore we have

$$\begin{aligned} A\psi(q,t) &= \sum_j c_j(t) A u_j(q), \\ &= \sum_{i,j} c_j(t) A_{ij} u_i \ . \end{aligned}$$

The matrix (A_{ij}) is referred to as the matrix representation of the operator A.

5.2.3. Density matrix

With the above notations, the mean value of A at the instant t is (see equation (5.2))

$$E\{A\} = \sum_{i,j} c_i^*(t) c_j(t) A_{ij}, \tag{5.5}$$

such that the relation

$$\langle u_j | \psi \rangle \langle \psi | u_i \rangle = c_i^*(t) c_j(t)$$

suggests to introduce the operator $\rho(t)$ defined by the equation

$$\rho(t) := |\psi\rangle\langle\psi|, \tag{5.6}$$

and of which the matrix elements are

$$\begin{aligned} \rho_{ji}(t) &= c_i^*(t)c_j(t) \\ &= \langle u_j|\rho(t)u_i\rangle \quad . \end{aligned} \tag{5.7}$$

DEFINITION 5.1. The operator $\rho(t)$ defined by the equation (5.6) is referred to as density operator, and the matrix (ρ_{ij}), in equation (5,7), is a density matrix.■

With this formulation, mathematical expectation can be expressed as the trace of a matrix. Indeed, using equations (5.4), (5.5) and (5.7) one can write,

$$\begin{aligned} E\{A\} &= \sum_{i,j}\langle u_j|\rho u_i\rangle\langle u_i|Au_j\rangle \\ &= tr\big(\rho(t)A\big), \end{aligned} \tag{5.8}$$

where *tr(.)* stands for *trace of.*.

DEFINITION 5.2. Following von Neumann [5.7,5.8], the amount of uncertainty defined by the density operator ρ, or the density matrix (ρ_{ij}), is measured by the quatum entropy

$$S(\rho) := -tr(\rho \ln \rho)\,. \blacksquare \tag{5.9}$$

This expression uses the notation of matrix functions, and is to be understood as follows. Let $\{\lambda_i\}$ denote the set of eigenvalues of ρ (they are positive) and assume that there is no multiple eigenvalue; then we have the quantum entropy

$$S(\rho) = -\sum_i \lambda_i \ln \lambda_i \cdot \tag{5.10}$$

When, for instance, there is one eigenvalue of order k, then the corresponding term in (5.10) is repeated k times (to be consistent with the Shannon entropy).

In the following we shall generalize this concept to operators which are not Hermitian and/or have no probabilistic meaning; but beforehand, we shall show how one can obtain the expression of the entropy of Shannon entropy of random variables by using the *maximum conditional entropy principle.*

5.3. A maximum entropy approach to entropy of random variables

5.3.1 Preliminary remarks

In order to obtain the explicit expression of the Shannon entropy of random variables, it is customary to select a set of prior desiderata which a measure of uncertainty should likely satisfy, and in this approach, the axioms which are by-now largely taken for granted, are those of Shannon himself [5.10] refined later by Fadeev [5.1], see the second chapter.

In a unified approach to information and quantum information we should follow the same or a similar way of arriving at the corresponding definition of quantum entropy; but unfortunately this (mathematical) functional scheme will not be useful in a non-probabilistic framework. This is the reason why we suggets the following variational derivation of Shannon entropy, which we shall be able to duplicate easily for quantum entropy of non-probabilistic matrices.

5.3.2. A new set of axioms for entropy of random variables

Another set of possible meaningful desiderata which should be satisfied by a measure of the uncertainty involved in a discrete probability distribution, is the following one.

(A1) As a direct consequence of Hartley's measure of uncertainty [5,2] (see Chapter 2), we shall assume that the amount of uncertainty contributed by the outcome x_i of the random variable X is $-\ln p_i$, and we shall measure the amount of uncertainty which charaterizes X by the weighted combination

$$U(X|\beta) \;:=\; \sum_{i=1}^{m} -\beta_i \ln p_i \;, \tag{5.11}$$

where $\beta := (\beta_1, \beta_2, ..., \beta_m)$, $\beta_i \rhd 0$, $i = 1,2,...,m$, is a vector parameter such that

$$\sum_{i=1}^{m} \beta_i \;=\; 1 \;. \tag{5.12}$$

(A2) Assume that β satisfies some mathematical constraints; then among all the admissible weighted combinations defined by the equations (5.11) and (5.12) and these

constraints we shall select the largest: the entropy $U(X)$ will be defined by the equality

$$\begin{aligned} U(X) &:= \max_{\beta} U\left(X|\beta\right) \\ &:= U\left(X|\beta^*\right). \end{aligned} \tag{5.13}$$

(A3) Assume that the final expression of $U(X)$ can be written in the form

$$U(X) = -\sum_{i=1}^{m} p_i f(p_i) \ln p_i \ , \tag{5.14}$$

where $f(.)$ is an unknown positive differentiable function to be determined; then applying the axioms (A1) and (A2) with the additional condition

$$-\sum_{i=1}^{m} \beta_i f(\beta_i) \ln \beta_i = h \ , \tag{5.15}$$

where h is a given constant, should yield the solution

$$\beta_i^* = p_i f(p_i), \tag{5.16}$$

at least for some special values of h.

On the practical meaning of these axioms. These axioms are supported by arguments of a physical nature which can be detailed as follows.

Axiom (A1) says that if we can guess a possible distribution of uncertainties which characterizes the outcomes x_i, $i = 1,2,...,$m; then the uncertainty associated with X is the mean value of this distribution, in some sense to be defined by a set of weighting coefficients. According to Hartley, the entropy defined by the uniform distribution $p_1 = p_2 = ... = p_m = 1/m$ is $\ln m = -\ln p_i$, and it is this remark which suggests the equation (5.11). In addition, the condition $\beta_i \triangleright 0$ emphasizes that all the components p_i should contribute to the uncertainty involved in X.

Axiom (A2) is exactly the maximum conditional entropy principle which we used to derive the entropy of non-random functions. Here, of course, we cannot consider it as a consequence of Shannon information theory, but rather we think of it as an assumption, which moreover is quite consistent with Jaynes' maximum entropy principle [5.4].

Axioms (A3) is a self-reference principle which can be understood as follows. Assume that we randomize the problem, and that we consider β as being a probability distribution; then the uncertainty involved in β should be

$$U(\beta) \;=\; -\sum_{i=1}^{m} \beta_i f(\beta_i) \ln \beta_i \;. \tag{5.17}$$

But in our embedding approach this assumption looks like quite natural, because, by this way, we directly refer to the informational framework to assume that the amount of uncertainty we have about β has a given fixed value h. The result thus obtained will depend upon h, and in varying the value of the latter, we shall therefore obtain a family of entropic expressions.

5.3.3. Shannon entropy via maximum entropy

We can now state the following result:

PROPOSITION 5.1. *Shannon entropy of random variables can be obtained as a consequence of the axioms (A1) to (A3) above.* ■

Proof. Firstly, we shall show that these axioms provide a family of generalized Shannon entropy of order c in the form

$$H_c(X) \;=\; -\frac{\sum_{i=1}^{m} p_i^c \ln p_i}{\sum_{i=1}^{m} p_i^c} \;. \tag{5.18}$$

To this end we shall proceed as follows.

(i) We define the Lagrangian

$$L \;:=\; \sum_{i=1}^{m} \left[-\beta_i \ln p_i + \lambda \beta_i f(\beta_i) \ln \beta_i + \mu \beta_i \right],$$

where λ and μ denote the multipliers, and its maximization provides the equation (we equate to zero the derivative w.r.t. β_i)

$$-\ln p_i + \lambda \left[f(\beta_i) + \beta_i f'(\beta_i) \right] \ln \beta_i + \lambda f(\beta_i) + \mu \;=\; 0 \;. \tag{5.19}$$

(ii) According to the axiom (A3), the solution of this equation (5.19) should be $\beta_i = p_i f(p_i)$, and inserting into (5.19)) we obtain the functional equation

$$[\lambda f(p_i f) - 1]\ln\ p_i + \lambda f(p_i f)\ln\ f + p_i f f'(p_i f)\ln\,(p_i f) + \mu\ =\ 0.$$

(iii) This equation suggests testing whether the equality

$$\begin{aligned} f(\beta_i)\ &=\ k \\ &=\ constant\ , \end{aligned}$$

would not be a possible solution; and making the substitution into (5.19) yields

$$\beta_i\ =\ p_i^{(1/k\lambda)-1-(\mu/k\lambda)}, \tag{5.20}$$

therefore the expression of $H_c(X)$ via the normalizing condition (5.12).

This completes the derivation of equation (5.18).

(iv) If we now restrict ourselves to the expression (5.14) in which $f(p_i)$ depends upon p_i only, then we have to make $c = 1$ into (5.18), therefore we obtain the expression of Shannon entropy of random variables.

Remarks and comments. Obviously, from a practical standpoint this result does not contribute new equations to the theory, since the generalized Shannon entropy of order c can be obtained via entropy of non random functions (see Chapter 3).

Nevertheless, from the theoretical standpoint this new variational approach provides additional support to the validity of the maximum entropy principle which, in this way, can be taken as the basic axiom of information theory. Indeed, by using this principle we shall be able to obtain the Shannon entropy of random variables, entropies of non-random functions, quantum entropies of square matrices, and von Neumann quantum entropy. In other words, we shall thus arrive at a unified approach to these different measures of uncertainty.

5.4. Quantum entropies of non-probabilistic square matrices

5.4.1. von Neumann quantum entropy of square matrices

PROPOSITION 5.2. *Let* A *denote a constant real valued n×n square matrix with the eigenvalues* $\lambda_1,\lambda_2,...,\lambda_n$; *then a measure of the amount of uncertainty it defines, which is fully consistent with von Neumann quantum entropy on the one hand, and with*

Shannon entropy of random variables on the other hand, is the quantum entropy of order c [5.5]

$$QH_c(A) \;:=\; \frac{\sum_{i=1}^{n} |\lambda_i|^c \ln|\lambda_i|}{\sum_{i=1}^{n} |\lambda_i|^c} \quad , \quad c \in \Re \cdot \tag{5.21}$$

When there are multiple eigenvalues, the corresponding terms in (5.21) are then duplicated as many times as indicated by their respective orders. ■

Proof. (i) Preliminary remark. We first bear in mind that the entropy of a random variable X and the entropy of its (cumulative) distribution function vary in opposite ways. One of them increases when the other decreases; one of them is maximum when the other one achieves its minimum value.

We should have the same property for quantum entropy of density matrices (in quantum mechanics) as compared with the quantum entropy of non probabilistic matrices.

(ii) This being so, let $H_1, H_2, \ldots, H_n$ denote the Shannon entropies of n independent random events $E_1, E_2, \ldots, E_n$; then the Shannon entropy of the n-tuple $(E_1, E_2, \ldots, E_n)$ is

$$H(E_1, E_2, \ldots, E_n) \;=\; \sum_{i=1}^{n} H_n \cdot \tag{5.22}$$

(iii) Let us restrict ourselves to the special case when all these entropies are positive, which amounts to suppose that we are dealing with discrete events; then we can introduce the number p_i defined by the equations

$$-\ln p_i \;=\; H_i, \quad i = 1, 2, \ldots, n \;, \tag{5.23}$$

and inserting into (5.22) yields

$$H(E_1, E_2, \ldots, E_n) \;=\; -\sum_{i=1}^{n} \ln p_i \cdot \tag{5.24}$$

Notice that p_i as so defined is not necessarily a probability, since then, generally one will have $\sum_{i=1}^{n} p_i \neq 1$.

(iv) When we write the equation (5.24), we implicitly assume that at the instant of the observation we simultaneously observe the issues of all the events E_i's.

This being so, consider the case where these events E_i's occur serially in time; for instance E_1 occurs at the instant *1*, E_2 occurs at the instant *2*, and so on. Then, according to the combination law of Shannon entropy of deterministic functions, we shall have to consider the entropy

$$H_c\left(E_1 \cup E_2 U \ldots \cup E_n\right) = -\frac{\sum_{i=1}^{n} e^{c(-\ln p_i)} \ln p_i}{\sum_{i=1}^{n} e^{c(-\ln p_i)}}. \tag{5.25}$$

In the special case when $\sum_i p_i = 1$, one find the meaningful equality

$$H_{-1}\left(E_1 \cup E_2 \cup \ldots \cup E_n\right) = H(X).$$

(v) We now consider the $n \times n$ square matrix A. Let $f(x)$ denote a continuously differentiable n-vector function; then as a result of the relation

$$H(Af(.);\Omega) = H(f(.);\Omega) + \ln|A|, \tag{5.26}$$

where $|A|$ denotes the absolute value of the determinant of A, $|A| = |\det A|$, one can think of the Shannon entropy of the constant matrix A as

$$H(A) := \ln|A|. \tag{5.27}$$

Assume that A has n different eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$; then one can write as well

$$H(A) = \sum_{i=1}^{n} \ln|\lambda_i|. \tag{5.28}$$

(vi) We now refer to the parlance of quantum mechanics, and we assume that each λ_i is associated with the state Σ_i of a physical system Σ. Then we can meaningfully make the formal identification

$$\begin{aligned} H(\Sigma) &= H(\Sigma_1, \Sigma_2, ..., \Sigma_n) \\ &= \sum_{i=1}^{n} H(\Sigma_i) \\ &= -\sum_{i=1}^{n} \ln|\lambda_i| \cdot \end{aligned} \tag{5.29}$$

(vii) By using (5.25), we arrive at the expression

$$H_{-1}(\Sigma_1 \cup \Sigma_2 \cup ... \cup \Sigma_n) = -\frac{\sum_{i=1}^{n} |\lambda_i| \ln|\lambda_i|}{\sum_{i=1}^{n} |\lambda_i|}, \tag{5.30}$$

which is identical to von Neumann quantum entropy of density matrices when $\lambda_i > 0$ for all i, and $\Sigma_i \lambda_i = 1$.

(viii) Next, according to remark (i) above, we reverse the sign of (5.30), and we apply the formalism of entropy of non-random functions to obtain (5.21).

(ix) The requirement regarding the meaning of equation (5.21) in the case where there are multiple eigenvalues is stated in order to have consistency with the entropy of non-random functions. More explicitly, it is easy to see that the equation (5.28) is still satisfied, in such a manner that one can duplicate the argument step by step, therefore we once more obtain the expression (5.21).

5.4.2. Renyi quantum entropy of square matrices

PROPOSITION 5.3. *Consider the constant square matrix A of proposition 5.2. A measure of the amount of uncertainty it defines, which is fully consistent with von Neumann quantum entropy on the one hand, and Renyi entropy of random variables on the other, is the Renyi quantum entropy of order s defined by the expression*

$$QH_{R,s}(A) := -\frac{1}{1-s}\ln\frac{\sum_{i=1}^{n}|\lambda_i|^s}{\sum_{i=1}^{n}|\lambda_i|} \quad , \quad s \rhd 0 \quad , \quad s \neq 1 \tag{5.31}$$

where each eigenvalue is repeated a number of times equal to its order. ■

The proof of this result is a direct duplication of the proof of the proposition 5.2, but in the framework of Renyi entropy.

5.4.3. Structural quantum entropy of square matrices

PROPOSITION 5.4. *Consider the constant square matrix A of proposition 5.2. A measure of the amount of uncertainty it defines, which is fully consistent with von Neumann quantum entropy on the one hand, and structural entropy of random variables on the other, is the structural quantum entropy of order s defined by the expression*

$$QH_{C,s}(A) := \frac{1}{e^{1-s}-1}\left[\frac{\sum_{i=1}^{n}|\lambda_i|}{\sum_{i=1}^{n}|\lambda_i|^s}-1\right] \quad , \quad s \rhd 0 \quad , \quad s \neq 1 \text{.} \blacksquare \tag{5.32}$$

Here again, the proof is a straightforward duplication of the proof of proposition 5.2, but in the framework of structural entropy.

5.4.4. Application of matrix functions

These entropies can be written in a more compact form in terms of matrix functions, but before doing so, we shall digress a bit on this topic (for the reader who is not familiar with it)

Background on matrix functions

Assume that the function *f(x)* of the scalar real valued variable *x* can be expanded in the series

$$f(x) = \sum_{i=0}^{\infty} a_i x^i \quad , \quad x \in \Re, \tag{5.33}$$

then we shall define *f(A)* by the expression

$$f(A) := \sum_{i=0}^{\infty} a_i A^i. \tag{5.34}$$

When A has no multiple eigenvalue one can rewrite (5.34) in the form

$$f(A) = \sum_{i=1}^{n} \frac{\prod_{j\neq i}(A-\lambda_j I_n)}{\prod_{j\neq i}(\lambda_i-\lambda_j)} f(\lambda_i), \tag{5.35}$$

which can be obtained also in computing the Vandermonde determinant

$$\begin{vmatrix} f(A) & f(\lambda_1) & f(\lambda_2) \ldots\ldots\ldots f(\lambda_n) \\ A^{n-1} & \lambda_1^{n-1} & \lambda_2^{n-1} \ldots\ldots\ldots \lambda_n^{n-1} \\ A^{n-2} & \lambda_1^{n-2} & \lambda_2^{n-2} \ldots\ldots\ldots \lambda_n^{n-2} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ A & \lambda_1 & \lambda_2 \ldots\ldots\ldots\ldots \lambda_n \\ I_n & 1 & 1 \ldots\ldots\ldots\ldots 1 \end{vmatrix} = 0, \tag{5.36}$$

where I_n denotes the $n \times n$ unit matrix.

Obviously (5.35) and (5.36) do not apply when there are multiple eigenvalues; but by using limit techniques one can obtain the following result.

Assume that λ is of order p. Instead of having p identical columns in (5.36), one has the normal column followed by columns which are the derivatives of the first one with respect to λ until the order $(p\text{-}1)$. For instance, when $p = 3$, one has the following Vandermonde matrix component

$$\begin{vmatrix} f(\lambda) & f'(\lambda) & f''(\lambda) \\ \lambda^{n-1} & (n-1)\lambda^{n-2} & (n-1)(n-2)\lambda^{n-3} \\ \lambda^{n-2} & (n-2)\lambda^{n-3} & (n-2)(n-3)\lambda^{n-4} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \lambda^3 & 3\lambda^2 & 6\lambda \\ \lambda^2 & 2\lambda & 2 \\ \lambda & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix} \tag{5.37}$$

In the special case where all the n eigenvalues are identical one finds:

$$f(A) = I_n f(\lambda) + (A - \lambda I_n) f'(\lambda) + \frac{1}{2}(A - \lambda I_n)^2 f''(\lambda) + .. \\ .. + \frac{1}{(n-1)!}(A - \lambda I_n)^{n-1} f^{(n-1)}(\lambda) \quad . \tag{5.38}$$

Another possible definition of $f(A)$ is the complex integral

$$f(A) = \frac{1}{2\pi i} \oint \frac{f(z)}{(zI_n - A)} dz, \tag{5.39}$$

in which the closed contour contains all the eigenvalues of A.

Application to quantum entropy.

Let $m(x)$ denote the function

$$m(x) := |x| \quad , \quad x \in \Re, \tag{5.40}$$

then, using the notation of matrix functions, one has the equalities

$$QH_c(A) = \frac{tr\left[m^c(A) \ln\ m(A)\right]}{tr\ m^c(A)}, \tag{5.41}$$

$$QH_{R,s}(A) = \frac{1}{1-s} \ln \frac{tr\,(m^s(A))}{tr\,(m(A))}, \tag{5.42}$$

$$QH_{C,s}(A) = \frac{1}{e^{1-s} - 1}\left[\frac{tr(m(A))}{tr(m^s(A))} - 1\right] \tag{5.43}$$

5.5. Some properties of quantum entropies of square matrices

In the present section, we give some properties of quantum entropies which can be proved very easily, as they refer to well known results on matrices.

(i) Let $|\lambda_\mu|$ and $|\lambda_M|$ denote respectively the minimum and the maximum value of $|\lambda_i|$ in the equation (5.21); then one has

$$\lim QH_c(A) = \ln |\lambda_M| \quad as \quad c \uparrow \infty, \tag{5.44}$$

$$= \ln|\lambda_m| \quad as \quad c \downarrow -\infty. \tag{5.45}$$

This result can be obtained as follows. First, for large values of c, one has

$$\sum_{i=1}^{n}|\lambda_i|^c = |\lambda_M|^c\left(1+\sum_{i\neq M}\left|\frac{\lambda_i}{\lambda_M}\right|^c\right)$$
$$\cong |\lambda_M|^c ,$$

and similarly, one can write

$$\sum_{i=1}^{n}|\lambda_i|^c \ln|\lambda_i| = |\lambda_M|^c \ln|\lambda_M|\left(1+\sum_{i\neq M}\frac{|\lambda_i|^c \ln|\lambda_i|}{|\lambda_M|^c \ln|\lambda_M|}\right),$$

therefore the equation (5.44)

A similar argument will provide the equation (5.45).

(ii) $QH_c(A)$ is an increasing function of c.

Indeed, the derivative of $QH_c(A)$ with respect to c is given by the expression

$$\frac{dQH_c(A)}{dc} = \frac{\sum_{i=1}^{n}|\lambda_i|^c \ln|\lambda_i|}{\sum_{i=1}^{n}|\lambda_i|^c} - [QH_c(A)]^2 , \tag{5.46}$$

which is positive by virtue of Schwarz' inequality.

(iii) Let A and B denote two $n\times n$ matrices such that $AB = BA$, and with the respective eigenvalues λ_i and μ_i, then one has the equality

$$QH_c(AB) = \frac{\sum_{i=1}^{n}|\lambda_i\mu_i|^c \ln|\lambda_i\mu_i|}{\sum_{i=1}^{n}|\lambda_i\mu_i|^c} , \tag{5.47}$$

$$= \frac{tr\left[m^c(AB)\ln(m(AB))\right]}{tr\,(m^c(AB))} . \tag{5.48}$$

This result is a consequence of the property that if $AB = BA$, then A and B have the same eigenvectors..

(iv) Let A and B denote a $n \times n$ and a $m \times m$ matrix respectively, and let $A \otimes B$ denote their Kroneker (or tensor) product; then one has the following relations, which are

$$QH_c(A \otimes B) = QH_c(A) + QH_c(B), \tag{5.49}$$
$$QH_{R,s}(A \otimes B) = QH_{R,s}(A) + QH_{R,s}(B) . \tag{5.50}$$

Indeed, in such a case, the $n \times m$ eigenvalues of $A \otimes B$ are $\lambda_i \mu_j$.

(v) Let a denote a scalar real valued constant, then one has

$$QH_c(aA) = QH_c(A) + \ln|a|, \tag{5.51}$$
$$QH_{R,s}(aA) = QH_{R,s}(A) + \ln|a| . \tag{5.52}$$

(vi) Assume that A is a density matrix ρ (of quantum mechanics), then the following equality holds:

$$QH_1(\rho) = -S(\rho).$$

This equation is exactly the parallel of the similar property satisfied by the entropy of a random variable X and the entropy of its (cumulative) distribution function $F(x)$, clearly $H(F(.);\Re) = -H(X)$.

(vii) Let A and B denote two $n \times n$ matrices, then one has the relations

$$QH_c(AB) = QH_c(BA), \tag{5.53}$$
$$QH_{R,s}(AB) = QH_{R,s}(BA), \tag{5.54}$$
$$QH_{C,s}(AB) = QH_{C,s}(BA) , \tag{5.55}$$

which are a result of the property that AB and BA have the same eigenvalues.

(viii) Define $A' := U^{-1}AU$, where U is an invertible matrix; then one has the equalities

$$QH_c(A') = QH_c(A), \tag{5.56}$$
$$QH_{R,s}(A') = QH_{R,s}(A), \tag{5.57}$$
$$QH_{C,s}(A') = QH_{C,s}(A). \tag{5.58}$$

(ix) When P is a positive definite matrix, one can write

$$QH_c(P) = \frac{tr(P^c \ln P)}{tr(P^c)}, \tag{5.59}$$

$$QH_{R,s}(P) = -\frac{1}{1-s}\frac{tr(P^s)}{tr(P)}, \tag{5.60}$$

$$QH_{C,s}(P) = \frac{1}{e^{1-s}-1}\left[\frac{tr(P)}{tr(P^s)}-1\right]. \tag{5.61}$$

(x) The usual limiting relations between Shannon entropy, Renyi entropy and structural entropy apply here to yield

$$\lim \; QH_{R,s}(A) = QH_1(A) \quad as \quad s \to 1, \tag{5.62}$$

$$\lim \; QH_{C,s}(A) = QH_1(A) \quad as \quad s \to 1. \tag{5.63}$$

(xi) Let A, B and C denote three $n \times n$ matrices, then their products ABC, CAB and BCA (and not BAC.; i.e. cyclic permutation only are allowed) are such that

$$QH_c(ABC) = QH_c(CAB) = QH_c(BCA), \tag{5.64}$$

$$QH_{R,s}(ABC) = QH_{R,s}(CAB) = QH_{R,s}(BCA), \tag{5.65}$$

$$QH_{C,s}(ABC) = QH_{C,s}(CAB) = QH_{C,s}(BCA). \tag{5.66}$$

This property is a consequence of the fact that these products have the same eigenvalues.

(xii) Remark that all the above relations, which involve the Renyi entropy, have their counterparts in terms of structural quantum entropy, and so by virtue of the obvious equation

$$QH_{R,s}(A) = \frac{1}{1-s}\ln\left[\left(e^{1-s}-1\right)QH_{C,s}(A)+1\right]. \tag{5.67}$$

5.6. Quantum entropy and complexity

5.6.1. On the practical meaning of quantum entropy

In the present section we shall make some remarks and comments which could be of help in clarifying the significance of the various measures of uncertainty above defined.

Uncertainty and eigenvalues.

In order to simplify the argument we shall assume that the matrix A under consideration has no multiple eigenvalue. In such a case one can expand it in terms of its basis matrices A_i, $i = 1,2,...,n$, in the form

$$A = \lambda_1 A_1 + \lambda_2 A_2 + ... + \lambda_n A_n \ , \tag{5.68}$$

where the $n \times n$ matrices A_i satisfy the conditions

$$A_i^n = A_i \ , \quad i = 1,2,..,n, \tag{5.69}$$

$$A_i A_j = 0 \ , \quad i \neq j \ , \tag{5.70}$$

$$|A_i| = 0 \ , \quad i = 1,2,..,n, \tag{5.71}$$

$$A_1 + A_2 + ... + A_n = I_n, \tag{5.72}$$

with I_n denoting the unity matrix of order n.

The analogy with the usual vector decomposition

$$v = X_1 v_1 + X_2 v_2 + ... + X_n v_n \ , \quad X_i \in \Re$$

is quite suggestive. If we assume that the bases A_i are fixed, we are then considering the variations of A with respect to its eigenvalues only, and we are led to introduce an entropy $H(\lambda_1,...,\lambda_n)$. And this point of view is not surprizing at all, since in the information-theoretic framework, the eigenvalues are the only parameters which are significant.

Another way to support this remark is the following one. Assume that we identify A with the n-tuple $(...,\lambda_i A_i,...)$,

$$A \equiv (\lambda_1 A_i,...,\lambda_n A_n),$$

then it is quite natural to consider the entropy of the matrix $\lambda_i A_i$ in the sense of equation (5.27), that is

$$\begin{aligned} H(\lambda_i A_i) &= \ln|\lambda_i A_i| \\ &= \ln|\lambda_i| + \ln|A_i| \end{aligned} \tag{5.73}$$

But according to equation (5.71) one has that $\ln|A_i| = -\infty$, in other words A_i does not contribute any amount of uncertainty to the observation of $\lambda_i A_i X$; and the only relevant contribution is owed to the gain λ_i, which amounts to considering the observable $\lambda_i X$.

This remark is of interest, because it could provide a way to generalize the present concept of quantum entropy whenever we have at hand a formal expansion similar to (5.68).

Quantum entropy and structure

The Shannon entropy of the random variable X is $H(X)$. If we now consider the product xY where x is a constant and Y is a scalar random variable, one has the equality

$$H(xY) = \ln|x| + H(Y), \tag{5.74}$$

in such a manner that the density of entropy of the multiplying coefficient $X\times(.)$ which takes on the value $x\times(.)$ is $ln|x|$, therefore the entropy

$$H(X\times(.)) := \int_{\Re} p(x)\ln|x|dx. \tag{5.75}$$

Notice the presence of the dot in the notation, which emphasizes that in the present case X is not considered as a random variable, but as a random multiplicative coefficient. And, indeed, we must carefully distinguish between the entropy of the random variable X and the entropy of the operator $(X\times(.))$.

The same remark applies to the function $f(x)$. On the one hand, we have the entropy of the random variable $f(X)$, and on the other, we have the entropy of the function $f(.)$ which is $H_c(f(.);\Omega)$.

We now consider the $n\times n$ matrix $A := (a_{ij})$. The entropy of A considered as a random vector is defined by the joint probability density $p(a_{11},...,a_{ij},...,a_{nn})$; and its entropy when it is considered as a multiplying coefficient is

$$H(A\times(.)) = \ln|A|. \tag{5.76}$$

In contrast, $QH_c(A)$ should rather be thought of as a measure of uncertainty associated with the internal structure of A; briefly, $H(A\times(.))$ is a macroscopic entropy, whilst $QH_c(A)$ is a microscopic entity. In this way $QH_c(A)$ appears to be more interesting than $H(A\times(.))$ for characterizing the internal structure of A.

5.6.2. On the definition of complexity

The quantum entropy $QH_c(A)$ of non-probabilistic matrices has been obtained as an extension of von Neumann quantum mechanical entropy, and at the present time there is a certain consensus amongst physicists to think that the latter could be considered as a measure of (structural) complexity of quantum systems.

But what do we mean when we refer to complexity? In fact, this term has been utsed in so many different frameworks that it is probable that it refers to a set of properties rather than to one given feature only. For our purpose we shall have in mind the following simple and straightforward statements, of which the generality is quite suitable and sufficient to support further analysis.

(i) The amount of information which is required to carefully describe the structure of a given system, increases with the complexity of the latter.

(ii) The number of parameters (or variables) which we need in order carefully to describe the structure of a given system increases with the complexity of the latter.

5.6.3. Some illustrative examples

Complexity of linear systems.

The solution of the n-dimensional linear differential equation

$$\dot{x}(t) = Ax(t) + u(t) \quad , \quad x(0) = x_0, \tag{5.77}$$

where A is a constant square matrix, is

$$x(t) = e^{At}x_o + e^{At}\int_0^t e^{-A\tau}u(\tau)d\tau \ , \tag{5.78}$$

in such a manner that we are led to consider the quantum entropy $QH_c\{\exp(At)\}$. In other words, analogously with quantum mechanics, we could measure the complexity of this system by $QH(e^A)$, in which c is equal to the unity.

For instance, according to this definition the complexity of the system at the instant t would be $QH(e^{At})$, a complexity which increases when the system is unstable. And this property is quite consistent with our common sense notion of complexity.

Complexity of Markov chains.

Let us consider the two-state Markov chain defined by the transition matrix

$$P := \begin{pmatrix} p & q \\ q & p \end{pmatrix} \quad , \quad p+q=1,$$

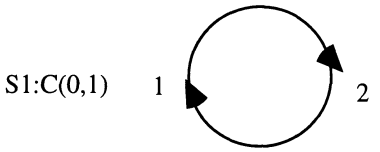
and assume that we measure its structural complexity by means of $QH(e^P)$, that is to say

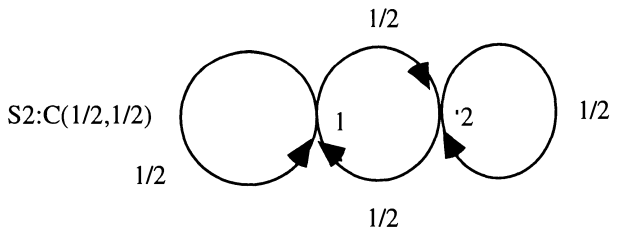
$$QH(e^P) = \frac{e^1 \ln 1 + e^{p-q}(p-q)}{e^1 + e^{p-q}}. \qquad (5.79)$$
$$=: C(p,q).$$

A simple calculation provides the complexities

$$C(0,1) = -0,1192 \; ; \; C(\frac{1}{2},\frac{1}{2}) = 0 \; ; \; C(1,0) = \frac{1}{2},$$

which are respectively associated with the transition graphs$S1$, $S2$ and $S3$ between the points 1 and 2, as pictured below.





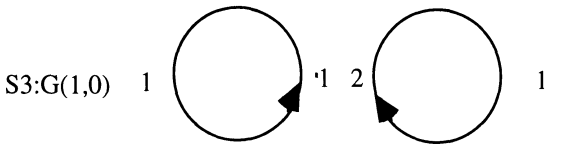


FIGURE 5.1.

We can comment on this result as follows. S2 is of zero complexity as a result of its uniform structure: clearly, uniformity does not generate complexity. The inequality $C(0,1) < \{C(1/2,1/2), C(1,0)\}$ could be thought of as a picture of the fact that $S1$ is the simpler graph that one can imagine to relate the states 1 and 2. In a like manner, $S3$ could be considered as being the most unexpected graph among the three ones.

Complexities of matrices.

In order to measure the (structural) complexity of a square matrix A, it is suggested to use its quantum entropy *QH(A).*

As an illustrative example, assume that A is a $n \times n$ matrix with real eigenvalues λ_i only, and such that $\lambda_1 + \lambda_2 + ... + \lambda_n = constant$. Then in the absence of more information about these eigenvalues, the simplest assumption we can make is that the characteristic equation of A is

$$a(\lambda - b)^n = 0, \tag{5.80}$$

which consistently provides the smaller quantum entropy *QH(A).*

5.7. Quantum entropies of distributed matrices

5.7.1. von Neumann quantum entropy of distributed matrices

PROPOSITION 5.5. *Let* $A(z) = (a_{ij}(z))$ *denote a* $n \times n$ *matrix which is continuous with respect to the distributed parameter* $z \in D \subset \Re^m$; *and let* $\lambda_i(z)$, $i = 1,2,...,n$, *denote its eigenvalues. Then a measure of the amount of uncertainty it involves on the domain D, and which is fully consistent with von Neumann quantum entropy of density matrices, is the quantum entropy of order* $(b,c(z))$, $b \in \Re$, $c(z) \in \Re$, *defined by the expression*

$$QH_{b,c(z)}(A_z; D)$$

$$= \frac{\int_D \exp\left[bQH_{c(z)}(A(z)) + H(X_A; z, c(z))\right] QH_{c(z)}(A(z))dz}{\int_D \exp\left[bQH_{c(z)}(A(z)) + H(X_A; z, c(z))\right]dz}, \tag{5.81}$$

where $H(X_A; z, c(z))$ *is the Shannon entropy of the random variable* X_A *associated with the discrete probability distribution*

$$p_{A,c(z)}(i) \;:=\; \frac{|\lambda_i(z)|^{c(z)}}{\sum_{i=1}^{n}|\lambda_i(z)|^{c(z)}} \tag{5.82}$$

with b and c(z) denoting two parameters which describe how the family of matrices is scanned in the observation process. . ■

Proof. The proof is based on a comparison with the entropy of distributed functions, see section 4.2, and reads as follows.

(i) We first remark that equation (4.27) holds when the entropy $H_{c(z)}(f(.,z);\Omega)$ is the discrete entropy $\sum_i |f_i|^c \ln|f_i| / \sum_i |f_i|^c$.

(ii) This being so, we consider the stairwise function

$$Y_A(x) \;:=\; \sum_{i=1}^{n}|\lambda_i(z)|Y(x-x_i) \quad , \quad x \in \Re \;,$$

where $Y(x\text{-}x_i)$ is the Heaviside step function at x_i.

According to the equation (3.101), one has the identity

$$H_c\big(Y_A(.)\big) \;=\; QH_c(A),$$

in other words, we are in the framework of entropy of distributed functions; we can then apply the equation (4.27), therefore the expression (5.81).

A useful special case.

Assume that $c(z)$ is a constant c in the equation (5.81); then, by using the explicit form of $H(X_A;z,c)$ in terms of $p_{A,c(z)}(i)$, one obtains the expression

$$QH_{b,c}(A_z;D) \;=$$

$$= \frac{\int_D \exp\big[(b-c)QH_c(A(z))\big]\left[\sum_{i=1}^{n}|\lambda_i(z)|^c \ln|\lambda_i(z)|\right]dz}{\int_D \exp\big[(b-c)QH_c(A(z))\big]\left[\sum_{i=1}^{n}|\lambda_i(z)|^c\right]dz} \tag{5.83}$$

$$= \frac{\int_D \exp\left[(b-c)QH_c(A(z))\right] tr\left[m^c(A(z))\ln m(A(z))\right]dz}{\int_D \exp\left[(b-c)QH_c(A(z))\right] tr\left[m^c(A(z))\right]dz}. \tag{5.84}$$

Assume further that one has the equality $b = c$; then equation (5.83) reduces to the simplified form

$$QH_{c,c}(A_z;D) = \frac{\int_D \sum_{i=1}^{n} |\lambda_i(z)|^c \ln|\lambda_i(z)|dz}{\int_D \sum_{i=1}^{n} |\lambda_i(z)|^c dz}, \tag{5.85}$$

$$= \frac{\int_D tr\left[m^c(A(z))\ln m(A(z))\right]dz}{\int_D tr\left[m^c(A(z))\right]dz}. \tag{5.86}$$

5.7.2. Renyi quantum entropy of distributed matrices

PROPOSITION 5.6. *Consider the distributed matrix A(z) of proposition (5.5), equation (5.81); then a measure* $H_{R,s}(A_z;D,v)$, $v \in \Re$, *of the amount of uncertainty it involves on the domain D, which, furthermore, is fully consistent with von Neumann quantum entropy on the one hand and Renyi entropy of random variables on the other , is defined by the expression*

$$H_{R,s}(A_z;D,v) = -\frac{1}{1-s}\ln\frac{\int_D P(z)dz}{\int_D Q(z)dz}, \tag{5.87}$$

with

$$P(z) := \left\{\left[W(A(z))\right]^{1-s} + v\exp\left((1-s)H_{R,s}(X_A;z)\right)\right\}^{\frac{1}{1-s}}\left[W(A(z))\right]^{-s}\sum_{i=1}^{n}|\lambda_i(z)|^s \tag{5.88}$$

and

$$Q(z) := \left\{[W(A(z))]^{1-s} + v\exp\big((1-s)H_{R,s})X_A;z)\big)\right\}^{\frac{1}{1-s}}[W(A(z))]^{1-s}, \tag{5.89}$$

where v *is a real-valued parameter;* $W(A(z))$ *is defined by the expression*

$$W(A(z)) := \sum_{i=1}^{n}|\lambda_i(z)|, \tag{5.90}$$

and $H_{R,s}(X_A,z)$ *is the Renyi entropy associated with the discrete probability distribution (5.82), that is to say*

$$H_{R,s}(X_A;z) := \frac{1}{1-s}\ln\frac{\sum_{i=1}^{n}|\lambda_i(z)|^s}{\left[\sum_{i=1}^{n}|\lambda_i(z)|\right]^s}. \quad \blacksquare \tag{5.91}$$

Proof. The proof is strictly parallel to the proof of Proposition 4.6 (Renyi entropy of distributed functions), but in the framework of quantum entropy.

A useful special case.

If we make $v = 0$ in the equation (5.87), we obtain the counterpart of the expression (4.54), which is

$$H_{R,s}(A_z;D) = -\frac{1}{1-s}\ln\frac{\int_D\left(\sum_{i=1}^{n}|\lambda_i(z)|^s\right)dz}{\int_D\left(\sum_{i=1}^{n}|\lambda_i(z)|\right)dz}. \tag{5.92}$$

$$= -\frac{1}{1-s}\ln\frac{\int_D m^s(A(z))dz}{\int_D m(A(z))dz}. \tag{5.93}$$

5.7.3. *Structural quantum entropy of distributed matrices*

PROPOSITION 5.7. *Consider the distributed matrice A(z) of proposition 5.5, equation (5.81); then a measure* $H_{C,s}(A_z;D,v)$ *of the amount of uncertainty it involves on the domain D, which, furthermore, is fully consistent with von Neumann quantum entropy on the one hand and with structural entropy of random variables on the other , is defined by the expression*

$$H_{C,s}(A_z;D,v) = \frac{1}{e^{1-s}-1}\left[\frac{\int_D Q(z)dz}{\int_D P(z)dz}-1\right], \tag{5.94}$$

where P(z) and Q(z) are given by the equations (5.85) and (5.86). ■

Proof. The proof is parallel to the proof of Proposition 4.7, but in the framework of quantum entropy.

The simplified form of (5.94), which corresponds to $v = 0$, is

$$H_{C,s}(A_z;D,v) = \frac{1}{e^{1-s}-1}\left[\frac{\int_D\left(\sum_{i=1}^{n}|\lambda_i(z)|\right)dz}{\int_D\left(\sum_{i=1}^{n}|\lambda_i(z)|^s\right)dz} - 1\right], \tag{5.95}$$

$$= \frac{1}{e^{1-s}-1}\left[\frac{\int_D tr(m(A(z)))dz}{\int_D tr\left(m^s(A(z))\right)dz} - 1\right]. \tag{5.96}$$

5.7.4. *On the meaning of these entropies*

As is evident, the significance of these expressions can be simply exhibited by combining the randomization approach, which we used to obtain the entropies of non-random functions, with the intrinsic features of the concepts of quantum entropies themselves. The crucial points of the derivation are, firstly, the randomization of the distributed parameter, and, secondly, the identification of the distributed eigenvalues with distributed functions.

Nevertheless, these expressions require that we have at hand the explicit definition of the matrix *A(z)* under consideration, in such a manner that, theoretically speaking, we can obtain its eigenvalues.

This feature gives rise to the following problem. Assume that all the data which are available are the values of the quantum entropy $QH_1(A(z))$, $z \in D$, only, and that we do not have at hand the exact definition of the matrix A(z)) itself. In such a case how can we define the quantum entropy of the family of matrices so indexed by *z*?

It is clear that we cannot explicitly refer to the eigenvalues of $A(z)$ in such a way that the quantum entropies so obtained will rather be global or external, in contrast to the precedings ones which then appear as analytic and internal.

In order to obtain these new expressions, we shall use again the basic approach to entropy via maximization of conditional entropy, and it is the purpose of the following section.

5.8. Entropy of random functions via the maximum entropy principle

5.8.1. General approach

Shannon entropy, Von Neumann quantum entropy and entropy of non-random functions can be obtained in exactly the same way, that is by maximizing some suitable conditional entropies; and this remark is of importance, for at least two reasons.

Firstly, it emphasizes that the definition of entropy of non-random functions and of quantum entropy of non-probabilistic matrices, as we have suggested them, are not definitions made for convenience only. On the contrary, they have the same significance and rationale as Shannon entropy of random variables and von Neumann internal quantum entropy of density matrices.

Secondly, the technique which we applied for obtaining the expressions of these entropies, and which is mainly the maximization of conditional entropies, can be systematically used as a general algorithm to generate new entropic expressions. We shall summarize this approach as follows.

A general scheme to generate entropic expressions.

Let X and Y denote two (discrete or continuous) random variables, and let $U(X)$, $U(Y)$ and $U(X,Y)$ denote their entropies (in the sense of their uncertainty measures) and joint entropy respectively. One has the inequality

$$U(X,Y) \leq U(X)+U(Y), \tag{5.97}$$

where the equality holds when and only when X and Y are independent. We may therefore consider $U(Y)$ as

$$U(Y) = \max_{p(.)} \left(U(X,Y)-U(X)\right), \tag{5.98}$$

where $p(x)$ denotes the probability density of X, subject to the condition

$$U(X) = h,$$

with h denoting a constant.

If we assume that we have at hand some definition for the difference $U(X,Y)$-$U(X)$, but none for $U(Y)$, then the relation (5.97) can be meaningfully used to define $U(Y)$.

This scheme is very general and applies to any measure of uncertainty provided that the relation (5.97) is satisfied. Moreover, very often, it will be possible to replace the difference $U(X,Y) - U(X)$ by the conditional uncertainty $U(Y|X)$, i.e.,

$$U(Y|X) = \int_{\Re^n} p(x)U(Y|x)dx, \tag{5.99}$$

where $p(x)$ is the probability density of X and $U(Y|x)$ is the conditional entropy of Y given that X is equal to x.

5.8.2. Illustrative examples

Example 1. Let $X \in \Re$ denote a continuous random variable with a probability $p(x)$, and B denote a discrete random coefficient which can take on the values b_1 and b_2 with the respective probabilities q_1 and q_2 , $q_1 + q_2 = 1$; and define the new variable $Y := BX$. The probability $r(y)$ of Y is

$$, \; q_1|b_1|^{-1} p(y/b_1) + q_2|b_2|^{-1} p(y/b_2) \tag{5.100}$$

therefore one can calculate the entropy

$$H(Y) = -\int_{\Re} r(y) \ln r(y) dy.$$

Unfortunately, this procedure does not easily provide the explicit expression for $H(Y)$ in such a way that it may be of interest to examine which kind of result one can obtain by using the maximization of conditional entropy.

(i) To this end let us consider the entropy

$$H(BY) = H(B) + H(Y|B)$$

$$= H(B) + q_1 H(Y|b_1) + q_2 H(Y|b_2), \tag{5.101}$$

with

$$H(B) = -q_1 \ln q_1 - q_2 \ln q_2. \tag{5.102}$$

Inserting the relation

$$H(Y|b_i) = H(X) + \ln|b_i| \quad , \quad i = 1,2,, \tag{5.103}$$

into (5.101) yields

$$H(B,Y) = H(B) + H(X) + q_1 \ln|b_1| + q_2 \ln|q_2|.$$

(ii) According to equation (5.98) we then have

$$H(Y) = H(X) + \max_{q_1,q_2}\left(q_1 \ln|b_1| + q_2 \ln|b_2|\right), \tag{5.104}$$

subject to the condition

$$H(B) = h, \tag{5.105}$$

$$q_1 + q_2 = 1. \tag{5.106}$$

(iii) The Lagrange multiplier technique provides the optimal (q_1^*, q_2^*) in the form

$$q_i^* = |b_i^*|^{c^*} / \left(|b_1|^{c^*} + |b_2|^{c^*}\right) \quad , \quad i = 1,2, \tag{5.107}$$

where c^* is the solution of the equation

$$-q_1^* \ln q_1^* - q_2^* \ln q_2^* = h, \tag{5.108}$$

therefore the expression of the entropy, which is

$$H(Y) = H(X) + \frac{|b_1|^{c^*} \ln|b_1| + |b_2|^{c^*} \ln|b_2|}{|b_1|^{c^*} + |b_2|^{c^*}}. \tag{5.109}$$

Example 2. Assume that the exact probability density function of the random variable Y is unknown, and that all we have at hand is the distribution of the conditional entropy $H(Y|x)$ where $x \in \Re^n$ is the value of the random variable X.

Entropy of Y.

The conditional entropy $H(Y|X)$ is

$$H(Y|X) = \int_{\Re^n} p(x)H(Y|x)dx, \tag{5.110}$$

and maximizing the latter with the additional condition $H(X)=h$ we find that there exists a real valued constant c such that

$$H(Y) = \frac{\int_{\Re^n} e^{cH(Y|x)}H(Y|x)dx}{\int_{\Re^n} e^{cH(Y|x)}dx}. \tag{5.111}$$

On the determination of the value of c.

We thus obtain a family of entropies indexed by the parameter c, and of course the question is to determine which value of c is the right one.

According to the optimization procedure this value is defined by the condition $H(X) = h$, where $H(X)$ is the Shannon entropy corresponding to the probability density

$$p(x) = \frac{\exp\big(cH(Y|x)\big)}{\int_{\Re^n} \exp\big(cH(Y|x)\big)dx}. \tag{5.112}$$

Making the substitution into (5.110) we thus obtain the implicit equation

$$c\int_{\Re^n} H(Y|x)e^{cH(Y|x)}dx = K(c)\big(\ln K(c) - h\big), \tag{5.113}$$

with

$$K(c) := \int_{\Re^n} e^{cH(Y|x)}dx. \tag{5.114}$$

The constant h can be selected irrespective of the distribution $H(Y|x)$ of conditional entropy, and for a given value of h $H(Y|X)$ achieves its maximum value , that is to say $H(Y)$, when X and Y are independent.

As a result it is not easy to determine the value of c in this way, and in order to circumvent this dificulty we shall refer to the entropy of order c of random variables in the equation (5.18). Indeed, comparing the two problems which are very similar, we are led to the following statement:

Identification principle.

Assume that X in the equation (5.111) is defined on the domain Ω; then a suitable value of c would be defined by the equation

$$\int_{\Omega} e^{cH(Y|x)} dx = 1. \quad \blacksquare \tag{5.115}$$

This principle stems from the requirement that the values of c in the expressions (5.18) and (5.111) should be provided by the same equation. And in the first case we have obtained this value in writing that

$$\sum_{i=1}^{n} p_i^c = 1. \tag{5.116}$$

5.9. External quantum entropies of distributed matrices

In this section we assume that the explicit definition of the distributed matrix $A(z)$ is not available, and that all we have at hand is the value of one of its entropies. Then by using the approach of the preceding section it is possible to define meaningfully quantum measures of the amount of uncertainty thus involved in the family of matrices, and this is summarized in the following result.

PROPOSITION 5.8. *Let $A(z)$ denote a $n \times n$ matrix which depends explicitly upon a distributed parameter $z \in D \subset \Re^m$; and assume that all we know about $A(z)$ is the value of its von Neumann quantum entropy $QH_{c(z)}(A(z))$. Then a measure of the amount of uncertainty it involves on the domain D, and which is fully consistent with von Neumann quantum entropy of density matrices, is the external von Neumann quantum entropy $QH'_{b,c(z)}(A_z; D)$ defined by the expression*

$$QH'_{b,c(z)}(A_z; D) := \frac{\int_D \exp\left[bQH_{c(z)}(A(z))\right] QH_{c(z)}(A(z))dz}{\int_D \exp\left[bQH_{c(z)}(A(z))\right]dz}, \tag{5.117}$$

where b is a real-valued constant. $\blacksquare$

PROPOSITION 5.9. *Consider the distributed matrix A(z) of proposition 5.8, and assume that all we know about it is the value of its Renyi quantum entropy* $QH_{R,s}(A(z))$. *Then a measure of the amount of uncertainty it involves on the domain D and which is fully consistent with von Neumann quantum entropy of density matrices on the one hand, and Renyi entropy of random variables on the other hands, is the external Renyi quantum entropy* $QH'_{R,s,s'}(A_z;D,u)$ *defined by the expression*

$$QH'_{R,s,s'}(A_z;D,u)$$

$$= \frac{1}{1-s'} \frac{\int_D \left[u+\exp\left((1-s')QH_{R,s}(A(z))\right)\right]^{\frac{s'}{1-s'}} \exp\left((1-s')QH_{R,s}(A(z))\right)dz}{\int_D \left[u+\exp\left((1-s')QH_{R,s}(A(z))\right)\right]^{\frac{s'}{1-s'}} dz}, \quad (5.118)$$

where u is a non-negative real-valued parameter, and (s,s') is such that $s,s' \triangleright 0$, $s,s' \neq 1$.■

PROPOSITION 5.10. *Consider the distributed matrix of proposition 5.8, and assume that all we know about it is the value of its structural quantum entropy* $QH_{C,s}(A(z))$. *Then a measure of the amount of uncertainty it involves on the domain D, which, furthermore, is fully consistent with von Neumann quantum entropy of density matrices on the one hand and structural entropy of random variables on the other, is the external structural quantum entropy* $QH'_{C,s,s'}(A_z;D,u)$ *defined by the expression*

$$QH'_{C,.s.,s'}(A_z;D,u) \quad (5.119)$$

$$= \frac{1}{e^{1-s'}-1} \frac{\int_D \left[u+e^{(1-s')QH_{C,s}(A(z))}\right]^{\frac{s'}{1-s'}} \left(e^{(1-s')QH_{C,s}(A(z))}-1\right)dz}{\int_D \left[u+e^{(1-s')QH_{C,s}(A(z))}\right]^{\frac{s'}{1-s'}} dz},$$

where u is a non-real-valued parameter, and (s,s") is such that $s,s' \triangleright 0$; $s,s' \neq 1$.■

The proof of these results is direct if one refers to the entropies of non-random functions defined in the third chapter. Indeed, everything happens as if we are determining the entropies of a deterministic function *f(z)* such that

$$\ln|f'(z)| := QH_{c(z)}(A(z)), \tag{5.120}$$

or

$$\ln|f'(z)| := QH_{R,s}(A(z)), \tag{5.121}$$

or

$$\ln|f'(z)| := QH_{C,s}(A(z)), \tag{5.122}$$

depending upon whether we are dealing with Shannon entropy, Renyi entropy, or structural entropy.

5.10. On the relation beween cross-entropy and mutual information

5.10.1. On the definition of cross-entropy

Our first remark is related to (Shannon) information theory. In this framework, the Kullback Leibler relative entropy, or divergence or cross-entropy

$$H(q,p) := \int_{\Re} q(x) \ln \frac{q(x)}{p(x)} dx \quad , \quad x \in \Re, \tag{5.123}$$

is non negative when $p(.)$ and $q(.)$ are two probability density functions.

Indeed, loosely speaking, this is due to the constraint

$$\int_{\Re} p(x)dx = 1, \tag{5.124}$$

which provides the admissible variations in the form

$$\int_{\Re} \delta p(x)dx = 0,$$

in such a manner that one has

$$\begin{aligned} H(p+\delta p, p) &\cong \int_{\Re} (p+\delta p)\frac{\delta p}{p} dx, \\ &\cong \int_{\Re} (\delta p)^2 dx. \end{aligned}$$

For arbitrary positive functions the condition (5.124) does not necessarily hold, and this is the reason why we introduced the *cross-entropic variance* $\sigma_H^2(q,p;\Re)$ defined by the expression

$$\sigma_H^2(q,p;\Re) \;:=\; \int_{\Re} q(x)\ln^2\frac{q(x)}{p(x)}dx - H^2(q,p)\;. \tag{5.125}$$

Our second remark refers to quantum information. von Neumann first defined quantum entropy of density matrices (loosely speaking, matrices with positive eigenvalues and unit trace) and later some authors [5.3,5.9] suggested a definition of quantum cross-entropy for two density matrices ρ and ρ' in the form

$$S(\rho|\rho') \;=\; tr\left(\rho(\ln\rho - \ln\rho')\right), \tag{5.126}$$

of which the explicit meaning is as follows. Let μ_i and μ'_j denote the positive eigenvalues of ρ and ρ' respectively; and assume that $\mu_i \neq \mu_j$ for allL *(i,j);* and likewise for μ'_i. First, one has

$$\begin{aligned} tr(\rho'\ln\rho') \;&:=\; \sum_i \mu'_i \ln \mu'_i \\ &=\; tr\left(\sum_i \mu'_i \ln \mu'_i R'_i\right). \end{aligned} \tag{5.127}$$

Next, expanding ρ and ρ' along with their respective constituent matrices, one obtains

$$\rho \;=\; \sum_{i=1}^{n}\mu_i R_i \;,\quad R_iR_j \;=\; 0\,,i\neq j \;;\quad R_i^n \;=\; R_i \;, \tag{5.128}$$

$$\rho' \;=\; \sum_{j=1}^{n}\mu'_j R'_j \;,\quad R'_iR'_j \;=\; 0\,,i\neq j \;;\quad (R'_j)^n \;=\; R'_j, \tag{5.129}$$

therefore

$$tr(\rho'\ln\rho) \;=\; tr\left(\sum_{j=1}^{n}\sum_{i=1}^{n}\mu'_j \ln \mu_i . R'_jR_i\right), \tag{5.130}$$

and

$$S(\rho,\rho') = tr\left(\sum_{j=1}^{n}\mu'_j \ln \mu'_j.R'_j - \sum_{j=1}^{n}\sum_{i=1}^{n}\mu'_j \ln \mu_i.R'_jR_i\right). \tag{5.131}$$

This expression can be thought of as a formal extension of the Kullback Leibler divergence, and it exhibits many interesting and meaningful properties. Nevertheless, it gives rise to the question of whether the quantum entropies should or should not be independent of the basis matrices. For instance, could we write the quantum cross-entropy in the alternative form

$$\tilde{S}(\rho',\rho) := \sum_{i=1}^{n}\mu'_i \ln \frac{\mu'_i}{\mu_i} \; ? \tag{5.132}$$

Indeed, $\tilde{S}(\rho',\rho)$ is positive or equal to zero, and this last case occurs when and only when one has $\mu'_i = \mu_i$ for every i, features which look to be quite sufficient for a suitable modelling, as far as the quantum entropy of a density matrix is defined by the eigenvalues of the latter only.

At first glance this point of view would be quite consistent with the derivation of quantum entropy as presented in the section 5.4 where we did not explicitly refer to the constituent matrices themselves because we did not need to do so! Indeed, in terms of probability the density matrices ρ and ρ' will be characterized by their eigenvalues, and they will be considered as being identical (from the information-theoretic standpoint) when these eigenvalues are equal irrespective of the eigenvectors.

In order to obtain more insight into this remark let us consider the special case when the density matrices ρ and ρ' in the expression (5.128) and (5.129) have two eigenvalues only to yield,

$$\begin{aligned} S(\rho,\rho') = {} & tr\left(\mu'_1 \ln \mu'_1.R'_1 - \mu'_1 \ln \mu_1.R'_1R_1 - \mu'_1 \ln \mu_2.R'_1R_2\right) \\ & + tr\left(\mu'_2 \ln \mu_2.R'_2 - \mu'_2 \ln \mu_2.R'_2R_2 - \mu'_2 \ln \mu_1.R'_2R_1\right) . \end{aligned} \tag{5.133}$$

This expression is equivalent to $\tilde{S}(\rho',\rho)$ in equation (5.132) when and only when one has

$$R'_1R_2 = 0, \qquad R'_2R_2 = 0, \tag{5.134}$$

$$tr(R_1'R_1) = 1, \qquad tr(R_2'R_1) = 1. \tag{5.135}$$

It is clear that these conditions are far from being satisfied in the general case, in other words the extension via matrix functions is probably more general than the extension which we would obtain by using a strict interpretation of von Neumann quantum entropy.

5.10.2. Further comments on the extension via matrix functions

(i) Consider the $n \times n$ matrix A with its basis expansion

$$A = \sum_{i=1}^{n} \lambda_i A_i, \tag{5.136}$$

then making explicit the equation (5.41) we obtain the expression

$$QH_c(A) = \frac{tr\left(\sum_{i=1}^{n} |\lambda_i|^c \ln|\lambda_i| . A_i\right)}{tr\left(\sum_{i=1}^{n} |\lambda_i|^c A_i\right)}, \tag{5.137}$$

which yields the equation (5.21) as a result of property

$$tr(A_i) = 1, \qquad i = 1,2,...,n. \tag{5.138}$$

Loosely speaking, one can suppose that all this simplification is caused by the equation

$$|A_i| = 0 \ , \ i = 1,2,..,n \ ,$$

which can be considered as illustrating that A_i does not involve any amount of information.

(ii) The second remark refers to matrices functions $f(A,B)$ where B is a $n \times n$ matrix with the basis expansion

$$B = \sum_{i=1}^{n} \lambda_i' B_i \cdot \tag{5.139}$$

Well obviously $f(A,B)$ is a new matrix with the eigenvalues $\lambda_i(f(A,B))$, $i = 1,2,...,n$; therefore we can define the corresponding quantum entropy

$$QH_c(f(A,B)) = \frac{\sum_{i=1}^{n} |\lambda_i(f(A,B)| \ln |\lambda_i(f(A,B))|}{\sum_{i=1}^{n} |\lambda_i f(A,B)|} \cdot \tag{5.140}$$

But if, instead, we want to calculate $QH_c(f(A,B))$ via the expansions (5.136) and (5.139) we shall then have to consider the expression

$$f(A,B) = \sum_{i=1}^{n} \sum_{j=1}^{n} f(\lambda_i, \lambda_j') A_i B_j , \tag{5.141}$$

where the matrices A_iB_j are not necessarily the bases of $f(A,B)$.

5.10.3. Cross-entropy and mutual information

Let us bear in mind that if $X \in \Re$ and $Y \in \Re$ are two random variables with the respective probability densities $p(x)$ and $q(y)$ and the joint probability density $r(x,y)$, then according to Shannon their mutual information $I(X,Y)$ is equal to

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{5.142}$$

$$= \int_{\Re^2} r(x,y) \ln \frac{r(x,y)}{p(x)q(y)} dxdy \tag{5.143}$$

$$= H(r, qp).$$

In other words, the divergence $H(q,p)$ can be thought of as a special case of mutual information, or again as an extension of the notion of mutual information to the one-dimensional case.

This relation gave rise to a long debate amongst scientists, some of them claiming that the basic concept of information is the divergence, whilst others consider mutual

information as the central notion. Here, in our modelling purpose, the point at issue is merely a matter of physical meaning. $I(X,Y)$ is actually a measure of information about something (the information involved in X about Y) whilst $H(q,p)$ is just an informational distance-like quantity which defines the deviation between $p(x)$ and $q(x)$. The difference between $H(q,p)$ and $I(X,Y)$ will be clearer if we rewrite the latter in the form

$$I(X,Y) = H(Y) - H(Y|X), \tag{5.144}$$

where $H(Y|X)$ is the conditional entropy of Y given X.

As a conclusion of these comments, when we try to define a concept of quantum divergence for non-probabilistic matrices we shall have in mind the following remarks.

(i) If we can meaningfully define the quantum entropy of a pair of matrices, then it will be more suitable to look for a generalization of the mutual information $I(X,Y)$.

(ii) If all we have at hand is the definition of the matrices, then there is no alternative but to generalize the divergence $H(q,p)$.

In this way we may expect to have several possible meaningful definitions of quantum divergence between density matrices on the one hand, and between non-probabilistic matrices on the other.

5.11. Quantum divergences of non-probabilistic matrices

5.11.1. Quantum divergence of density matrices

The above remarks suggest various possible definitions for the divergence of density matrices in quantum mechanics, which are summarized as follows.

We consider two density matrices ρ and ρ' which are expanded in terms of their respective constituent matrices in the forms

$$\rho = \mu_1 R_1 + \mu_2 R_2 + ... + \mu_n R_n, \tag{5.145}$$
$$\rho' = \mu'_1 R'_1 + \mu'_2 R'_2 + ... + \mu'_n R'_n. \tag{5.146}$$

The eigenvalues are pairwise indexed in such a manner that one can write

$$\rho = V\, diag(\mu_1, \mu_2, ..., \mu_n)\, V^{-1}, \tag{5.147}$$
$$\rho' = V'\, diag(\mu'_1, \mu'_2, ..., \mu'_n)(V')^{-1}\ . \tag{5.148}$$

MODEL 1 (Ref [5.3]). The quantum cross-entropy $S(\rho',\rho)$ is defined by the expression (5.126). ■

MODEL 2. Let us introduce the conditional density matrix

$$(\rho|\rho') := \mu_1 R_1' + \mu_2 R_2' + ... + \mu_n R_n' , \tag{5.149}$$

then the quantum cross-entropy is defined as

$$\begin{aligned} S_2(\rho',\rho) &:= tr\left(\rho' \ln \rho' - \rho' \ln (\rho|\rho')\right), \\ &= S\left(\rho',(\rho|\rho')\right) . \blacksquare \end{aligned} \tag{5.150}$$

MODEL 3. Assume that ρ and ρ' are respectively defined on the Hilbert spaces E_n and E_n', and let ρ'' denote a third density matrix defined on the cartesian or tensor product $E_n \otimes E_n'$; then the quantum divergence between ρ'' and $\rho' \otimes \rho$ is defined as

$$S_3(\rho'',\rho' \otimes \rho) := |QH(\rho) + QH(\rho') - QH(\rho'')|, \tag{5.151}$$

$$= |QH(\rho \otimes \rho') - QH(\rho'')|. \blacksquare \tag{5.152}$$

5.11.2. Quantum cross-entropic variance of density matrices

The cross-entropic variance of random variables as defined by the equation (5.125) can be extended to density matrices as follows, by using the matrix notation

$$\begin{aligned} tr_\Omega A &= tr_\Omega (a_{ij}) , \\ &= \sum_{i \in \Omega} a_{ii} . \end{aligned} \tag{5.153}$$

MODEL 1. The quantum cross-entropic variance of the density matrices ρ and ρ' (see equations (5.145,5.146) on the set Ω of indexes, is defined by the expression

$$\begin{aligned} Q\sigma_H^2(\rho',\rho;\Omega) := & \frac{tr_\Omega\, \rho'(\ln \rho' - \ln \rho)^2}{tr_\Omega\, \rho'} \\ & - \left[\frac{tr_\Omega\, \rho'(\ln \rho' - \ln \rho)}{tr_\Omega\, \rho'}\right]^2 . \blacksquare \end{aligned} \tag{5.154}$$

MODEL 2. This second model is defined as the quantum cross-entropic variance between the density matrices ρ and $(\rho|\rho')$, that is to say

$$Q\sigma_H^2(\rho',(\rho|\rho');\Omega) := \frac{tr_\Omega\, \rho'\left[\ln \rho' - \ln(\rho|\rho')\right]^2}{tr_\Omega\, \rho'} - \left[\frac{tr_\Omega\, \rho'\left(\ln \rho' - \ln[\rho|\rho']\right)}{tr_\Omega\, \rho'}\right]^2 . \blacksquare \tag{5.155}$$

5.11.3. *Quantum cross-entropic variance of non probabiistic matrices*

These models of quantum cross-entropic variance for density matrices can be directly extended to non-probabilistic matrices by using the matrix function $m(A)$ introduced in the section 5.3, equation (5.40).

We then consider two $n \times n$ matrices A and B, and we have their quantum cross-entropic variances as follows:

MODEL 1.

$$Q\sigma_{H,c}^2(A,B) = \frac{tr\left[m^c(A)(\ln(m(A)) - \ln(m(B))^2\right]}{tr(m^c(A))} - \left[\frac{tr\left(m^c(A)[\ln(m(A)) - \ln(m(B))]\right)}{tr(m^c(A))}\right]^c . \blacksquare \tag{5.156}$$

MODEL 2 [4.6]

$$Q\sigma_{H,c}^2(A,(B|A)) = \frac{tr\left[m^c(A)\left(\ln(m(A)) - \ln(m(B|A))\right)^2\right]}{tr(m^c(A))} - \left[\frac{tr\left(m^c(A)[\ln(m(A)) - \ln(m(B|A))]\right)}{tr(m^c(A))}\right]^2 . \blacksquare \tag{5.157}$$

5.12. Singular quantum entropies of square matrices

5.12.1. Singular values and singular value decomposition of matrices

The concept of singular value and of singular value decomposition are very useful in the general study of dynamical systems, and we herein bear in mind the essentials of the topic.

Let A be a complex $n \times m$ matrix of rank r, and let A^* denote its conjugate transpose. The matrix A^*A is Hermitian and positive semi definite, it then has an orthogonal basis of eigenvectors

$$A^*A\, v_i \;=\; \lambda_i(A^*A)\, v_i,$$

where the eigenvalues $\lambda_{i(A^*A)}$ of (A^*A) are real.

By definition, the singular values $s_{i(A)}$ of A are

$$s_i(A) \;:=\; \left[\lambda_i(A^*A)\right]^{1/2},\ i = 1,2,..,r \tag{5.158}$$

This being so, we introduce the vectors

$$u_i \;:=\; \frac{1}{s_i} A v_i \quad ,\quad i = 1,2,\ldots r,, \tag{5.159}$$

and we notice that one has the equality

$$\begin{aligned} u_i^* u_j \;&=\; (s_i s_j)^{-1}\, v_i^* A^* A v_j \\ &=\; \lambda_j (s_i s_j)^{-1} v_i^* v_j \end{aligned} \tag{5.160}$$

for every i and j, in such a manner that the set $\{u_1, u_2, ..., u_r\}$ is orthonormal.

Then one can show that A can be written as the sum of rank one matrices

$$A \;=\; \sum_{i=1}^{n} s_i u_i v_i^*, \tag{5.161}$$

which is sometimes referred to as the *singular value decomposition* of A.

From the practical standpoint, one of the main interests of considering singular values can be summarized in the following remarks. Firstly, the largest singular value is exactly the operator norm of A with respect to the Euclidean norm on vectors; and, secondly, the singular values are very easy to manipulate since they are positive. In addition, the concept of singular value applies to both square and rectangular matrices. It follows that it is recommended to use singular values in some problems related to the stability of dynamical systems.

5.12.2. Singular quantum entropies of non probabilistic matrices

The basis $\{u_i v_i^*\}$ in equation (5.161) is orthonormal, and analogously with the decomposition (5.68), we are led to duplicate the quantum entropic expressions (5.21), (5.31) and (5.32) in terms of singular values. We then obtain a new family of quantum entropies referred to as *singular quantum entropies.*

DEFINITION 5.3. The von Neumann singular quantum entropy of order c of the $n \times m$ matrix A is

$$Q_s H_c(A) := \frac{1}{2} \frac{tr\left[(A^*A)^{c/2} \ln(A^*A)\right]}{tr\left[(A^*A)^{c/2}\right]}, c \in \Re, \tag{5.162}$$

its Renyi singular quantum entropy of order s is

$$Q_s H_{R,s}(A) := -\frac{1}{1-s} \ln \frac{tr\left[(A^*A)^{s/2}\right]}{tr\left[(A^*A)^{1/2}\right]} \quad , \quad s \rhd 0 \quad , \quad s \neq 1 \ , \tag{5.163}$$

and its structural singular quantum entropy of order s is

$$Q_s H_{C,s}(A) = \frac{1}{e^{1-s}-1} \left[\frac{tr\left(\left(A^*A\right)^{1/2}\right)}{tr\left(\left(A^*A\right)^{s/2}\right)} - 1 \right], \quad s \rhd 0, \quad s \neq 1 \blacksquare \tag{5.164}$$

These expressions are direct applications of the equations ((5.41), (5.42), (5.43)) as a consequence of the eigenvalue decomposition

$$m(A^*A) = \sum_{i=1}^{r} s_i^2 (A^*A)_i \cdot \tag{5.165}$$

A prospect of the application of this concept of singular quantum entropy is the analysis of the complexity of dynamical systems.

5.13. Quantum entropies of non probabilistic operators

Up to now we have considered matrices only, but it is clear that all the above results directly apply to non-probabilistic operators defined on Hilbert spaces, that is to say operators which have any meaning in terms of probability. The main difference with matrices is that, sometimes, one may have infinite sets of eigenvalues indexed by continuous parameters. Some examples will illustrate the matter.

Example 1. *Quantum entropy of differential operators.* $M(D)$ and $N(D)$ denoting two polynomials of the differential operator $D := d(.)/dt$, we consider the differential equation

$$M(D).y(t) = N(D).x(t) \quad , \quad t \geq 0, x \in \Re, y \in \Re, \tag{5.166}$$

with zero initial conditions. We rewrite

$$y(t) = \frac{N(D)}{M(D)}.x(t),$$

where $x(t)$ is a given function, and $y(t)$ is the solution to be determined. By using Laplace's transform (the complex variable s here is not to be confused with the order of the Renyi entropy $H_{R,s}(X)$! As is said in computer programming, it is only a local variable)

$$X(s) := \int_0^\infty e^{-st} x(t) dt, \tag{5.167}$$

we obtain the operational equation

$$\begin{aligned} Y(s) &= \frac{N(s)}{M(s)} X(s) \\ &= G(s)X(s), \end{aligned} \tag{5.168}$$

and it is well known that the eigenvalues of the operator $G(D) = N(D)/M(D)$ are $\exp(i\omega t)$; $i^2 = -1$, $\omega \in \Re$, with the corresponding eivalues $G(i\omega)$. We then have the quantum entropy

$$QH_c(G(D);\Omega) = \frac{\int_\Omega |G(i\omega)|^c \ln|G(i\omega)| d\omega}{\int_\Omega |G(i\omega)|^c d\omega} . \tag{5.169}$$

In automatic control theory, $G(i\omega)$ is referred to as the transfer function of the system described by the differential equation (5.166).

Example 2. *Quantum entropy of n-dimensional differential operators.* We now assume that we are dealing with a set of differential equations in such a manner that, formally, we shall once more consider the equation (5.168), in which now $G(s)$ is a $n \times n$ matrix (transfer matrix in automatic control). According to the section 5.7, equation (5.84), we can then write the quantum entropy of $G(D)$ in the form

$$QH_{b,c}(G(D);\Omega)$$

$$= \frac{\int_\Omega \exp\left[(b-c)QH_c(G(i\omega))\right] tr\left[m^c(G(i\omega)) \ln m(G(i\omega))\right] d\omega}{\int_\Omega \exp\left[(b-c)QH_c(G(i\omega))\right] tr\left[m^c(G(i\omega))\right] d\omega} . \tag{5.170}$$

Example 3. Quantum entropy of integral operators. We consider the one-dimensional Fredholm integral equation

$$y(x) = f(x) + \int_a^b K(x,\xi) y(\xi) d\xi . \tag{5.171}$$

If we denote by λ_i, $i = 1,2,\ldots,n$ the eigenvalues of $K(.)$ on the interval $[a,b]$, then its quantum entropy is

$$QH_{c,c}(K(.);[1,n]) = \frac{\sum_{i=1}^{i=n} |\lambda_i(K)|^c \ln|\lambda_i(K)|}{\sum_{i=1}^{i=n} |\lambda_i(K)|^c} . \tag{5.172}$$

Example 4. *Singular quantum entropy in automatic control.* According to the equation (5.170), in which we set $b = c$, a simplified expression for the singular quantum entropy of the matrix G(s) is

$$Q_s H_{c,c}(G(D);\Omega) = \frac{\int_\Omega tr\left[(G^*G)^{c/2}\ln(G^*G)\right]d\omega}{\int_\Omega tr\left[(G^*G)^{c/2}\right]d\omega} . \tag{5.173}$$

This expression exhibits a weighted combination of the singular values of the transfer matrix, which could be of help in some problems related to stability, complexity of structure, and reduction of dynamical systems for instance.

In all the preceding examples we have referred only to the von Neumann quantum entropies of the various operators, but obviously one can write their Renyi entropies and their structural entropies in exactly the same way.

5.14. Concluding remarks

In this chapter we have proposed new concepts of quantum entropies which apply to matrices which are not necessarily density matrices in the sense of quantum mechanics, that is to say, positive trace class matrices with unit trace, therefore the term of non-probabilistic matrices to emphasize this feature. Obviously, the first quantum entropy thus introduced can be thought of as a generalization of von Neumann quantum mechanical entropy, but in the present framework this property is only incidental.

Indeed, and this remark is of paramount importance, quantum entropy has been obtained as a consequence of Shannon information theory on the one hand, and of the information theory of non-random functions on the other, via a principle of maximum conditional entropy.In other words, while, as is usually understood, Shannon entropy of random variables can be thought of as a special case of von Neumann quantum entropy, here we have shown that in contrast the expression of the latter can be derived from the former.

We thus have a new unified framework which encompasses von Neumann quantum information, Shannon information, and a probability-free approach to modelling information involved in non random-functions and in non-probabilistic matrices.

A new theory is interesting when and only when it can be of help in understanding and solving some practical problems. Accordingly, the quantum entropies above appear to be quite suitable measures of complexity, they exhibit new weighted

combinations of eigenvalues; and, of course, it is tempting to examine what kind of results one can obtain when they are used in the analysis of dynamical systems, for instance.

References

5.1 FADEEV, D.K.; On the concept of entropy of a finite probabilistic scheme (in Russian), *Uspehi Mat. Nauk (N.S.)*, No 1 (67), pp 227-231, 1956

5.2 HARTLEY, R.V.; Transmission of information, *Bell System Techn. J.*, Vol, 7, pp 535-563 1968

5.3 HUMEGAKI, H.; Conditional expectation in an operator algebra IV, *Kodai Math Sem. Rep.*, Vol 14, pp 59-85, 1962

5.4 JAYNES, E.T.; Information theory and statistical mechanics I, *Physical Review*, Vol 106, pp 620-630, 1957

5.5 JUMARIE, G.; Quantum entropies of non probabilistic matrices, *Journal of Mathematical Physics,* Vol 32, No 32, No 11, pp 2967-2972, 1991

5.6 JUMARIE, G.; Information and quantum information of non probabilistic matrices. Theory and applications, *J. Systems Analysis Modelling Simulation*, Vol 9, No 3, pp 179-213, 1992

5.7 NEUMANN, J. Von; Die Eindeutigkeit der Schrodingerschen Operatoren, *Math. Annalem*, Vol 104, pp 570-578, 1931

5.8 NEUMANN, J. Von; Die mathematischen Grundlagen der Quantenmechanik, Springer Verlag, Berlin, 1932

5.9 OHYA, M.; On compound state and mutual information in quantum information theory, *IEEE Trans. Information Theory,* Vol IT-29, No 5, pp 770-774, 1983

5.10 SHANNON, C.E.; A mathematical theory of communication, *Bell System Tech. J.*, Vol 2, pp 379-423, pp 623-659, 1948

Chapter 6

Complex-Valued Fractional Brownian Motion of Order n. Part I

Statistics is the first of the inexact sciences.
Edmond and Jules de Goncourt

6.1. Introduction

6.1.1. Purpose of the present chapter

In the present chapter we shall consider the problem of modelling complex-valued fractional Brownian motion of order n (C-$(fBm)_n$ in the following).

Brownian motion and Gaussian white noise are two companion stochastic processes. They are intimately linked, and one of them can be used to define the other. In the engineering mathematics framework one is used firstly to define Gaussian white noise, and then, Brownian motion appears as its integral. In contrast, mathematicians firstly define Brownian motion, of which the derivative provides Gaussian white noise.

We have exactly the same routes to fractional Brownian motion.. Mandelbrot [6.8] suggested a model in the form of an integral of Gaussian white noise, which is related to the theory of derivatives of fractional order; and later Sainty [6.9], in the complex plane, constructed another model by directly extending Levy's ideas.

Here, we shall work with the engineering mathematics point of view. In this framework, all we need (to define fractional Brownian motion) is a stochastic process which exhibits properties similar to those of Gaussian white noise, but which, in addition, is such that its first moments are zero up to the order n. And such a process is exactly provided by the complex roots of the unity!

In the following, we shall consider various possible approaches to define C-$(fBm)_n$. We shall firstl build a model by using a Gaussian white noise which rotates on the complex roots of the unity. Unfortunately this derivation does not work for odd order ($n = 2k+1$); and to circumvent this problem we shall use a rotating semi-Gaussian white noise. Then, extending a one-dimensional random walk, we shall obtain a first model of C-$(fBm)_n$ by using a random walk in the complex plane which involves radial steps. And then we shall generalize this approach by using a random walk with non-radial steps,

G. Jumarie, *Maximum Entropy, Information Without Probability and Complex Fractals*
© Springer Science+Business Media Dordrecht 2000

referred to as an approach to $C\text{-}(fBm)_n$ with complex variance. But before doing so, we set out the heuristic of our approach.

6.1.2. Presentation of the intuitive idea

(i) Again in the engineering mathematics framework (or elementary mathematical physics!), consider the stochastic differential equation

$$\dot{x}(t) = w(t) \quad , \quad pr\{x(0) = 0\} = 1, \tag{6.1}$$

where $w(t)$ is a Gaussian white noise with zero mean and variance $\sigma^2(t)$. Formally, we shall write

$$x(t) = \int_0^t w(\tau)d\tau \ , \tag{6.2}$$

and a simple calculation yields the mathematical expectation

$$E\{x(t)\} = 0, \tag{6.3}$$

with the variance

$$\begin{aligned} Var\{x(t)\} &= E\left\{\int_0^t\int_0^t w(\tau)w(s)d\tau ds\right\} \\ &= \int_0^t \sigma^2(\tau)d\tau. \end{aligned} \tag{6.4}$$

(ii) This being so, we write $x(t)$ in the form

$$x(t) = \int_0^t dx(\tau)d\tau \ , \tag{6.5}$$

where the increments $dx(\tau)$'s are mutually independent, to have the equality

$$Var\{x(t)\} = \int_0^t Var\{dx(\tau)\} \ ; \tag{6.6}$$

(iii) and comparing (6.4) with (6.6) yields

$$Var\{dx(t)\} = E\{dx^2(t)\},$$
$$= \sigma^2(t)dt \quad . \tag{6.7}$$

As a result, it is quite meaningful to rewrite (6.1) in the form

$$x(t) = \int_0^t w(t)(dt)^{1/2}. \tag{6.8}$$

which is the Maruyama's notation (see, for instance, Levy [6.6]).

If we want to use similar properties to construct $C(fBm)_n$, we need a white noise of which the (n-1) first moments are zero, and fortunately it is provided by the complex roots of the unity.

But before describing our model, for the sake of completeness, in the next section we shall give a brief background of some stochastic processes of fractional order.

Warning to the reader

Throughout the book, exactly as in computer programming, all the variables are local variables, and in the following, in order to comply with standard notations, we shall introduce the complex variable $z = x+iy$ which, of course, is not to be confused with the distributed parameter z in Chapter 4.

6.2. Review of some stochastic processes of fractional order

6.2.1 Modelling via derivatives of fractional order

One of the most popular models of fractional Brownian motion, or at least the first, was introduced by Mandelbrot and Van Ness [6.7] who defined it by the expression (the symbol := means that the left side is defined by the right hand one),

$$b_\alpha(t) := \frac{1}{\Gamma\left(\alpha+\frac{1}{2}\right)}\int_{-\infty}^{t}(t-\tau)^{\alpha-\frac{1}{2}}w(\tau)d\tau \quad , \quad 0 \lhd \alpha \lhd 1, \tag{6.9}$$

where $w(t)$ is a Gaussian white noise with zero mean, and $\Gamma(.)$ is the Eulerian function

$$\Gamma(x) := \int_0^\infty e^{-u}u^{x-1}du \; . \tag{6.10}$$

The basic property of $b_\alpha(t)$ is that the variance of its increment satisfies the condition

$$E\left\{\left|b_\alpha(t)-b_\alpha(\tau)\right|^2\right\} \propto |t-\tau|^{2\alpha} \tag{6.11}$$

which can be re-written in the form

$$\left\langle\left|b_\alpha(t)-b_\alpha(\tau)\right|\right\rangle \propto |t-\tau|^{2\alpha}, \tag{6.12}$$

where the bracket $\langle(.)\rangle$ denotes ensemble averages over several path samples of $b_\alpha(t)$ (remark the slight difference with mathematical expectation).

$b_\alpha(t)$ exhibits the property of *self-affinity*: clearly, given a scalar factor ρ one has the equalities

$$E\left\{\left|\Delta b_\alpha(\rho t)\right|^2\right\} \propto \rho^{2\alpha} E\left\{\left|\Delta b_\alpha(t)\right|\right\}, \tag{6.13}$$

and

$$E\left\{\left|\Delta(\rho^\alpha b_\alpha(t)\right|^\alpha\right\} \propto \rho^{2\alpha} E\left\{\left|\Delta b_\alpha(t)\right|\right\}. \tag{6.14}$$

6.2.2 Modelling via mixing independent processes with short range dependence

The following model has been proposed by Cox [6.1].

For each v, $0 < v < 1$, let $x(t,v)$ be a stationary process with zero mean. Let $\{N(u)\}$ denote a Poisson process of rate ψ, and consider the new process $x_c(t)$

$$x_c(t) := \int_0^1 x(tv,v)v^{\frac{\alpha}{2}-\frac{1}{2}}dN(v), \quad 0 \lhd \alpha \lhd 1. \tag{6.15}$$

which is a stationary process with zero mean. If σ^2 and $\rho_{11}(t)$ are the variance and the correlation of $x(t)$ respectively, and if one defines

$$A(\alpha) := \int_0^\infty u^{\alpha-1}\rho_{11}(u)du\ , \tag{6.16}$$

then the covariance of $x_C(t)$ is

$$E\{x_c(t)x_c(t+\tau)\} = \psi\sigma^2 A(\alpha)\tau^{-\alpha}. \tag{6.17}$$

$x_C(t)$ has an asymptotic self-similarity property in the following sense. Assume that the process is averaged over periods of length T into a discrete-time process, clearly

$$x_{c,T}(t) := \frac{1}{T}\int_{tT}^{(t+1)T} x_c(u)du \quad , \quad (t = 0,\pm1,\pm2,...); \tag{6.18}$$

then one has the equality

$$Var\{x_{c,T}(t)\} \propto T^{-\alpha} \quad as \quad T \uparrow +\infty \tag{6.19}$$

6.2.3 Modelling via Mittag Leffler function

Sainty [6.9] proposed a model of *C-(fBm)$_n$* of which the basic features can be summarized as follows.

(i) *Rademacher's random variable.* Let

$$\omega_k(n) := \exp\left\{\frac{2ik\pi}{n}\right\} \quad , \quad i^2 = -1 \quad , \quad k = 0,1,2,..,n-1, \tag{6.20}$$

denote the n roots (of order n) of the unity, and define the random variable $R(n)$ (referred to as Rademacher's random variable) which takes on the values $\omega_k(n)$, $k = 0,1,2,..,n-1$ with the uniform probability $1/n$.

(ii) *n-Gaussian random variable* . Let $X(n)$ denote the real positive random variable defined by the moment generating function

$$E\{e^{-\lambda X(n)}\} = E_{1/n}\{-\lambda / K_n\} \quad , \quad \lambda \geq 0 \quad , \quad K_n := (n!)^{1/n}, \tag{6.21}$$

where $E_{1/n}$ is the Mittag Leffler function of order n (see for instance Zolotarev [6.10])

$$E_{1/n}(z) := \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(1+k/n)} \quad , \quad z \in C \cdot \tag{6.22}$$

Then the standard n-Gaussian random variable $G(n)$ is defined by the equation

$$G(n) \overset{law}{\approx} R(n)X(n) \quad . \tag{6.23}$$

(iii) *Complex-valued Brownian motion.* Loosely speaking, the *C-(fBm)*$_n$ $x_s(t,n) \in C$ is defined by the following properties (which are a generalization of Levy's well known ideas [6.6])

(Property 1). $x_s(0,n) = 0$ almost surely.

(Property 2). $x_s(t,n)$ has independent increments.

(Property 3). These increments satisfy the condition

$$x_s(t,n) - x_s(\tau,n) \overset{law}{\approx} (t-\tau)^{1/n} Z(n) \quad , \tag{6.24}$$

with

$$Z(n) := \lim_{r \to +\infty} r^{-1/n} \left(\sum_{k=1}^{r} G_k(n) \right) \quad , \tag{6.25}$$

where $\{G_k(n)\}$ is a family of standard n-Gaussian r.v.'s.

Remarks and comments. This model duplicates exactly the construction of Brownian motion (of order 2) by Levy; but its formalism is somewhat intricate.

As such, it looks like a formal generalization of which the relation with real physical problems is not transparent. For instance, what is the exact meaning of *X(n)?* When, where, and in which real practical problems are we dealing with *X(n)?*

And it is exactly the advantage of the white noise as compared with the Brownian motion. The former can be thought of as a sequence of independent random impulses, and this is very simple to picture from an engineering (or physical) standpoint. How does one extend this simple pattern in a modelling of *C(fBm)*$_n$? In the following, in the framework of engineering mathematics we shall show that by using a white noise which rotates on the finite net defined by the complex roots of the unity, it is possible to get some results in this way.

6.3. Complex-valued Gaussian white noise of even order

Firstly, we introduce the following

DEFINITION 6.1. A Rademacher white noise *R(t,n)* of order *n* is a complex valued stochastic process which satisfies the following condition: at each instant *t*, *R(t,n)* is a Rademacher r.v. of order *n*, which, moreover, is such that

$$E\{R^j(t,n)R^j(\tau,n)\} = \delta_{j,n}\delta(t-\tau), \tag{6.26}$$

where $\delta_{j,n}$ is the Kronecker delta, clearly $\delta_{j,n} = 0$ when $j \neq n$ whilst $\delta_{n,n} = 1$. ■

DEFINITION 6.2. [6.3] Let *w(t)* denote a scalar-valued Gaussian white noise with zero mean and variance $\sigma^2(t)$, and let *R(t,2k)* denote a Rademacher white noise of order *2k*. Assume that *w(t)* and *R(t,2k)* are independent. Then a model of complex Gaussian white noise of order *n*=2k, $w(t,2k) \in \Re^{1/2k}$, $\Re^{1/2k} = \{z \in C, z^{2k} \in \Re\}$, is defined by the equation

$$w(t,2k) = R(t,2k)w(t). \blacksquare \tag{6.27}$$

LEMMA 6.1. *The following relation holds, that is*

$$E\{w^j(t,2k)\} = 0, \quad j = 1,2,...,2k-1, \tag{6.28}$$

$$= \frac{(2k)!}{2^k k!}\sigma^{2k}(t). \blacksquare \tag{6.29}$$

Indeed, since *R(t,n)* and *w(t)* are independent, one can write the equality

$$E\{w^j(t,n)\} = E\{R^j(t,n)\}E\{w^j(t)\},$$

and by virtue of the properties of the complex roots of the unity, one has

$$E\{R^j(t,n)\} = \delta_{j,n}.$$

Furthermore, it is well known that $E\{w^j(t)\}$ is exactly equal to the right sides of equations (6.28) and ((6.29).

This lemma explains why we restrict ourselves to the case $n=2k$ only. Since one has $E\{w^j(t,2k+1)\}= 0$ for every $j = 1,2,3,....$

Remarks and comments . The equation (6.27), which defines $w(t,n)$ exhibits the basic difference with Sainty's modelling, see equation (6.24). Indeed $X(n)$, in equation (6.24) is positive, whilst $w(t)$ may be positive or negative. $G(n)$ is defined in $\Re_+^{1/n} = \{z \in C; z^n \in \Re\}$ and $w(t,n)$ is defined in $\Re^{1/n} = \{z \in C; z^n \in \Re\}$. This difference, in our point of view, will considerably simplify the formalism.

In the following, to shorten the notation we shall often use n instead of $2k$, whenever this will not be confusing to the reader.

6.4. Brownian motion of even order via rotating Gaussian white noises

6.4.1. Main definition

Duplicating the usual construction of Brownian motion (of order two), we introduce the following

DEFINITION 6.3. [6.3] A continuous stochastic process $\beta(t,n)$, $t > 0$, is a Brownian motion of order $n = 2k$ if

(i) $\beta(0,n) = 0$ almost surely,

(ii) $\beta(t,n)$ has stationnary independent increments, and

(iii) For every $t > 0$, the differential $d\beta(t,n)$ is given by the expression

$$d\beta(t,n) = R(t,n)w(t)(dt)^{1/n}. \quad \blacksquare \tag{6.30}$$

According to equations (6..28) and (6.29) one has

$$E\{d\beta^j(t,2k)\} = 0 \ , \quad j = 1,..,2k-1 \ , \tag{6.31}$$

$$E\{d\beta^{2k}(t,2k)\} = \frac{(2k)!}{2^k k!}\sigma^{2k}(t)dt \tag{6.32}$$

$$=: K(n)\sigma^n(t)dt \ , \quad n = 2k. \tag{6.33}$$

DEFINITION 6.4. Consider the partition

$$0 \lhd t_1 \lhd ..... \lhd t_N = t$$

and define the difference

$$\Delta\bar{\beta}(t_i,n) := R(t_i,n)w(t_i)\left(t_{i+1}-t_i\right)^{1/n} \quad , \quad n=2k, \tag{6.34}$$

then $\beta(t,n)$ is defined as the limit

$$\beta(t,n) := \lim \sum_{i=0}^{N-1} \Delta\bar{\beta}(t_i,n) \quad when \quad \sup(t_{i+1}-t_i)\downarrow 0, \tag{6.35}$$

which we shall write

$$\beta(t,n) = \int_0^t d\beta(\tau,n). \blacksquare \tag{6.36}$$

Strictly speaking, we should define the sense in which this limit is considered, i.e. almost surely or in probability or in mean square, but since, here, we shall make a formal use of this equation only, this will not be necessary.

6.4.2. Some properties of $\beta(t,2k)$

LEMMA 6.2. The following equality clearly holds,

$$E\left\{\beta^j(t,2k)\right\} = 0 \quad , \quad j=1,2,...,2k-1 \ , \tag{6.37}$$

$$= K(2k)\int_0^t \sigma^{2k}(\tau)d\tau \quad , \quad j=2k \ . \blacksquare \tag{6.38}$$

Proof. One refers to the equation (6.35), and one uses the independence of the $\Delta\bar{\beta}(t_i,n)$'s on the one hand, together with the properties (6.31-6.33) on the other..

LEMMA 6.3. *Each of the following processes is a complex Brownian motion defined in* $\Re^{1/2k}$.

$$\beta(t,n)-\beta(\tau,n) \quad , \quad t,\tau \geq 0 \qquad \text{(additivity)} \tag{6.39}$$

$$\omega_k(n)\beta(t,n) \quad , \quad t\geq 0 \quad , \quad k\in N \qquad \text{(isotropy)} \tag{6.40}$$

$$c\beta(t/c^n,n) \quad , \quad t\geq 0 \quad , \quad c \triangleright 0 \qquad \text{(scaling)} \tag{6.41}$$

$$h^{-1/n}\left[\beta(t+h\tau,n)-\beta(t,n)\right] \quad , \quad h,t,\tau \triangleright 0 \quad \text{(self-similarity)} \tag{6.42}$$

$$\omega_k(n)h^{-1/n}\left[\beta(t+h\tau)-\beta(t)\right] \quad , \quad k\in N \quad , \quad h,t,\tau \triangleright 0$$

$$\text{(isotropy and self-similarity)} \ . \blacksquare \tag{6.43}$$

Proof. (6.39) is obvious. (6.40) results from the properties of $\omega_k(n)$. (6.41) can be obtained as follows. Define

$$y(t,n) := c\beta(t/c^n),$$

then one has

$$\begin{aligned} dy(t) &= cR\left(\frac{t}{c^n},n\right)w\left(\frac{t}{c^n}\right)\left(\frac{dt}{c^n}\right)^{1/n} \\ &= R\left(\frac{t}{c^n},n\right)w\left(\frac{t}{c^n}\right)(dt)^{1/n} \end{aligned}$$

(6.42) is a result of (6.39) and (6.41), while (6.43) is a consequence of (6.39-6.41).

6.4.3. Fokker Planck equation of $\beta(t,2k)$

DEFINITION 6.5. Analogously with probability density of real-valued random variables, we introduce a weighting function $p(\beta,t)$ defined on $\Re^{1/n}$,which is the probability density of $\beta(t)$, satisfying the conditions

$$\int_{\Re^{1/n}} p(\beta,t)d\beta = 1, \tag{6.44}$$

$$\int_{\Re^{1/n}} \beta^j p(\beta,t)d\beta = m_j(t) \tag{6.45}$$

where the mathematical expectations $m_j(t)$'s are given function of time. ■

The partial differential equation satisfied by $p(\beta,t)$ is provided by the following result.

LEMMA 6.4. *The probability density $p(\beta,t)$ of $\beta(t,2k)$ is the solution of the generalized Fokker Planck equation*

$$\frac{\partial}{\partial t}p(\beta,t) = \frac{K(n)}{n!}\sigma^n(t)\frac{\partial^n}{\partial\beta^n}p(\beta,t) \quad , \quad n=2k. \blacksquare \tag{6.46}$$

Proof. (i) Firstly, one remarks that $\beta(t,n)$, $n = 2k$, satisfies the stochastic differential equation

$$d\beta(t,n) = R(t,n)w(t)(dt)^{1/n} \quad , \quad n=2k. \tag{6.47}$$

Furthermore, according to (6.31-6.33), in terms of generating moment functions, one has

$$E\left\{e^{-vd\beta(t)}\middle|\beta(t)=\beta\right\} = 1+\frac{v^n}{n!}K(n)\sigma^n(t)dt+o(dt) \quad , \quad n=2k, \tag{6.48}$$

where we assume that the term $o(dt)$ is uniform in β.

(ii) Consider the moment generating function of $\beta(t,n)$, namely,

$$\phi(v,t) = E\left\{e^{-v\beta(t,n)}\right\}$$

By differentiating under the integral , we have

$$\frac{\partial\phi}{\partial t} = \int_{\Re^{1/n}} e^{-b\beta}\frac{\partial}{\partial t}p(\beta,t)d\beta. \tag{6.49}$$

(iii) Alternatively, we can obtain another expression for $\partial\phi/\partial t$ by noticing that

$$\begin{aligned}\frac{\partial\phi}{\partial t} &= \lim_{\Delta t\downarrow 0}\frac{1}{\Delta t}\left[\phi(v,t+\Delta t)-\phi(v,t)\right]\\ &= \lim_{\Delta t\downarrow 0}\frac{1}{\Delta t}E\left\{e^{-v[\beta(t,n)+\Delta\beta(t,n)]}-e^{-v\beta(t,n)}\right\}.\end{aligned} \tag{6.50}$$

(iv) The expected value on the right side can be written

$$\begin{aligned}&E\left\{\left[e^{-v\Delta\beta(t,n)}-1\right]e^{-v\beta(t,n)}\right\}\\ &\quad= \int_{\Re^{1/n}} E\left\{\left[e^{-v\Delta\beta(t,n)}-1\right]e^{-v\beta(t,n)}\middle|\beta(t,n)=\beta\right\}p(\beta,t)d\beta\\ &\quad= \int_{\Re^{1/n}} E\left\{\left[e^{-v\Delta\beta(t,n)}-1\right]\middle|\beta(t,n)=\beta\right\}e^{-v\beta}p(\beta,t)d\beta.\end{aligned} \tag{6.51}$$

(v) Now, by using (6.48) and letting $\Delta t\downarrow 0$, we find that

$$\frac{\partial\phi}{\partial t} = \int_{\Re^{1/n}}\frac{v^n}{n!}K(n)\sigma^n(t)e^{-\beta v}p(\beta,t)d\beta, \tag{6.52}$$

and on integrating by part,

$$\frac{\partial\phi}{\partial t} = \int_{\Re^{1/n}}\frac{1}{n!}\frac{\partial^n}{\partial\beta^n}\left[K(n)\sigma^n(t)p(\beta,t)\right]d\beta. \tag{6.53}$$

Equating (6.49) and (6.53), we obtain the result.

6.5. Brownian motion of odd order and rotating semi-Gaussian white noises

6.5.1. On the modelling of Brownian motion of order two

(i) The main difference between Sainty's modelling and the present approach lies in the following remark. Sainty's model involves the standard n-Gaussian r.v.

$$G(n) = R(n)X(n) \quad , \quad X(n) \geq 0, \tag{6.54}$$

while $\beta(t,n)$ involves the r.v.

$$G_w(n) = R(n)w \quad , \quad w \in R \ . \tag{6.55}$$

This difference explains why Sainty's Brownian motion is defined for any n while $\beta(t,n)$ is defined for $n=2k$ only. Extending our modelling to the order $n=2k+1$ can be made as follows.

(ii) Firstly, one notices that the usual Brownian motion (i.e. of order two) can be generated by the stochastic differential equation (with Maruyama's notation)

$$d\beta(t,2) = R(t,2)w(t)(dt)^{1/2}, \tag{6.56}$$

as a result of $R(t,2)w(t)$ being itself is a Gaussian white noise.

Another modelling is defined by the stochastic differential equation

$$d\beta(t,2) = R(t,2)|w(t)|(dt)^{1/2}, \tag{6.57}$$

and the latter will provide a generalized model of $C\text{-}(fBm)_n$, for any $n \geq 2$.

6.5.2. Complex-valued Brownian motion of order n

The equation (6.57) provides a general modelling of $C\text{-}(fBm)_n$ as follows.

DEFINITION 6.6. The continuous stochastic process $b(t,n)$, $t > 0$, $n = 2,3,4,5,$ is a Brownian motion of order n defined in $\Re^{1/n}$ if it satisfies the following conditions:

(i) $b(0,n) = 0$ almost surely,

(ii) $b(t,n)$ has stationary independent increments,

(iii) for every $t > 0$ the differential $db(t,n)$ is given by the expression

$$db(t,n) = R(t,n)|w(t)|(dt)^{1/n} \quad , \quad n = 2,3,4,5,..., \tag{6.58}$$

where $R(t,n)$ is a Rademacher white noise of order n, and $w(t)$ is a Gaussian white noise with zero mean and variance $\sigma^2(t)$. ■

In the sequel we shall introduce the coefficient $Q(n)$ which is equal to

$$Q(n) := K(n) \quad , \quad n = 2k \tag{6.59}$$

$$:= 2^{k+1}k!/\sqrt{2\pi} \quad , \quad n = 2k+1. \tag{6.60}$$

The following results can be easily obtained by duplicating the arguments in the preceding sections.

LEMMA 6.5. (Parallel to Lemma 6.1). *One has the equalities*

$$E\{db^j(t,n)\} = 0 \quad , \quad j = 1,..,n-1, \tag{6.61}$$

$$E\{db^n(t,n)\} = Q(n)\sigma^n(t)dt \; . \; \blacksquare \tag{6.62}$$

LEMMA 6.6. *Similar to Lemma 6.2, but stated in terms of* b(t,n). ■

6.6. Random walk with radial steps in the complex plane

Assume that at each instant 0,1,2,... the random variable $\tilde{R}_k(n) \in C$ takes on the values $\omega_0(n), \omega_1(n), ..., \omega_{n-1}(n)$, (see equation (6.20)), with the respective probabilities $p_0, p_1, ..., p_{n-1}$, $p_0 + p_1 + ... + p_{n-1} = 1$. Given a randomly selected value $\omega_j(n)$ for $\tilde{R}_k(n)$, we consider the step length $|\Delta z|$, $z \in C$ in the direction defined by $\omega_j(n)$. For each time instant k, we define the random variable

$$\tilde{w}_k := \begin{cases} +|\Delta z| \quad , \quad pr\{w_k = +|\Delta z|\} = p, \\ -|\Delta z| \quad , \quad pr\{w_k = -|\Delta z|\} = q, \end{cases} \tag{6.63}$$

with $p+q = 1$.

The random variable $\tilde{R}_k(n)\tilde{w}_k(n)$ defines a random walk in the complex plane, and assuming that we start from the origin $w_O = 0$, the position of the moving point $z = x + iy$ at the instant j is given by the expression

$$\tilde{z}_j = \sum_{k=0}^{j-1} \tilde{R}_k \tilde{w}_k \quad , \quad j = 1,2,\ldots, \tag{6.64}$$

which is merely the solution of the difference equation

$$\tilde{z}_{j+1} = \tilde{z}_j + \tilde{R}_j \tilde{w}_j . \tag{6.65}$$

In the following we shall consider the special case where $p_O = p_1 = \ldots = p_{n-1} = 1/n$ and $p = q = 1/2$, to then write $R_k(n)$, w_k and z_j, and

$$z_{j+1} = z_j + R_j w_j \quad . \tag{6.66}$$

If v denotes the variable of Fourier's transform, then the characteristic function of z_j is given by the expression

$$\begin{aligned} \varphi_{z_j}(v) &:= E\left\{e^{ivz_j}\right\} \\ &= \left[\frac{1}{2n}\sum_{k=0}^{n-1}\left(\exp\{iv\omega_k|\Delta z|\} + \exp\{-iv\omega_k|\Delta z|\}\right)\right]^j , \end{aligned} \tag{6.67}$$

and one can then state the following lemma:

LEMMA 6.7. *For small* $|\Delta z|$ *one has the equivalence*

$$\varphi_{z_j}(v) = \left(1 + \frac{1}{n!}(iv)^n |\Delta z|^n\right)^j + o\left(|\Delta z|^{2nj}\right) \quad , \tag{6.68}$$

where $o(.)$ *denotes Landau's symbol..*■

Proof. One expands the exponentials in Taylor series and one then uses the properties of the complex roots of the unity.

6.7. Brownian motion of order n via random walk in the complex plane

6.7.1. Derivation of the model

One then has the following results:

PROPOSITION 6.1. [6.4,6.5]. *The limit of the complex random walk* zj *described by equation* (6.66) *defines a stochastic process of which the probability density* $p(z,t)$ *is solution of the heat equation of order n*

$$\frac{\partial p(z,t)}{\partial t} = (-1)^n \frac{\sigma^n}{n!} \frac{\partial^n p(z,t)}{\partial z^n} \quad , \quad z \in C \, , \tag{6.69}$$

where σ *denotes a positive constant..*■

Proof. (i) We have to take the limit of (6.68) as $|\Delta z| \downarrow 0$, and to this end,we use the standard approach. We define $j = t/\Delta t$ to write

$$\varphi_{z_j}(v) = \exp\left\{\frac{(iv)^n}{n!}\frac{|\Delta z|^n}{\Delta t}t\right\} . \tag{6.70}$$

In order to have a sensible result, we shall require that

$$\frac{|\Delta z|^n}{\Delta t} \to \sigma^n \qquad as \qquad \Delta t \downarrow 0,$$

to have

$$\varphi_{z_j}(v) \to \varphi_z(v),$$

with

$$\varphi_z(v) = \exp\left\{\frac{(iv)^n}{n!}\sigma^n t\right\} . \tag{6.71}$$

(iii) We then obtain the density

$$p(z,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left\{-ivz + \frac{(iv)^n}{n!}\sigma^n t\right\} dv \tag{6.72}$$

which satisfies the equation (6.69).

6.7.2. Relation with the approach via rotating Gaussian white noise

Define the coefficient $K(n)$ by the expressions

$$K(n) = \frac{(2k)!}{2^k k!} \quad , \quad n = 2k \quad , \tag{6.73}$$

$$= \frac{2^{k+1} k!}{\sqrt{2\pi}} \quad , \quad n = 2k+1, \tag{6.74}$$

then one can show (the simple proof is left to the reader) that the probability density $p(b,t)$ of $b(t,n)$ as defined by equation (6.72) is solution of the partial differential equation

$$\frac{\partial p}{\partial t}(b,t) = (-1)^n \frac{K(n)}{n!} \sigma^n \frac{\partial^n p}{\partial b^n}(b,t), \tag{6.75}$$

where σ is the standard deviation of $w(t)$.

Equations (6.69) and (6.75) are the same, given the scaling coefficient $K(n)$, and as a result one can claim that the processes $b(t,n)$ defined as the solution of the equation (6.30) one the one hand, and as the limit of the complex random walk (6.65) on the other, are the same in law.

Convention and notation. In the following we shall assume that the characteristic function $\varphi_b(v)$ of the *C-(fBm)*$_n$ $b(t,n)$ is defined by the expression

$$\varphi_b(v) = \exp\left\{(iv)^n \frac{K(n)}{n!} \sigma^n t\right\}, \tag{6.76}$$

which will provide the mathematical expectation

$$E\left\{b^{nk}(t,n)\right\} = \frac{K^k(n)}{(n!)^k} \sigma^{nk} t^k \, . \tag{6.77}$$

This amount to take the limit of the complex random walk in Section 6.6, in assuming that

$$|\Delta z|^n / \Delta t \;\rightarrow\; K(n)\sigma^n \quad as \quad \Delta t \downarrow 0.$$

As such, equation (6.76) can be thought of as a unifying model which comprises equation (6.30) on the one hand, and the approach via complex random walk on the other.

6.8. Fractional Brownian motion with complex variance

6.8.1. Derivation of the model

Assume that, in the definition of the random walk in section 6.6., instead of referring to the step length $|\Delta z|$, we consider the complex step $\Delta z := \Delta x + i\,\Delta y$ itself to define the random variable

$$w_k' = \begin{cases} +\Delta z \;, & pr\{w_k = \Delta z\} = 1/2, \\ -\Delta z \;, & pr\{w_k = -\Delta z\} = 1/2. \end{cases} \tag{6.78}$$

A direct duplication of the above calculus then provides the characteristic function of the sum z_j (see equation (6,66)) in the form

$$\varphi_{z_j}(v) = \left(1 + \frac{1}{n!}(iv)^n(\Delta z)^n\right)^j + o\left(|\Delta z|^{2nj}\right), \tag{6.79}$$

and in order to get a sensible result when taking the limit with $|\Delta z| \downarrow 0$, we shall need to assume that one has the limiting equality

$$\frac{(\Delta z)^n}{\Delta t} \;\rightarrow\; \sigma^n \qquad as \qquad \Delta t \downarrow 0, \tag{6.80}$$

where σ is now a complex-valued parameter, say

$$\sigma := \sigma_x + i\,\sigma_y, \tag{6.81}$$

$$= |\sigma| \exp\{i\theta\} \tag{6.82}$$

We thus obtain, in a quite straightforward manner, the parallel of Proposition 6.1, which now reads:

PROPOSITION 6.2. *The limit of the random walk described by the equation*

$$z_{j+1} = z_j + R_j w'_j, \tag{6.83}$$

in the complex plane, defines a complex-valued stochastic process b(t,n) of which the probability density p(b,t) is the solution of the heat equation (6.75) *in which* σ *is defined by the equations* (6.81), (6.82). ■

6.8.2. A simplified random walk

It appears that the random walk which we have so introduced to define the *C-(fBm)*$_n$ can be simplified and this remark is of paramount importance since it enlightens the practical significance of the model.

(i) *C-(fBm)*$_n$ *with real-valued nth-moment*

Assume that we replace the equation (6.66) which defines the random walk by the new equation

$$z_{j+1} = z_j + R_j|\Delta z|, \tag{6.84}$$

its characteristic function is then

$$\varphi_{z_j}(v) = \left[\frac{1}{n}\sum_{k=0}^{n-1}\exp\{iv\omega_k|\Delta z|\}\right]^j, \tag{6.85}$$

and a simple calculation shows that the proposition 6.1 still holds. In other words, this random walk, which is simpler than the preceding one, gives the same result.

(ii) *C-(fBm)*$_n$ *with complex-valued n-th moment*

The same remark applies to the generating of *C-(fBm)*$_n$ with complex-valued *n*-th moment. Here, instead of equation (6.84), we shall use the model

$$z_{j+1} = z_j + R_j \Delta z \quad , \tag{6.86}$$

and we shall obtain exactly the same results.

In the next section, we shall try to guess the mathematical expression of the probability density of *b(dt,n)*, where *dt* is a small increment of time. The interest of this

derivation is to provide results on the probability transition of $b(t,n)$, that is to say, the probability density of $b(t+dt,n)-b(t,n)$.

6.9. Determination of the probability density of $b(dt,n)$

6.9.1. Generalized semi-Gaussian density of order n

DEFINITION 6.7. The positive real-valued random variable X is said to be a generalized semi-Gaaussian random variable of order n, $n = 1,2,...,$ when its probability is

$$f_n(x) = \frac{1}{n^{1/n}\sigma\Gamma\left(1+\frac{1}{n}\right)}\exp\left\{-\frac{x^n}{n\sigma^n}\right\} \quad , \quad \sigma \triangleright 0 \ .\blacksquare \tag{6.87}$$

A simple calculation yields the expectation

$$E\{X^j\} = \frac{n^{j/n}\sigma^j\Gamma\left(\frac{j+1}{n}\right)}{\Gamma(1/n)} \quad , \quad j = 0,1,2,.. \tag{6.88}$$

hence the expected result

$$E\{X^n\} = \sigma^n. \tag{6.89}$$

This probability density can be obtained by a routine calculation, starting from the density $f_n(x) = K\exp\{-\alpha x^n/\sigma^n\}$ and determining the constant K and α in order to have the equality $E\{X^n\} = \sigma^n$. When $n=2$, one has

$$f_2(x) = \frac{2}{\sigma\sqrt{\pi}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \tag{6.90}$$

and in this way, $f_n(x)$ can be thought of as a generalization of the semi-normal density. Let us point out that $f_n(x)$ is also the probability density which maximixes the Shannon entropy of X given the values of their first n-1 moments.

6.9.2. General expression of the probability density of b(dt,n)

Our main purpose, in the present section, is to obtain the general expression of the probability density $p_n(\beta,dt)$ of *b(dt,n);* and in this way we have the following

LEMMA 6.8. *The probability density of b(dt,n) can be written in the general form*

$$p_n(\beta,dt) = \frac{1}{n}\sum_{k=0}^{n-1}\frac{1}{\omega_k(n)\sigma(dt)^{1/n}}\, g\left(\frac{\beta}{\omega_k(n)\sigma(dt)^{1/n}}\right)_{\beta\in D_k}, \tag{6.91}$$

where the function $g(x)$, $x \geq 0$, $g(x) \geq 0$, *satisfies the condition*

$$\int_0^\infty g(x)dx = 1. \blacksquare \tag{6.92}$$

Proof. (i) Define $\Re_+^{1/n} := \{D_1, D_2, ..., D_n\}$ where D_k is the semi-axis defined as $\{z \in D_k, z = \omega_k\xi, \xi \in \Re_+\}$. On a general standpoint, one can write $p_n(\beta,dt)$ in the form

$$p_n(\beta,dt) = \sum_{k=0}^{n-1} r_k g_k(\beta,dt), \tag{6.93}$$

where $g_k(\beta,dt)$ is a probability density defined on D_k, and where r_k is the probability with which D_k is selected. Here, one has $r_k = 1/n$, $k = 0,1,..,n-1$.

(ii) According to the normalizing condition

$$\int_{D_k} g_k(\beta,dt)d\beta = 1,$$

one can write as well

$$\int_0^\infty \omega_k(n) g_k(\omega_k\xi,dt)d\xi = 1,$$

and owing to the geometrical symmetry of the problem it follows that

$$\omega_k g_k(\omega_k\xi,dt) = g(\xi) \quad , \quad \xi \in R_+ \quad , \quad k = 0,1,..,n-1.$$

Clearly, one should have the equality

$$g_k(\beta,dt) = \frac{1}{\omega_k} g\left(\frac{\beta}{\omega_k}\right). \tag{6.94}$$

(iii) Normalizing with the standard deviation $\sigma(dt)^{1/n}$ provides the result.

Example 1. According to the model of $C\text{-}(fBm)_n$ in the form of a Gaussian white noise which rotates on the grid defined by the complex roots of the unity, everything looks like as if we were using a Gaussian white noise on each D_k. According to the calculation above, we shall have the density

$$p_n(\beta,dt) = \frac{1}{n}\sum_{k=0}^{n-1} \frac{2}{\sqrt{2\pi}\,\omega_k(n)\sigma(dt)^{1/n}} \exp\left\{-\frac{\beta^2}{2\omega_k^2(n)\sigma^2(dt)^{2/n}}\right\}_{\beta\in D_k}, \tag{6.95}$$

.....

therefore the mathematical expectation (see equation (6.88))

$$E\left\{b^j(dt,n)\right\} = \frac{2^{j/2}\Gamma((j+1)/2)}{n\sqrt{\pi}}\,\sigma^j\left(\sum_{k=0}^{n-1}\omega_k^j(n)\right)(dt)^{j/n}, \tag{6.96}$$

with

$$\sum_{k=0}^{n-1}\omega_k^j(n) = 0, \quad j=1,\ldots,n-2, \tag{6.97}$$

$$= n, \quad j=0,n. \tag{6.98}$$

6.9.3. C-(fBm)$_n$ and semi-Gaussian density of order n

Assume now that we add the following assumption:

(H) The function $g(.)$ in (6.94) is formally independent of $\omega_k(n)$, clearly one has the equality

$$g\left(\beta/\omega_k(n)\sigma(dt)^{1/n}\right) = g\left(\beta/(dt)^{1/n}\right). \tag{6.99}$$

In some sense, this assumption could be thought of as illustrating the circular symmetry of the problem. The function $g(\beta)$ is the same for all k, and the choice of D_k is defined only by the corresponding weighting coefficient.

Under this assumption, we have the general expression

$$p_n(\beta,dt) = \frac{1}{n}\sum_{k=0}^{n-1}\frac{1}{\omega_k(n)\sigma(dt)^{1/n}}\, g\left(\frac{\beta^n}{\sigma^n dt}\right)_{\beta\in D_k} \tag{6.100}$$

Example 2. If we require that by formally making n=2 in ((6.100), we should get the expression (6.95), we are led to refer to the semi-Gaussian density of order n (see equation (6.87)), and to define the corresponding density in the form

$$q_n(\beta,dt) = \frac{1}{n}\sum_{k=0}^{n-1}\frac{1}{n^{1/n}\Gamma\left(1+\frac{1}{n}\right)\omega_k(n)\sigma(dt)^{1/n}}\exp\left\{-\frac{\beta^n}{n\sigma^n dt}\right\}_{\beta\in D_k} . \tag{6.101}$$

Here again, according to (6.88) one has the equality

$$E\left\{b^j(dt,n)\right\} = \frac{n^{j/n}\Gamma\left((j+1)/n\right)}{n\Gamma(1/n)}\sigma^j\left(\sum_{k=0}^{n-1}\omega_k^j(n)\right)(dt)^{j/n} \tag{6.102}$$

therefore the n-th moment

$$E\left\{b^n(dt,n)\right\} = \sigma^n dt \ . \tag{6.103}$$

To summarize, in the present section in referring to the physical nature of the C-$(fBm)_n$ process, we have obtained candidate mathematical expressions for the probability density of its increment, and mainly we have obtained densities involving the exponential functions $\exp(z^2)$ and $\exp(z^n)$. The fact that we have several possible probability densities (and not only one!) should not be disturbing at all. This merely illustrates the property that the heat equation has different solutions, depending upon the initial condition which is selected. This might also explain why Sainty's model can be expanded in a quite different way by using the Mittag Leffler function. There remains to relate this probability density to the density of $b(t,n)$ itself, and to this end we shall need the Chapman Kolmogorov equation.

are led in quite a natural way to ask ourselves whether the preceding results, that is to say the above probability densities of the transition *db(t,n),* are consistent with the heat equation.

6.10. On the fractal nature of the heat equation of order n

6.10.1. Preliminary background

In Section 6.7, we have shown that the probability density $p(z,t)$ of the C-$(fBm)_n$ satisfies the heat equation of order n, which we re-write here for convenience

$$\frac{\partial p}{\partial t}(z,t) = (-1)^n \frac{\sigma^n}{n!} \frac{\partial^n p}{\partial z^n}(z,t) , \tag{6.104}$$

and the question which then arises is the following: Can we describe, how to establish a link between the solution of this equation and the result of the preceding section? Is the solution of equation (6.104) in the form $\exp(z^2)$ or $\exp(z^n)$?

We shall show that this solution may have either expression above, depending upon the initial and boundary conditions, and to this end we shall begin with a special case.

6.10.2. Analysis of a special case

In order to get more insight in the problem, we firstly consider the differential equation

$$p(z) = \frac{\sigma^n}{n!} \frac{\partial^n p}{\partial z^n}(z) , \quad z \in C \tag{6.105}$$

where σ a real-valued number which may positive or negative.

Derivation of $\exp\{z^n\}$ solution

We select the following initial conditions:

$$p(0) = 1, \tag{6.106}$$

$$\left[\frac{\partial^j p(z)}{\partial z^j}\right]_{z=0} = 0 , \quad j = 1,...,n-1. \tag{6.107}$$

The condition (6.106) is written analogously with the stochastic initial condition $pr\{z(0)=0\} = 1$. The condition (6.107) is suggested by the circular symmetry of the problem.

The characteristic roots r_k, $k = 1,\ldots,n$, of the equation (6.105) are defined by the equation

$$1 = \frac{\sigma^n}{n!} r^n ,$$

therefore

$$r_k = \frac{(n!)^{1/n}}{\sigma} \omega_k(n) \quad , \quad k = 0,1,\ldots,n-1, \tag{6.108}$$

and the general solution

$$p(z) = \sum_{k=0}^{n-1} \exp\left\{ \frac{(n!)^{1/n}}{\sigma} \omega_k(n) z \right\}, \tag{6.109}$$

where the constants λ_k, $k = 0,..,n\text{-}1$ are defined by the initial conditions (6.106) and (6.107). A simple calculation yields

$$p(z) = \frac{1}{n} \sum_{k=0}^{n-1} \exp\left\{ \frac{(n!)^{1/n}}{\sigma} \omega_k(n) z \right\} . \tag{6.110}$$

and the following results is simple to obtain.

LEMMA 6.9. *p(z) in equation* (6.110) *can be re-written in the form*

$$p(z) = \frac{1}{n} \left[(n-1) + \exp\left\{ \frac{z^n}{\sigma^n} \right\} \right] . \blacksquare \tag{6.111}$$

Proof One expands the exponential in (6.110) to have

$$p(z) = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{j=0}^{\infty} \frac{(n!)^{j/n}}{\sigma^j} \omega_k^j(n) z^j , \tag{6.112}$$

and then one uses the equality

$$\sum_{k=0}^{n-1} \exp\left\{ i \frac{2jk\pi}{n} \right\} = \frac{1 - e^{i(2j\pi)}}{1 - e^{i(2j\pi/n)}}. \tag{6.113}$$

Derivation of $\exp\{z^2\}$ *solution*

In order to exhibit a complete and meaningful relation beetween the results of the preceding section (6.10) and the heat equation of order n, we must show that the equation (6.105) can have also an $\exp\{z^2\}$-solution. This can be done as follows.

Assume that the coefficients λ_k , $k = 0,..,n\text{-}1$ in (5.6) are selected so as to satisfy the conditions

$$\sum_{k=0}^{n-1} \lambda_k = 1, \tag{6.114}$$

$$\sum_{k=0}^{n-1} \lambda_k \omega_k^{2j-1} = 0 \quad , \quad 2j \le n-1, \tag{6.115}$$

$$\sum_{k=0}^{n-1} \lambda_k \omega_k^{2j} = 1 \quad , \quad 2j \le n-1 \cdot \tag{6.116}$$

Then, substituting these equations into the expression (6.109), namely

$$p(z) = \sum_{j=0}^{\infty} \frac{(n!)^{j/n}}{\sigma^j} \frac{z^j}{j!} \left(\sum_{k=0}^{n-1} \lambda_k \omega_k^j \right), \tag{6.117}$$

yields the result.

6.10.3. Application to the heat equation

We now consider the equation

$$\frac{\partial p}{\partial t}(z,t) = \frac{\sigma^n}{n!} \frac{\partial^n p}{\partial z^n}(z,t), \tag{6.118}$$

where σ is still a positive or negative real number. As usual we seek a solution in the form

$$p(z,t) = e^{st} f(z), \tag{6.119}$$

and in substituting into (6.118), we find that $f(z)$ is solution of the equation

$$f(z) = \left(\frac{\sigma^n}{s}\right)\frac{1}{n!}\frac{\partial^n f(z)}{\partial z^n}, \tag{6.120}$$

that is to say equation (6.105).

Clearly, the solution of the heat equation is of a fractal nature, and it is this remark which is the point of departure of Hochberg's paper on the sign measure of probability [6.2]. He considered the equation of even order (for reasons of convergence)

$$\frac{\partial}{\partial t} p(x,t) = (-1)^{n+1}\frac{\partial^{2n}}{\partial x^{2n}} p(x,t), \tag{6.121}$$

and analyzed the fractal properties of its solution $p(x,t)$.

6.11. Concluding remarks

In the present chapter we have shown how one can obtain various models of complex-valued Brownian motions by using the complex roots of unity. In the first approach we have used a family of Gaussian white noises indexed by the complex roots of unity, and then, in a quite natural way, we have replaced the white noise by a constant step, and so we arrived at a model involving a random walk in the complex plane. A straightforward and meaningful generalization of the one-dimensional random walk!

Here we have restricted ourselves to the construction of this new process only, and in the next chapter, we shall derive all its basic properties which will define the framework for a new stochastic calculus of order n .

References

6.1 COX, D.R.; Long-range dependence, non-linearity and time irreversibility, *J. Time Series Analysis,* Vol 12, No 4, pp 329-335, 1991

6.2 HOCHBERG, G.; Signed measure of probability, *Ann. of Probability,* Vol 6, No 3, pp433-458, 1978

6.3 JUMARIE, G.; A new approach to complex-valued fractional Brownian motion via rotating white noise, *Chaos, Solitons and Fractals*, Vol 9, No 6, pp 881-893, 1998

6.4 JUMARIE, G.; A new approach to fractional Brownian motion of order n via random walk in the complex plane, *Chaos, Solitons and Fractals*, Vol 10, No 7, pp 1193-1212, 1999
6.5 JUMARIE, G.; Complex-valued measures. An approach via random walk in the complex plane, *Statistics and Probab. Letters*, Vol 42, No 1, pp 61-67, 1999
6.6 LEVY, P.; *Stochastic Processes and Brownian Motion* (in French), Gauthier Villars, Paris, 1965
6.7 MANDELBROT, B.B.; and van NESS, J.W.; Fractional Brownian motion, fractional noises and applications, *SIAM Rev.*, Vol 10, No 4, pp 422-437, 1968
6.8 MANDELBROT, B.B.; *The Fractal Geometry of Nature*, W,H, Freeman, New York, 1982; reprint, Plenum Publishing, New York, 1992
6.9 SAINTY, P.; Construction of a complex valued fractional Brownian motion of order N, *J. Math. Phys.*, Vol 33, No 9, pp 3128-3149, 1992
6.10 ZOLOTAREV, Y.; *Transactions of Mathematical Monographs*, Vol 65, Amer. Mathematical Society, Providence, 1986

Chapter 7

Complex-Valued Fractional Brownian Motion of Order n. Part II

There is nothing constant in the universe. All ebb and flow, and every form that is born bears in its womb the seeds of change.

Ovid

7.1. Introduction

In the preceding chapter we have constructed a model of $C\text{-}(fBm)_n$ as the limit of random walks in the complex plane, and here we shall expand the theory with two objectives in mind. On the one hand, we shall put in evidence some special features of this new stochastic process, and on the other hand we shall define a new framework for an Itô stochastic calculus of order n. But before we give a summary of the main formulae which we shall need to this end.

Convention and notation. In the following, we shall refer to the $C\text{-}(fBm)_n$'s $b(t,n)$ and $\beta(t,n)$ by using the same new generic process $b(t,n)$ of which the characteristic function $\varphi_b(v) \equiv \varphi(v)$ is defined by the expression

$$\varphi(v) = \exp\left\{(iv)^n \frac{\sigma^n}{n!} t\right\}, \tag{7.1}$$

therefore the mathematical expectation

$$E\left\{b^{nk}(t,n)\right\} = \frac{\sigma^{nk}}{(n!)^k} t^k \ . \tag{7.2}$$

Probability densities. The probability density $p_n(b,t)$ defined by (7.1) is given by the expression

$$p_n(b,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left\{-ivb + \frac{(iv)^n}{n!}\sigma^n t\right\} dv \ , \tag{7.3}$$

which satisfies the partial differential equation

G. Jumarie, *Maximum Entropy, Information Without Probability and Complex Fractals*
© Springer Science+Business Media Dordrecht 2000

$$\frac{\partial}{\partial t} p_n(b,t) = (-1)^n \frac{\sigma^n}{n!} \frac{\partial^n}{\partial b^n} p_n(b,t). \tag{7.4}$$

The expression of $p_n(b,t)$ depends upon the value of n, and one has

$$p_n(b,t) = \begin{cases} \frac{1}{\pi}\int_0^{\infty} \exp\left\{-\sigma^n \frac{v^n}{n!} t\right\} \cos(vb)dv\ , \\ \quad .... \qquad\qquad n = 2(2k+1)\ , b \in C \qquad (7.5) \\ \frac{1}{\pi}\int_0^{+\infty} \cos\left(vb - \sigma^n \frac{v^n}{n!} t\right) dv\ , \\ \qquad\qquad n = 2(2k)+1\ , b \in C \qquad (7.6) \\ \frac{1}{\pi}\int_0^{\infty} \cos\left(vb + \sigma^n \frac{v^n}{n!} t\right) dv. \\ \qquad\qquad n = 2(2k+1)+1\ ,\ b \in C \qquad (7.7) \end{cases}$$

Equation (7.5) is close to Hochberg's density [7.2] which reads

$$\begin{aligned} p_H(b,t) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left\{-ib\xi - \frac{\xi^{2m}}{(2m)!} t\right\} d\xi \\ &= \frac{1}{\pi}\int_0^{\infty} \exp\left\{-\frac{\xi^{2m}}{(2m)!} t\right\} \cos b\xi\, d\xi, \end{aligned} \tag{7.8}$$

but while (7.8) holds for both $n = 2k$ and $n = 2k+1$, it is not so for (7.5). This discrepancy should not be understood as an inconsistency, but rather is merely a consequence of the term $(iv)^n$ in (7.3). The presence of $(i)^n$ is owed to the modelling via the complex random walk, whilst the density (7.8) is formally selected as a solution of the heat equation.

Sainty's density [7.10] is

$$p_S(b,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left\{-ib\xi + (-i)^n \frac{\xi^n}{n!} t\right\} d\xi\ , \tag{7.9}$$

(7.3) and (7.9) are equal when $n = 2k$, but are different whenever $n = 2k+1$. In other words, the complex random walk provides Sainty's model only when n is even. This difference is pictured by the corresponding heat equations which are respectively

$$\frac{\partial p_n}{\partial t}(b,t) = \frac{(-1)^n}{n!}\frac{\partial^n p_n}{\partial b^n}(b,t), \tag{7.10}$$

and

$$\frac{\partial p_S}{\partial t}(b,t) = \frac{1}{n!}\frac{\partial^n p_S}{\partial b^n}(b,t). \tag{7.11}$$

7.2. Stability in law of complex-valued Brownian motion of order *n*

7.2.1. Preliminary background

DEFINITION 7.1. If X, X_1 and X_2 are three normalized (i.e., with zero mean and unit standard deviation) independent real-valued random variables which have the same probability density; then we shall say that this density is L-stable (stable in the Levy sense) if there exists three scaling factor s, s_1 and s_2 such that

$$sX = s_1X_1 + s_2X_2, \tag{7.12}$$

where $s = s(s_1,s_2)$ is a function of s_1 and s_2. ■

According to Cauchy, the symmetrical distributions which satisfy this equation are such that the following condition is satisfied

$$s^{\upsilon} = (s_1)^{\upsilon} + (s_2)^{\upsilon}, \tag{7.13}$$

and Levy showed that it is sufficient and necessary that

$$0 \lhd \upsilon \leq 2. \tag{7.14}$$

The corresponding probability density is defined by the equation

$$p(x) = \frac{1}{\pi}\int_0^{\infty} \exp\{-\xi^{\upsilon}\}\cos(\xi x)d\xi\ , \tag{7.15}$$

with the companion characteristic function

$$\phi_x(v) = \exp\{-|v|^{\upsilon}\}. \tag{7.16}$$

7.2.2. Application to C-Brownian motion of order n

One has the following

LEMMA 7.1. *The probability distribution of the complex-valued fractional Brownian motion of order n is L-stable.* ■

Proof. It is sufficient to show that if $b_1(t,n)$ and $b_2(t,n)$ are two independent *C-(fBm)*$_n$, their sum $Y := b_1(t,n)+b_2(t,n)$ is also a *C-(fBm)*$_n$.

Indeed, according th the equation (7.1), the characteristic function of Y is

$$\begin{aligned} E\{\exp(ivY)\} &= E\{\exp(ivb_1)\}E\{\exp(ivb_2)\} \\ &= \exp\left\{(iv)^n \frac{\sigma_1^n}{n!}t\right\}\exp\left\{(iv)^n \frac{\sigma_2^n}{n!}t\right\} \\ &= \exp\left\{(iv)^n \frac{1}{n!}(\sigma_1^n+\sigma_2^n)t\right\}. \end{aligned} \tag{7.17}$$

Clearly, $Y(t)$ is a *C-(fBm)*$_n$ with the parameter

$$\sigma^n = \sigma_1^n + \sigma_2^n. \tag{7.18}$$

Further remarks and comments .

(i) Formally, if we assume that υ in equ. (7.13) can be larger than 2, then the characteristic functions $\phi_z(v,n)$ (equ. (7.1)) and $\phi_x(v)$ (equ. (7.16)), are equal when $n = 2(2k+1)$,

$$\phi_x(v) = \phi_z(v,2(2k+1)) , \tag{7.19}$$

in other words, we thus obtain an extension of Levy's result, which applies when $\upsilon > 2$. The condition $0 < \upsilon \leq 2$ is necessary to ensures that the probability density $p(x)$, in equ. (7.15), is a positive function. When $\upsilon > 2$, then $p(x)$ may be positive or negative and we come across signed measure of probability. But do not forget that $b(t,n)$ is complex-valued!!

(ii) For Gaussian r.v.'s, the scaling equation (7.13) is $\sigma^2 = \sigma_1^2 + \sigma_2^2$. In other words (7.18) can be considered as an extension which applies to rotating white noise (we recall that, in our first model, we defined *(fBm)*$_n$ as rotating Gaussian white noise).

(iii) According to the equation (7.18) the (usual) central limit theorem, that is to say the Lindeberg's theorem, does not apply for the *C-(fBm)*$_n$, since then the sum Y,

$$Y = \sum_{j=1}^{\infty} b_j(t,n) , \tag{7.20}$$

will define a *C-(fBm)n* whenever one has the equality

$$\sum_{j=1}^{\infty} \sigma_j^n \lhd \infty. \tag{7.21}$$

As a matter of fact, we have a central limit of order n, or again a central limit theorem in $\Re^{1/n}$, and in the next section, we shall consider this question.

7.3. Central limit theorems of order n

THEOREM 7.1. *(Weak central limit theorem). Let* $x_1, x_2, ..., x_n, ..$ *denote an infinite set of random variables which are mutualy independent in* $\Re^{1/n}$, *with the same distribution. Assume*

$$E\{x_j^q\} = 0 , \quad q = 1,2,..,n-1, \tag{7.22}$$

$$E\{x_j^n\} = \sigma^n . \tag{7.23}$$

Define

$$S_k := x_1 + x_2 + .. + x_k , \tag{7.24}$$

and

$$S_k^* := S_k / \sigma(k)^{1/n} . \tag{7.25}$$

Then as $k \uparrow \infty$, *the distribution of* S_k^* *tends to the distribution defined by the characteristic function* $i \exp\{(iv)^n / n!\}$. ■

Proof. According to the properties (7.22) and (7.23), we can approximate the characteristic function of xj by the n-order Taylor expansion

$$\phi_{x_j}(v) = 1 + \frac{(iv)^n}{n!}\sigma^n + o(v^{n+1}), \tag{7.26}$$

where *o(.)* denotes Landau's symbol. From (7.26) we have

$$\phi_{x_j/\sigma(k)^{1/n}}(v) = 1+\frac{(iv)^n}{n!k}+o\left(\frac{v^{n+1}}{k^{1+(1/n)}}\right), \tag{7.27}$$

therefore

$$\begin{aligned}\phi_{S_k^*}(v) &= \left[1+\frac{(iv)^n}{n!k}\right]^k + k\,o\left(\frac{v^{n+1}}{k^{1+(1/n)}}\right)\\ &= \left[1+\frac{(1v)^n}{n!k}\right]^k + o\left(\frac{v^{n+1}}{k^{1/n}}\right).\end{aligned} \tag{7.28}$$

Taking the limit as $k \uparrow \infty$ yields the result.

THEOREM 7.2. *(Mixing of distributions) . Let* $x_1, x_2, ..., x_n, ...$ *denote an infinite set of random variables which are mutually independent in* $\Re^{1/n}$. *Assume*

$$E\{x_j^q\} = 0 \quad , \quad q = 1,2,..,n-1, \tag{7.29}$$

$$E\{x_j^n\} = \sigma_j^n. \tag{7.30}$$

Define

$$S_k := x_1 + x_2 + .. + x_k, \tag{7.31}$$

$$s_k^n = \sigma_1^n + \sigma_2^n + ... + \sigma_k^n, \tag{7.32}$$

and

$$S_k^* := S_k / s_k \ . \tag{7.33}$$

Assume that for any given small $\varepsilon > 0$ *and k sufficiently large, one has*

$$\frac{1}{k} - \varepsilon \le \left(\frac{\sigma_j}{s_k}\right)^n \le \frac{1}{k} + \varepsilon \quad , \quad j = 1,2,..,k \tag{7.34}$$

Then as $k \uparrow \infty$ *the distribution of* S_k^* *tends to the distribution defined by the characteristic function* $exp\{(iv)^n/n!\}$. ■

Proof (i) Beginning as in the proof of Theorem 7.1 we have the characteristic functions

$$\phi_{x_j}(v) = 1+\frac{(iv)^n}{n!}\sigma_j{}^n + o(v^{n+1}), \tag{7.35}$$

$$\phi_{x_j/s_k}(v) = 1+\frac{(iv)^n}{n!}\left(\frac{\sigma_j}{s_k}\right)^n + o\left[\left(\frac{v}{s_k}\right)^{n+1}\right], \tag{7.36}$$

therefore

$$\phi_{S_k^*}(v) = \prod_{j=1}^{k}\left[1+\frac{(iv)^n}{n!}\left(\frac{\sigma_j}{s_k}\right)^n\right]. \tag{7.37}$$

(ii) Now, we consider the difference

$$(\Delta\phi)_k := \phi_{S_k^*}(v) - \exp\left\{\frac{(iv)^n}{n!}\right\} \; ; \tag{7.38}$$

and in order to obtain its limit as $k \uparrow \infty$, we refer to the logarithms difference

$$\begin{aligned}(\Delta \ln)_k &:= \ln\phi_{S_k^*}(v) - \frac{(iv)^n}{n!} \\ &= \sum_{j=1}^{k}\ln\left[1+\frac{(iv)^n}{n!}\left(\frac{\sigma_j}{s_k}\right)^n\right] - \frac{(iv)^n}{n!} \\ &\cong \sum_{j=1}^{k}\frac{(iv)^n}{n!}\left(\frac{\sigma_j}{s_k}\right)^n - \frac{1}{2}\frac{(iv)^{2n}}{(n!)^2}\left(\frac{\sigma_j}{s_k}\right)^{2n} - \frac{(iv)^n}{n!} + o\left[\left(\frac{\sigma_j}{s_k}\right)^{3n}\right],\end{aligned} \tag{7.39}$$

which provides, by virtue of the equality (7.32),

$$(\Delta \ln)_k \cong -\frac{1}{2}\frac{(iv)^{2n}}{(n!)^2}\sum_{j=1}^{k}\left(\frac{\sigma_j}{s_k}\right)^{2n} + o\left[\left(\frac{\sigma_j}{s_k}\right)^{3n}\right].$$

(iii) This being the case, according to (7.32), one has the inequality

$$\frac{1}{k^2}-\frac{2\varepsilon}{k} \le \left(\frac{\sigma_j}{s_k}\right)^{2n} \le \frac{1}{k^2}+\frac{2\varepsilon}{k},$$

therefore the bound

$$\left|(\Delta \ln)_k\right| \leq \frac{|v|^{2n}}{(n!)^2}\left(\frac{1}{k}+2\varepsilon\right)$$

which can be made small at will.

Further remarks and comments

(i) The condition (7.34) is more or less equivalent to the classical Lindeberg condition in the (usual) central limit theorem on the convergence to the normal law (see for instance Feller [7.1]); and states that the magnitudes of the moments σ_j^n, $j = 1,2,\ldots,$ have comparable sizes: one of them is not more important than the other ones.

Indeed, the Lindeberg's condition (which applies to the case when $n = 2$) states that for any given positive ε and k sufficiently large, one has

$$\sigma_j / s_k \lhd \varepsilon \,. \tag{7.40}$$

If instead, we write

$$\sigma_j / s_k \cong \varepsilon \,, \tag{7.41}$$

then, equation (7.32), with $n = 2$, yields

$$\varepsilon^2 \cong 1/k \,, \tag{7.42}$$

that is to say the condition (7.34).

(ii) Hochberg [7.2] and Sainty [7.10] stated only the parallel of Theorem 7.1 for their probabilities densities, but it is clear that theorem 7.2 applies as well, and this is mainly due to the definition of the corresponding characteristic functions which read $\phi(v) = \exp\{\lambda(iv)\}$, $\lambda \in C$.

(iii) On the practical standpoint, the theorem 7.2 is of some importance, in the sense that it can support the use of C-$(fBm)_n$ processes in the modelling of problems which deal with supremal or overall processes resulting from the aggregation of infimal or individual decisions, like in stock exchange for instance.

7.4. Itô's stochastic integral of order *n*

7.4.1. Main definition

Stochastic integrals of step functions

Let $T := [a,c]$, and let $\{b(t,n),\ t \in T\}$ be a $C\text{-}(fBm)_n$. Suppose that the random function $g(t,\alpha)$, which depends upon the random event α, is defined on T. Partition T: $a = t_0 < t_1 < \ldots < t_n = c$.

Consider the step function

$$g(t,\alpha) := \begin{array}{l} 0 \quad , \quad t \lhd t_0 \\ g(t_i,\alpha) \quad , \quad t_i \le t \lhd t_{i+1} \\ 0 \quad , \quad t \ge t_n \end{array} \tag{7.43}$$

where $g(t_i,\alpha)$ is independent of $\{b(t_k,n)\text{-}b(t_j,n) : a \le t_i \le t_j \le t_k \le c\}$, and satisfies the condition $E\{|g(t,\alpha)|^n\} < \infty$. For such step functions, the Itô integral is defined by

$$\int_T g(t,\alpha)db(t,n) := \sum_{j=0}^{n-1} g(t_j,\alpha)\left[b(t_{j+1},n) - b(t_j,n)\right]. \tag{7.44}$$

Stochastic integral of general functions

Assume that $g(t,\alpha)$ is not a step function, and let $g(t,\alpha)_k$ denote a sequence of step functions converging to $g(t,\alpha)$ in the sense that

$$\int_T E\left\{\left|g(t,\alpha) - g(t,\alpha)_k\right|^n\right\}dt \ \downarrow\ 0 \quad as \quad k \uparrow \infty\ . \tag{7.45}$$

Then one set

DEFINITION 7.2 The Itô stochastic integral of order n of $g(t,\alpha)$ is defined as the limit

$$\int_T g(t,\alpha)db(t,n) := (\lim)_n \int_T g(t,\alpha)_k db(t,n) \quad as \quad k \uparrow \infty, \tag{7.46}$$

where $(\lim)_n$ holds for limit in the sense of the moment of order n (see equation (7.45).■

With this definition one has the following results, which are brought together for the conveniency of the reader.

7.4.2. Main results

LEMMA 7.2. *Let the random functions $f(t,\alpha)$ and $g(t,\alpha)$ satisfy the conditions of Definition 7.1. Then their Itô stochastic integrals of order n are defined by (7.46) and one has*

$$E\left\{\left[\int_T g(t,\alpha)db(t,n)\right]^k\right\} = 0 \quad , \quad k=1,2,...,n-1, \tag{7.47}$$

$$E\left\{\left[\int_T g(t,\alpha)db(t,n)\right]^k\left[\int_T f(t,\alpha)db(t,n)\right]^{n-k}\right\} = \sigma^n\int_T E\left\{g^k f^{n-k}\right\}dt \quad \blacksquare \tag{7.48}$$

In some instances Itô integral of order n is the $(\lim)_n$ of Rieman Stieljes sums, as shown by the following

LEMMA 7.3. *Let $E\{|g(t,\alpha)|^n\} < \infty$ for all $t \in T$. Suppose $g(t,\alpha)$ is continuous on T in the sense that $E\{|g(t+\tau,\alpha)-g(t,\alpha)|^n\}$ tends to zero with τ. Partition T : $a = t_0 < t_1 < ... < t_n = c$, and define $\rho := \max(t_{j+1} - t_j)$. Then the Itô's stochastic integral of $g(t,\alpha)$ equals the following n-order limit*

$$(\lim)_n \sum_{j=0}^{n-1} g(t_j,\alpha)\left[b(t_{i+1},n) - b(t_i,n)\right] = \int_T g(t,\alpha)db(t,n). \quad \blacksquare \tag{7.49}$$

LEMMA 7.4. *Let $g(t,\alpha)$ be such that $E\{|g(t,\alpha)|^n\} < \infty$. Suppose $g(t,\alpha)$ is continuous in the sense of the moment of order n (see Lemma 7.3). Then*

$$\int_T g(t,\alpha)(db(t,n))^n = \sigma^n\int_T g(t,\alpha)dt. \quad \blacksquare \tag{7.50}$$

Indications on the proofs.

Proof of lemma 7.2. (i) By using the properties of $b(t,n)$ and the independency of $f(t,\alpha)$ (resp. $g(t,\alpha)$) one can show easily that the lemma holds when f and g are step functions.

(ii) In the general case of non step-wise functions, we consider two sequences of step functions $\{f_j(t,\alpha)\}$ and $\{g_j(t,\alpha)\}$ which converge to $f(t,\alpha)$ and $g(t,\alpha)$ respectively as j increases.

(iii) Using these step functions, the proof of equation (7.47) is direct. In order to obtain (7.48), we calculate the difference

$$
\begin{aligned}
&\left|E\left\{g_j^k f_j^{n-k} - g^k f^{n-k}\right\}\right| \\
&\qquad = \left|\int_T E\left\{g_j^k (f_j^{n-k} - f^{n-k}) + (g_j^k - g^k) f^{n-k}\right\} dt\right| \\
&\qquad \le \int_T \left|E\left\{g_j^k (f_j^{n-k} - f^{n-k})\right\}\right| dt + \int_T \left|E\left\{(g_j^k - g^k) f^{n-k}\right\}\right| dt \\
&\qquad \le M_1 \int_T \left|E\left\{g_j^k (f_j - f)\right\}\right| dt + M_2 \int_T \left|E\left\{(g_j - g) f^{n-k}\right\}\right| dt \\
&\qquad \le M_1 E^{\frac{k}{n}}\left\{(g_j^k)^{\frac{n}{k}}\right\} E^{\frac{n-k}{n}}\left\{\left|f_j - f\right|^{\frac{n}{n-k}}\right\} \\
&\qquad\quad + M_2 E^{\frac{n-k}{n}}\left\{(f^{n-k})^{\frac{n}{n-k}}\right\} E^{\frac{k}{n}}\left\{\left|g_j - g\right|^{\frac{n}{k}}\right\} \quad \to \quad 0 \quad (j \to \infty)
\end{aligned}
$$

The powers n/k and $n/(n - k)$ are the conjugate exponents of Schwarz inequality, and the convergence to zero is based on the fact that all the moments of order $k \le n$ exist.

Proof of lemma 7.3. One defines

$$\tilde{g}(t,\alpha) \quad := \quad g(t_i,\alpha) \quad , \quad t_i \le t \lhd t_{i+1},$$

and one then has

$$
\begin{aligned}
\Delta \quad &:= \quad \int_T g(t,\alpha) db(t,n) - \sum_{j=0}^{n-1} g(t_j,\alpha)\left[b(t_{j+1},n) - b(t_j,n)\right] \\
&= \quad \int_T \left[g(t,\alpha) - \tilde{g}(t,\alpha)\right] db(t,n),
\end{aligned}
$$

therefore

$$
\begin{aligned}
E\left\{\left|\Delta^n\right|\right\} \quad &= \quad \sigma^n \int_T E\left\{\left|g(t,\alpha) - \tilde{g}(t,\alpha)\right|^n\right\} dt \\
&\to 0 \quad as \quad \rho \to 0. \ .
\end{aligned}
$$

Proof of lemma 7.4. (i) We have to prove that

$$E\left\{|\Delta|^n\right\} \quad := \quad E\left\{\left|\sum_{j=0}^{m-1} g_j\left(\Delta b_j^n - \sigma^n \Delta t_j\right)\right|^n\right\} \quad \to \quad 0 \quad as \quad \rho \to 0$$

(ii) To this end, we remark that one has the equality

$$E\{|\Delta|^n\} = E\left\{\sum_{j=0}^{m-1}|g_j|^n\left|\Delta b_j^n - \sigma^n \Delta t_j\right|\right\}$$

and that the mutual independence condition provides

$$E\{|\Delta|^n\} = \sum_{j=0}^{m-1} E\{|g_j|^n\}E\{|\Delta b_j^n - \sigma^n \Delta t_j|\}.$$

(iii) Moreover, a simple calculation yields

$$E\{|\Delta b_j^n - \sigma^n \Delta t_j|^n\} = 2(\sigma^n \Delta t_j)^n ,$$

therefore

$$E\{|\Delta|^n\} = 2(\sigma^n)^n \sum_{j=1}^{m-1} E\{|g_j|^n\}(\Delta t_j)^n,$$

$$\leq 2\sigma^n \rho^{n-1} \sum_{j=1}^{m-1} E\{|g_j|^n\}\Delta t_j \rightarrow 0 \quad as \quad \rho \rightarrow 0 ,$$

as a result of the inequality

$$\int_T E\{|g(t,\alpha)|^{dt}\}dt < \infty .$$

Example . *Modelling of the process exp{b(t,n)}*

Let us consider the scalar process defined by the expression

$$u(t) = \exp\{b(t,n)\} , \quad t \geq 0 , \quad b(0,n) = 0 \quad wp1. \tag{7.51}$$

Taylor's expansion truncated at the order n yields ($db^j = (db)^j$)

$$du(t) = u(t)(db + \frac{1}{2}db^2 + ... + \frac{1}{n!}db^n) + o(db^{n+1}), \tag{7.52}$$

and on integrating we have

$$\begin{aligned} u(t) &= 1 + \sum_{j=1}^{n-1} \frac{1}{j!}\int_0^t u db^j(t,n) + \frac{1}{n!}\int_0^t u db^n(t,n) \\ &= 1 + \sum_{j=1}^{n-1} \frac{1}{j!}\int_0^t u\, db^j(t,n) + \frac{1}{n!}\sigma^n \int_0^t u(t)dt \cdot \end{aligned} \tag{7.53}$$

It follows that the differential representation of (7.51) is

$$du(t) = \sum_{j=1}^{n-1} \frac{1}{j!} u(t)(db(t,n))^j + \frac{1}{n!}\sigma^n u(t)dt \cdot \tag{7.54}$$

7.5. Stochastic differential equation of order *n*

7.5.1. Definition of the model

Deterministic Systems. Let us consider a dynamical system described by the 2D state vector *(x,y)* in the phase plane. We define the complex state

$$z(t) := x(t) + iy(t), \tag{7.55}$$

and we assume that its dynamics can be suitably defined by a differential equation in the form

$$\dot{z}(t) = f_1(z,t) \quad , \quad z \in C. \tag{7.56}$$

Stochastic systems. Due to some random disturbances, the actual system is not driven by the equation (7.56), but rather by the stochastic differential one

$$\dot{z}(t) = f_1(z,t) + \sum_{j=2}^{n} f_j(z,t)w(t,j), \tag{7.57}$$

where the *w(t,k)*'s are mutually independent fractional Gaussian white noises of order k, $k = 2,3,..,n$, satisfying the condition

$$E\{w^k(t,k)\} = \sigma^k(t) \quad , \quad k = 2,3,..,n, \tag{7.58}$$

Analogously with the differential equation $dx/dt = f(x,t) + g(x,t)w(t,2)$, we shall assume that the functions $f_k(z,t)$, $k = 1,2,..,n$ are sufficiently smooth on the mathematical standpoint. For instance, in the theory of Itô's stochastic differential equation it is customary to suppose that the functions $f(x,t)$ and $g(x,t)$ satisfy uniform Lipschitz conditions on u and t respectively

7.5.2. On the meaning of the stochastic differential equation

To write equation (7.57), we used the formalism of the engineering mathematics literature, and there remains to clarify its significance. To this end we firstly assume that the functions $f_k(z,t)$, $k = 2,3,..,n$, do not depend upon the parameter z, to yield

$$\dot{z}(t) = f_1(t) + \sum_{j=2}^{n} f_j(t)w(t,j). \tag{7.59}$$

In such a case, the differential dz can be expanded in the form

$$dz = \sum_{j=1}^{n} dz_j, \tag{7.60}$$

where z_1 is the contribution of $f_1(t)$ and z_j is the contribution of $f_j(t)w(t,j)$, $j = 2,3,..,n$. Clearly, with Maruyama's notation one can set

$$dz_1 = f_1(t)dt, \tag{7.61}$$

and

$$dz_j = f_j(t)w(t,j)(dt)^{1/j}, \tag{7.62}$$

and re-write the equation (7.59)) in the more meaningful form

$$dz = f_1(t)dt + \sum_{j=2}^{n} f_j(t)w(t,j)(dt)^{1/j}, \tag{7.63}$$

or, in terms of complex-valued fractional Brownian motion of order j,

$$dz = f_1(t)dt + \sum_{j=2}^{n} f_j(t)db(t,j)\cdot \tag{7.64}$$

In a like manner we shall rewrite equation (7.57) in the forms

$$dz = f_1(z,t)dt + \sum_{j=2}^{n} f_j(z,t)w(t,j)(dt)^{1/j}, \tag{7.65}$$

and

$$dz = f_1(z,t)dt + \sum_{j=2}^{n} f_j(z,t)db(t,j)\cdot \tag{7.66}$$

There remains now to examine how we can define the solution of this equation, and to this end, we can extend step by step Itô's theory of stochastic differential equations; via the approximation scheme

$$z_n(t) = z(t_0) + \int_{t_0}^{t} f_1(z_{n-1},\tau)d\tau + \sum_{j=2}^{n}\int_{t_0}^{t} f_j(z_{n-1},\tau)db(\tau,j) \tag{7.67}$$

which will be convergent, provided that $f_1,f_2,\ldots,f_n$ satisfy mathematical conditions which are sufficiently mild.

7.6. Itô's lemma of order *n*

LEMMA 7.5. [7.4,7.5] *Let $z(t)$ denote the solution of the stochastic differential equation (7.65), and let $\phi(z,t)$ be a complex-valued function, continuously differentiable with respect to t and having continuous partial derivatives w.r.t. z up to the order n. Then the (stochastic) differential $d\phi$ of ϕ is*

$$d\phi(z,t) = \left[\phi_t + \phi_z' f_1 + \sum_{j=2}^{n}\frac{1}{j!}\phi_z^{(j)} f_j^j \sigma_j^j(t)\right]dt + \phi_z\sum_{j=2}^{n} f_j db_j \tag{7.68}$$

$$+\sum_{j=2}^{n}\sum_{m_2,\ldots,m_n}\frac{1}{j!}\phi_z^{(j)}\binom{j}{0\,m_2\,\ldots\ldots m_n}f_2^{m_2}\ldots f_n^{m_n}(db_2)^{m_2}\ldots(db_n)^{m_n},$$

$$m_2+\ldots.+m_n=j\quad;\quad\frac{m_2}{2}+\ldots.+\frac{m_n}{n}\lhd 1$$

where ϕ_t *is the partial derivative w.r.t. t and* $\phi_z^{(j)}$ *is the partial differential of order j w.r.t. z.*

With Maruyama's notation we shall rewrite this equation in the form

$$d\phi(z,t) = \left[\phi_t+\phi_z f_1+\sum_{j=2}^{n}\frac{1}{j!}\phi_z^{(j)}f_j^{\dagger}\sigma_j^{j}(t)\right]dt+\phi_z\sum_{j=2}^{n}f_j w(t,j)(dt)^{1/j} \tag{7.69}$$

$$+\sum_{j=2}^{n}\sum_{m_2,\ldots,m_n}\frac{1}{j!}\phi_z^{(j)}\binom{j}{0\,m_2\,\ldots\ldots m_n}f_2^{m_2}..f_n^{m_n}w^{m_2}(t,2)..w^{m_n}(t,n)(dt)^{\frac{m_2}{2}+\ldots.+\frac{m_n}{n}},$$

$$m_2+\ldots.+m_n=j\quad,\quad\frac{m_2}{2}+\ldots.+\frac{m_n}{n}\le 1$$

■

Proof. (i) As a special case assume that $\phi(z,t) = z^m$, where $m > n$ is a positive integer. We can then make the explicit calculation of the differential $d(z^m)$, and in the result so obtained we shall drop the term of order larger than n, since they contribute terms of order $(dt)^{1+c}$ where c is a positive real number.

(ii) According to the lemma (7.4), equation (7.50), one has the equality

$$f(t,\omega)db^n(t,n) = \sigma^n(t)f(t,\omega)dt\ , \tag{7.70}$$

in such a manner that we can make the corresponding substitution in the result obtained above.

(iii) We now refer to the Stone Weierstrass theorem of approximation of functions on a compact set by a sequence of polynomials functions in the sense of the uniform convergence, to obtain the result.

As an example, for the equation

$$dz = f_1(z,t)dt+f_2(z,t)db(t,2)+f_3(z,t)db(t,3),$$

one has

$$d\phi(z,t) = (\phi_t + \phi_z f_1 + \frac{1}{2}\phi_{zz} f_2^2 \sigma^2 + \frac{1}{3!} f_3^3 \sigma_3^3)dt + \phi_z(f_2 db(t,2) + f_3 db(t,3)) + \frac{1}{2!}\phi_{zz}\left[f_3^2 (db(t,3)^2 + 2 f_2 f_3 db(t,2) db(t,3)\right].$$

7.7. Basic lemma of Itô's stochastic calculus of order n

LEMMA 7.6. *Let $\psi(x)$ be a n times continuously differentiable real scalar function of the real variable x, and let L(a,c) denote the path generated by b(t.n) on the time interval T := [a,c], c > a. Then one has the equality*

$$\oint_{L(a,c)} \sum_{j=1}^{n-1} \frac{1}{j!} \psi_x^{(j)}(b)(db(t,n))^j = \psi[b(c,n)] - \psi[b(a,n)] - \frac{1}{n!}\sigma^n \oint_{L(a,c)} \psi_x^{(n)}[b(t,n)]dt \,, \tag{7.71}$$

where the integral should be understood as path integrals along trajectories generated by b(t,n), and $\psi_b^{(j)}$ holds for $d^j\psi/db^j$. ■

Proof. (i) According to Itô's lemma of order *n*, the differential $d\varphi(b(t,n))$ is

$$d\psi(b) = \sum_{j=1}^{n-1} \frac{1}{j!} \psi_b^{(j)} db^j(t,n) + \frac{1}{n!}\sigma^n \psi_b^{(n)}(b)dt \,. \tag{7.72}$$

(ii) We integrate this equality along a given generated path to obtain the result.

We then deduce the following

Corollary 5.1. If one considers the integral of any function of *b(t,n)* in the sense of the mathematical expectation $E\{(.)^n\}$, then equation (7.71) reads

$$\int_a^c \sum_{j=1}^{n-1} \frac{1}{j!} \psi_x^{(j)}(b)(db(t,n))^j = \psi[b(c,n)] - \psi[b(a,n)] - \frac{K(n)}{n!}\sigma^n \int_a^c \psi_x^{(n)}[b(t,n)]dt. \tag{7.73}$$

7.8. Generalized Fokker-Planck equation

7.8.1. Probability density and complex variable

In order to derive the parallel of the Fokker-Planck equation for the stochastic differential equation of order n (7.66), we shall assume that we can define a density function $p(z,t)$ for $z(t)$, which satisfies the conditions

$$p(+\infty,t) = p(-\infty,t) = 0, \tag{7.74}$$

$$\int_{R^2} p(z,t)dxdy = 1, \tag{7.75}$$

$$\int_{R^2} z^j p(z,t)dxdy = m_j(t) \quad , \quad j \in N^* ; \tag{7.76}$$

where the right hand side term $m_j(t)$ is bounded and identified with the ensemble average $\langle z^j \rangle$

With this definition, the variance $Var\{z\}$ of z will be given by the expression

$$Var\{z\} = E\left\{\left[z - E\{z\}\right]^2\right\} \tag{7.77}$$

where $E\{.\}$ denotes the mathematical expectation

$$E\{z\} = \int_{R^2} zp(z)dxdy, \tag{7.78}$$

and one can verify that, for any complex-valued constant α, one has the equality

$$Var\{\alpha z\} = \alpha^2 Var\{z\}. \tag{7.79}$$

7.8.2. Fokker-Planck equation of order n

LEMMA 7.7. *The complex-valued density $p(z,t)$ of the state $z(t)$ in the stochastic differential equation of order n (7.65) and (7.66) is the solution of the generalized Fokker-Planck equation*

$$\frac{\partial p(z,t)}{\partial t} = -\frac{\partial}{\partial z}\left[f_1(z,t)p\right] + \sum_{j=2}^{n}(-1)^j \frac{1}{j!}\sigma_j^j(t)\frac{\partial^j}{\partial z^j}\left[f_j^j(z,t)p\right]. \blacksquare \tag{7.80}$$

Proof. (i) We apply the Itô lemma of order n to the function $f(z,t) = z^k$, $k > n$, to obtain

$$\frac{\partial z^k(t)}{\partial t} = \sum_{j=1}^{n} \frac{1}{j!} k(k-1)..(k-j+1)\sigma^j(t) f_j^j(z,t) z^{k-j}(t) \cdot \tag{7.81}$$

(ii) Multiplying both sides by $p(z,t)$ and integrating yields (after noticing that $\partial m_k(t)/\partial t = dm_k/dt$)

$$\frac{d}{dt} m_k(t) = \sum_{j=1}^{n} \frac{1}{j!} k(k-1)...(k-j+1)\sigma_j^j(t) \int_{R^2} z^{k-j} f_j^j(z,t) p(z,t) dxdy, \tag{7.82}$$

and

$$\frac{d}{dt} m_k(t) = \int_{R^2} u^k \frac{\partial p}{\partial t} dxdy \ .$$

We rewrite (7.81) in the form

$$\int_{R^2} \sum_{j=1}^{n} \left[z^k \frac{\partial p}{\partial t} - \frac{1}{j!} k(k-1)..(k-j+1)\sigma_j^j(t) z^{k-j} f_j^j(z,t) p(z,t) \right] dxdy = 0 \cdot \tag{7.83}$$

(iii) And integrating by parts with respect to the term z^{k-j} provides

$$\int_{R^2} z^k \left[\frac{\partial p}{\partial t} + \frac{\partial}{\partial z}(f_1 \, p(z,t)) - \sum_{j=2}^{n} (-1)^j \frac{1}{j!} \sigma_j^j(t) \frac{\partial^j}{\partial z^j} \left(f_j^j \, p(z,t) \right) \right] dxdy = 0 \tag{7.84}$$

(iv) Now we extend the solution sought in assuming that the equation (7.84) should be satisfied for all positive integer k, and since z^k, $k \geq 0$, is a complete basis of polynomials in C equation (7.80) is necessarily satisfied.

7.8.3. Dynamical equations of the state moments

LEMMA 7.8. *The dynamical equations of the moments of z(t) read as follows:*

$$\dot{m}_1(t) = \int_{R^2} f_1(z,t)p(z,t)dxdy, \tag{7.85}$$

$$\begin{aligned}\dot{m}_k(t) &= \int_{R^2} kz^{k-1} f_1(z,t)p(z,t)dxdy \\ &+ \sum_{j=2}^{n} Y(k-j)\frac{1}{j!}k(k-1)...(k-j+1)z^{k-j}\sigma_j^j(t)\int_{R^2} f_j^j(z,t)p(z,t)dxdy\end{aligned} \tag{7.86}$$

where Y(k - j) is the so-called Heaviside function: clearly, Y(k - j) = 1 when k ≥ j and Y(k-j) = 0 whenever k < j. ■

Proof. It is sufficient to multiply both sides of equation (7.80) by z^k and to integrate over $\Re^2$.

7.9. Relation with Kramers-Moyal expansion

7.9.1. Background on Kramers-Moyal expansion

Main Result Let us consider a scalar-valued Markovian stochastic process *x(t)* with the state probability density *p(x,t)* and the state transition probability density *q(x,t+dt|x',t).* Assume that all the transition moments are known to satisfy the equality

$$\begin{aligned}E\{(x'-x)^j\}_t &:= \int_R q(x',t+dt|x,t)p(x,t)dx \ , \\ &= a_j(x,t)dt \quad , \quad j=1,2,3,...\end{aligned} \tag{7.87}$$

Then the density *p(x,t)* is defined by the partial differential equation

$$\frac{\partial p(x,t)}{\partial t} = \sum_{j=1}^{n}\frac{(-1)^j}{j!}\frac{\partial^j}{\partial x^j}\left[a_j(x,t)p(x,t)\right]. \tag{7.88}$$

Proof. The proof, which is interesting in itself, some interest, can be decomposed into the following steps.

(i) One firstly assumes that the Chapman-Kolmogorov equation applies to yield

$$p(x,t+\tau) = \int_R q(x,t+\tau|x',t)p(x',t)dx' \ . \tag{7.89}$$

where *q(.)* denotes the density of conditional transition probability defined above.

(ii) The characteristic function $\varphi(u,x',t,\tau)$ of the transition probality is

$$\varphi(v,x',t,\tau) := \int_R e^{iv(x-x')} q(x,t+\tau|x',t)dx$$

$$= 1+\sum_{j=1}^{\infty}\frac{(iv)^j}{j!}a_j(x',t)dt, \tag{7.90}$$

and on taking the inverse transformation, we have

$$q(x,t+dt|x',t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iv(x-x')}\left[1+\sum_{j=1}^{\infty}\frac{(iv)^j}{j!}a_j(x',t)dt\right]dv. \tag{7.91}$$

(iii) But one has the equalities $(j \geq 0)$

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}(iv)^j e^{-iv(x-x')}dv = -\frac{\partial^n}{\partial x^n}\delta(x-x'), \tag{7.92}$$

and

$$\delta(x-x')f(x') = \delta(x-x')f(x), \tag{7.93}$$

in such a manner that the equation (7.90) yields

$$q(x,t+dt|x',t) = \left[1+\sum_{j=1}^{\infty}\frac{(-1)^j}{j!}\frac{\partial^j}{\partial x^j}a_j(z,t)\right]\delta(x-x'). \tag{7.94}$$

(iv) Inserting (7.94) into (7.89) provides the result (7.88).

Let us point out that this expansion can be obtained, in the information theoretic framework, as a consequence of the maximum entropy principle [7.3] .

PAWULA'S THEOREM. The problem, of course, is to check that $p(x,t)$, as defined by the Kramers-Moyal extension, is a positive function. Pawula's theorem (see for instance Risken [7.9]) *states that for having a positive $p(x,t)$, the right hand side term of (7.88) may stop either after the first term or the second one; and if it does not stop after the second term it must contain an infinite number of terms.*

7.9.2. Relation with complex Brownian motion of order n

According to Pawula's theorem a real stochastic process either has its first and second transition moments of order *o(dt),* or has all of them equal to *o(dt);* otherwise, it is not physically meaningful. Nevertheless, if we work in the complex plane then the Kramers Moyal expansion can make sense. Indeed, according to the above derivation, it can be thought of as defining the complex-valued probability density of a stochastic process of which the first n transition moments are generated by a sequence of complex-valued Brownian motions of order j, $j = 2,3,\ldots,n$. Clearly, the complex Brownian motion allows us to consider the Kramers-Moyal expansion with a viewpoint a little bit different from the usual one regarding classical Markovian processes.

But the opposite is also true. The Kramers-Moyal expansion is obtained by using the Chapman-Kolmogorov equation, and as a result, given the above equivalence between the two derivations, one can infer that the Chapman-Kolmogorov equation still applies in the complex plane to yield

$$E\left\{(z'-z)^j\right\} = \int_{\Re^2} q(z',t+dt|z,t)p(uz,t)dxdy, \tag{7.95}$$

7.10. Feynman-Kac formula

In the present framework of stochastic calculus of order n, we can generalize the so called Feynman-Kac formula [7.7] as follows:

LEMMA 7.9. *Let f and g denote two $C \to C$ analytic functions verifying the conditions*

$$\sum_{j=0}^{n}\left|f^{(j)}(z)\right| \lhd M , \tag{7.96}$$

and

$$\sum_{j=0}^{n}\left|g^{(j)}(z)\right| \lhd M , \tag{7.97}$$

for any $z \in C$. Define $\xi(t)$ as the solution of Itô's stochastic differential equation of order n,

$$d\xi(t) = a_1 dt + \sum_{j=1}^{n} a_j db(t,j) \quad , \quad a_j \in R, j = 1,2..,n, \tag{7.98}$$

where the a_j are constants, and the $b(t,j)$ are independent C-$(fBm)_n$ with unit variance. Then the solution $z(x,t)$: $\Re \times \Re \to C$ of the equation

$$\frac{\partial z}{\partial t}(x,t) + \sum_{j=1}^{n} \frac{(-1)^j}{j!} a_j{}^j \frac{\partial^j z}{\partial x^j}(x,t) + g(x)z(x,t) = 0, \tag{7.99}$$

$$z(x,0) = f(x) \ , \tag{7.100}$$

is defined by the expression

$$z(x,t) = E_\xi \left\{ f(\xi) \exp\left\{ -\int_0^t g(\xi,\tau) d\tau \right\} \right\}, \tag{7.101}$$

where the expectation is computed by replacing the integral by a Riemann sum, and then going on to the limit. ■

Formal derivation. Given that we are working in a probabilistic framework, we can duplicate step by step the proof of the original Kac formula [7.6], see also Krylov [7.7], and here we shall not repeat it here. Instead, we shall describe a formal derivation which can be of help to the practioner.

(i) According to Itô's lemma of order n, one has the equality

$$dz(\xi,t) = \left[a_1 z_t + \sum_{j=1}^{n} \frac{(-1)^j}{j!} a_j{}^j z^{(j)}(\xi,t) \right] dt + terms\ in\ [db(t,j)]^k .$$

This allows us to define the stochastic derivative of $z(\xi,t)$ w.r.t. time, along a given $\xi(t)$-path, by the expression

$$\frac{d_S z}{dt}(\xi,t) := a_1 \frac{\partial z}{\partial t} + \sum_{j=1}^{n} \frac{(-1)^j}{j!} a_j{}^j z^{(j)}(\xi,t). \tag{7.102}$$

(ii) With this notation the equation (7.99) can be rewritten in the form

$$\frac{d_S z}{dt}(\xi,t) = -g(\xi)z(\xi,t), \tag{7.103}$$

of which the formal solution is

$$z(\xi,t) = f(\xi)\exp\left\{-\int_0^t g(\xi(\tau))d\tau\right\}. \tag{7.104}$$

(iii) This being the case, the differential $dz(\xi,t)$ of the expression so obtained will involve undesirable terms in $(db)^k$, therefore the need to take the mathematical expectation of the right side term of (7.104).

Notice that this formal derivation is quite similar to the formal derivation of Itô's lemma via Taylor's expansion.

7.11. Dynkin's formula

The following result, of the type of Dynkin's formula, can be useful in the analysis of stochastic systems subject to *C-(fBm)$_n$*.

LEMMA 7.10. *Let $z(t) \in C$ be the unique solution of the Itô stochastic differential equation*

$$dz(t) = f(z,t)dt + g(z,t)db(t,n), \tag{7.105}$$

where $b(t,n)$ is a C-(fBm)$_n$ with the characteristic function $\exp\{(iv)^n\sigma^n(t)t/n!\}$. Let (z,t) denote a C-valued function, continuously differentiable in t and having continuous partial derivations with respect to z, up to the order n. Then the following equality holds, that is

$$E\{h(z(t),t)\} - E\{h(z(s),s)\} = \int_s^t E\left\{\frac{\partial h}{\partial \tau}(z,\tau) + \frac{(-1)^n}{n!}\sigma^n(\tau)\frac{\partial^n h}{\partial u^n}(z,\tau)g^n(z,\tau)\right\}d\tau \quad . \blacksquare \tag{7.106}$$

Proof. One applies Itô's lemma of order n; then one integrates with respect to time, and on taking the average on each side we obtain the result.

Formal derivation. The following formal derivation can be useful, as far as it may be of help in extending this result to stochastic dynamics involving several mutually independent *C-(fBm)$_n$* 's..

(i) One expands $h(z,t)$ in a Taylor series up to the order n, to have

$$dh \cong h_t dt + \sum_{j=1}^{n} \frac{1}{j!} \frac{\partial^j h}{\partial u^j} \left(f\, dt + g\, db(t,n) \right)^j . \tag{7.107}$$

(ii) We replace $(db(t,n))^n$ by $E\{(db(t,n))^n\} = (-1)^n \sigma^n(t) dt/n!$, and we delete the dt terms of order larger than the unity, to have

$$dh = (h_t + \frac{1}{n!} h_u^{(n)} g^n \sigma^n) dt + \sum_{j=1}^{n-1} \frac{1}{j!} h_u^{(j)} g^j (db)^j . \tag{7.108}$$

(iii) We integrate both sides; we take the mathematical expectation, we notice that $E\{(db)^j\} = 0$, $j = 1,...,n-1$, and we obtain the result.

This result will be of help in analysing the stability of fractal stochastic systems of order n by using stochastic Liapunov function (see for instance [7.8] in the case $n = 2$).

7.12. Concluding remarks

In the present chapter we have defined the basic elements of a theory of Itô's stochastic differential equations of order n (that is to say, subject to fractional Brownian motion of order n). The working framework is that of engineering mathematics with its applied mathematics point of view regarding stochastic dynamics..

In fact, one of the purposes of $C\text{-}(fBm)_n$ is to take account of collisions which may occur between more than two particles in statistical thermodynamics. Of course, this is not to construct a revisited theory of thermodynamics, but rather to have at hand a model which seems to be more meaningful when one applies thermodynamics to systems of which the elements are not particles, but, for instance, human actors, in which case collisions of order n, $n \geq 3$, are quite relevant.

In addition, at first glance the present model in the phase plane should be of some prospects in the analysis of fractal images.

For the applications of this fractional stochastic differential equation, we need some more results related to this stochastic fractional dynamics, and in the following we shall consider some of these topics, such as for instance, the stability and the informational content of the complex state $z(t)$.

References

7.1 FELLER, W.; *An Introduction to Probability Theory and its Applications,* John Wiley, NY, 1966

7.2 HOCHBERG, G.; Signed measure of probability, *Annals of Probability,* Vol 6, No 3, pp 433-458, 1978

7.3 JUMARIE, G.; Generalization of the diffusion equation by using the maximum entropy principle, *J. Math. Physics,* Vol 26, No 6, pp 1173-1177, 1985

7.4 JUMARIE, G.; Stochastic fractional dynamics in the phase plane I. Complex - valued noises of order n, *Systems Analysis Modelling Simulation,* Vol 33, pp 239-258, 1998

7.5 JUMARIE, G.; Stochastic fractional dynamics in the phase plane II (Complex-valued probability density and Kramers-Moyal expansion)

7.6 KAC, M.; On distributions of certain Wiener functionals, *Trans. Amer. Math. Soc.,* Vol 65, pp 1-13, 1949

7.7 KRYLOV, V.Yu.; Some properties to the distribution corresponding to the equation $\partial u/\partial t = (-1)^{q+1}\partial^{2q}u/\partial x^{2q}$, *Soviet. Math. Dokl.* , Vol 1, pp 760-763, 1960

7.8 KUSHNER, H.J.; *Stochastic Stability and Control,* Academic Press, NY, 1967

7.9 RISKEN, H.; *The Fokker Planck Equation*, Springer Verlag, Berlin, 1984

7.10 SAINTY, P.; Construction of a complex-valued fractional Brownian motion of order n, *J. Math. Physics*, Vol 33, No 9, pp 3128-3149, 1992

Chapter 8

Information Thermodynamics and Complex-Valued Fractional Brownian Motion of Order *n*

There is nothing more practical than a good theory.
Ludwig Boltzman

8.1. Introduction

Since the first reflections of Schrödinger about life [8.14] there has been a considerable amount of literature on self-organizing systems, and more especially on the various relationships between information, on the one hand, and such topics as order, disorder, complexity, and organization on the other. Loosely speaking, these books and articles can be divided into two kinds: those which are qualitative and philosophical, and those which deal with systems science and physics.

In the physical approach the mainstream is based upon the link which may exist between informational entropy (I-entropy) and thermodynamic entropy (T-entropy); but whilst this relation is clearly identified for closed systems (that is to say, systems which have no exchange with their environment), it is not yet completely clarified in the case of open systems. More explicitly, problems arise with the sign of the variation of the entropy.

Indeed, according to the second law of thermodynamics the T-entropy of a closed system is a non-decreasing function of time, and on the surface this is quite consistent with I-entropy. But with open systems, the T-entropy may increase or on the contrary decrease, in other words in the information-theoretic approach, one should have at hand dynamical models in which the I-entropy can decrease as time increases. One way of achieving this result would be to assume the existence of some external control, but it does not seem that it is so easy to construct suitable modellings of this approach in information-theoretic frameworks.

Fortunately, it appears that the introduction of external constraints provides a way of circumventing this problem, and in what follows, we shall show how this can be implemented by using the maximum entropy principle (the maxent principle). The

G. Jumarie, *Maximum Entropy, Information Without Probability and Complex Fractals*
© Springer Science+Business Media Dordrecht 2000

keywords herein are information, self-organization, thermodynamics, and fractals. By using the maxent principle Haken [8.4] put in evidence the link between self-organization and information-theoretic approach to thermodynamics. He showed that besides the usual thermodynamic equilibrium, defined by uniform density, temperature, pressure, concentrations, and so on, there exists a large domain of higher order equilibria, where all these local properties may be non-uniform. Here, we shall show that if one applies the maxent principle in the complex plane with the given values of the mathematical expectations as constraints, one obtains the probability density of the complex-valued Gaussian white noise of fractional order. In other words, one comes across a possible relation between thermodynamics and fractals.

The chapter comprises two parts. In the first, after some prerequisites for the convenience of the reader, we shall show how, in the maxent approach, the creation of constraints increases the self-organizing properties of systems.We shall take this opportunity to show how one can use the maxent principle to solve the Fokker-Planck equation. And in the second part, we shall deal with thermodynamics and fractals.

8.2. Informational entropy of open systems

8.2.1. Internal and external thermodynamic entropy

In the macroscopic approach to the analysis of physical systems consider a thermodynamic mass medium Σ. Let dQ denote the amount of energy received as heat by this system in a small reversible (that is to say, slow enough to be close to equilibrium at each instant) transformation, and let T denote its absolute (Kelvin) temperature. The thermodynamic entropy S (once again, we apologize to the reader, but the notations are local through the book, as in computer programming, and S here is not to be confused with the von Neumann quantum entropy, in Chapter 5) of Σ is defined by the differential equality

$$dS = dQ/T. \tag{8.1}$$

In the case of transformations far from equilibrium, the balancing equation of entropies is

$$dS = d_iS + d_eS, \tag{8.2}$$

where d_iS is the increase of entropy generated by the irreversible process inside the system, whilst d_eS denotes the amount of entropy contributed by its environment.

According to the second law of thermodynamics in physics, one has the inequality

$$d_iS \geq 0, \tag{8.3}$$

whilst d_eS may be positive or negative.

In thermodynamics, if p is the probability distribution of the different energetic micro-states of which the statistical average defines the macro-state of the system Σ, then one can show that S_i and H are related by the equality

$$S_i = kH, \tag{8.4}$$

where $k = 1{,}380.10^{-23}$ joules/ K° is the so called Boltzmann's constant.

When $d_eS = 0$ one has $dS = d_iS$, and we shall say that we are dealing with a closed system. Otherwise the system is referred to as an open system.

8.2.2. Open systems, order, disorder, organization

Following several authors [8.15,8.2,8.1,8.10], we shall measure the *disorder* of n distinguishable states $\Sigma_1, \Sigma_2,\ldots,\Sigma_n$ which are occupied by the system Σ with the respective probabilities $p_1,p_2,\ldots,p_n$ defining the informational entropy $H(p) \equiv H(\Sigma)$, by the expression

$$D(\Sigma) := H(\Sigma)/H_m \tag{8.5}$$

where $H_m = \ln n$ is the maximum value of the entropy; which occurs when all the probabilities are equal to $1/n$. In a like manner their *order* will be characterized by the quantity

$$R(\Sigma) := 1-\left(H(\Sigma)/H_m\right), \tag{8.6}$$

which is exactly the redundancy first introduced by Shannon and Weaver [8.13].

As a result we shall say that Σ is self-organizing whenever it satisfies the condition

$$\dot{R}(\Sigma,t) \ \triangleright \ 0, \tag{8.7}$$

or, in a like manner, when one has the condition

$$\frac{\dot{H}_m}{H_m} \ \triangleright \ \frac{\dot{H}}{H}. \tag{8.8}$$

Strictly speaking, since H_m takes on the values ln n, for positive integers, the derivative dH_m/dt in (8.8) should be replaced by the finite difference ΔH_m = ln (n + 1) - ln n. In other words, the system would be self-organizing provided that the following condition is satisfied:

$$\dot{H} \ \triangleleft \ 0 \quad \textit{whenever } H_m \textit{ is constant}, \tag{8.9}$$

and otherwise

$$\frac{\ln(n \pm 1) - \ln n}{\ln n} \ \triangleright \frac{\dot{H}}{H}, \tag{8.10}$$

when the number of states increases from n to (n+1), or on the contrary decreases from n to *(n-1)*.

The practical significance of condition (8.8) gives rises to some difficulties which we shall try to clarify as follows. As is mentionned above, when H_m is constant the system is self-organizing whenever it satisfies the condition $dH(\Sigma,t) < 0$; in other words, if we come back to the thermodynamic framework we cannot simply make the identification $kdH(\Sigma,t) \equiv d_i(\Sigma,t)$, but rather we are compelled to set:

$$kdH(\Sigma,t) \ = \ d_iS(\Sigma,t) + d_eS(\Sigma,t).$$

The problem which then arises can be stated as follows: how can we determine the respective informational counterparts of $d_iS(\Sigma,t)$ and $d_eS(\Sigma,t)$? Or again, in the informational definition of the system under consideration, how can we introduce its exchanges with the environment? Or else, how can we modify the definition of the probability function $(p_1,p_2,...,p_n)$ in order to account for the exchanges between the system and its environment? We shall see that the maxent principle provides some hints for a possible answer. The key is that the maximum value of the entropy should decrease

with time, and this is achieved by introducing new constraints on the definition of the system.

8.3. Maximum entropy principle and self-organization

8.3.1. Self-organization and constraint

The basic result is the following:

LEMMA 8.1. *Assume that we apply the maximum entropy principle with the J constraints* $f_j(p_1, p_2,..., p_n)$, $j = 1, 2, ..., J$ *to obtain the entropy* H_J^*. *Assume further that we remove one of these constraints, say the last one* f_J, *and that we apply the maximum entropy principle with the remaining constraints* $f_1, f_2,...,f_{J-1}$ *to obtain* H_{J-1}^*. *Then the following inequality holds*:

$$H_J^* \leq H_{J-1}^* \ .\blacksquare \tag{8.11}$$

Proof. This result is quite obvious:

(i) The function $H(p)$ is convex (it exhibits a mountain landscape) with the maximum valued achieved at $p_{\max} = (1/n, 1/n, ..., 1/n)$.

(ii) The constraints $\{f_1=0,...,f_{J-1}=0\}$ defines the domain D_{J-1} of the admissible p, which contains the corresponding optimal p_{J-1}^*, and, moreover, one has $p_{J-1}^* \neq p_{\max}$.

(iii) This being the case, the constraint $f_J = 0$ defines another domain D_J of admissible p, and the solution p_J^* of the new problem with the J constraints is such that $p_J^* \in D_J \cup D_{J-1}$. Two instances can then occur. Either $p_J^* \in D_J \cap D_{J-1}$, in which case one has the equality $p_J^* = p_{J-1}^*$ which provides $H_{J-1}^* = H_J^*$; or $p_{J-1}^* \notin D_J \cap D_{J-1}$, therefore, since $H(p)$ is a convex function, $H_J^* \lhd H_{J-1}^*$.

8.3.2. Application to a modelling of self-organization

This result provides a straightforward model of self-organization capability by the creation and the transfer of constraints on the informational entropy of the whole system.

A system would be organizing mainly because there is creation of constraints which causes its informational entropy to decrease. In the same way, it would be self-organizing whenever there is self-creation of constraints. In other words, the level, the grade of the organization capability of the system, would be directly characterized by the constraints which act on it.

The question which then arises is related to the possible origin of these constraints, and at first glance, we can make the following suggestions.

(i) Firstly, one may assume that these constraints originate from some physical bounds on the environment of the system, such as limitation of resources, for instance.

(ii) Secondly, in the special case of multi-level hierarchical systems the lower level is subject to the constraints generated by the upper level, and one may assume that it is the upper level subsystem itself which creates these constraints.

8.4. Analysis of a system comprising two components

Dynamics of a model defined by two subsystems.

We consider two subsystems A and B which can occupy the states $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_m$ with the probabilities $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_m$ respectively, when they are isolated from one another. A and B are then defined by their respective entropies

$$H(A) = -\sum_{i=1}^{n} p_i \ln p_i, \tag{8.12}$$

$$H(B) = -\sum_{j=1}^{m} q_j \ln q_j \cdot \tag{8.13}$$

$H(A)$ (and/or $H(B)$) may be equal to its maximum value $\ln n$, or on the contrary one may have $H(A) \lhd \ln n$. This case corresponds to the instance when A is subject to some exchanges with its environment, while A and B are mutually isolated from one another (model Σ_a). This is in contrast with the configuration Σ_b in which A and B have exchange with their respective environments on the one hand, and mutually between themselves on the other hand.

The informational entropy of the configuration Σ_a is $H(\Sigma_a) = H(A) + H(B)$, and that of Σ_b is $H(\Sigma_b) = H(A,b) \leq H(A) + H(B)$. According to the second law of thermodynamics, if the system is in the initial state Σ_b then in the absence of any further constraint it will tend to converge to the state Σ_a Clearly, the system is then disorganizing.

Introduction of a new constraint.

Assume that while the system A is in the state Σ_a a new constraint takes place in the form of the given mathematical expectation

$$\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} f(x_i, y_j) = \bar{f}, \tag{8.14}$$

where $\bar{f}$ has a given fixed value, $f(x,y)$ is a given real-valued function of the real variable (x,y), and r_{ij} is the mutual probability function of the pair (A,B).

For instance, one may assume that $f(x,y)$ is a function of the available resources which are necessary to ensure the survival of the system, and that x_i is the contribution of A when it is in the state A_i, whereas y_j is that of B in the state B_j (configuration Σ_d). Equation (8.14) then defines a prescribed value for the mean value of this resource function.

The informational entropy $H(\Sigma_d)$ is then defined by the expression

$$H(\Sigma_d) = -\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} \ln r_{ij}, \tag{8.15}$$

and self-organization will occur whenever the new constraint is such that

$$H(\Sigma_d) \lhd H(\Sigma_b). \tag{8.16}$$

8.5. Maximum path entropy principle and Fokker-Planck equation

8.5.1. Background on some results

Let $x(t) \in \Re^n$, $x^T := (x_1,x_2,...x_n)$ denote a stochastic process with a probability density $p(x,t)$ and the increment $x'(t,\tau) := x(t+\tau)-x(t)$. Define the conditional probability density $q(x',\tau|x,t)$ of x' at $t+\tau$ given the state x at t; define the multi-index $k := (k_1,k_2,...,k_n)$, $k_i \geq 0$ for every i, define

$$\varpi^k(x) := x_1^{k_1} x_2^{k_2} ... x_n^{k_n}, \tag{8.17}$$

and the moments

$$m_k(t) := \langle \varpi^k(x) \rangle. \tag{8.18}$$

The conditional expectation of $\varpi^k(x')$ given x at t will be denoted by $\langle \varpi^k(x')|x,t \rangle$ and furthermore the expression $|k| \geq K$ will be a shorthand for $k_1 + k_2 + ... + k_n \geq K$. We can now state the following result.

PROPOSITION 8.1. *Consider the continuous stochastic process above, and assume that it is Markovian and satisfies the following conditions:*

$$\langle x_i' | x,t \rangle = \tau f_i(x,t); \quad i = 1,2,..,n, \tag{8.19}$$

$$\langle x_i' x_j' | x,t \rangle = \tau g_{ij}(x,t); \quad i,j = 1,2,..,n, \tag{8.20}$$

$$\langle \omega^k(x) | x,t \rangle = o(\tau^2); \quad k \geq 3. \tag{8.21}$$

Then the moment m_k's are given by the following dynamical equation

$$\dot{m}_i(t) = \langle f_i(x,t) \rangle; \quad i = (\delta_{i1}, \delta_{2i}, ..., \delta_{ni}) \quad , \quad |i| = 1 \tag{8.22}$$

$$\begin{aligned} \dot{m}_k(t) &= \left\langle \sum_{i=1}^{n} f_i \partial_i \varpi^k(x) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} \partial_{ij} \varpi^k(x) \right\rangle, \\ &=: \langle P_k(x,t) \rangle \quad , \quad k \geq 2, \end{aligned} \tag{8.23}$$

with $\partial_i := \partial / \partial x$ and $\partial_{ij} := \partial^2 / \partial x_i \partial x_j$ and where δ_{ij} is the Kronecker delta. ■

PROPOSITION 8.2. *Assume that the following conditions are satisfied:*

(A1) *Conditions of proposition (8.1);*

(A2) *$f_i(x,t)$ and $\partial_j f_i(x,t)$ are continuous and bounded for every (i,j) and every $(x,t) \in \Re^{n+1}$;*

(A3) *$g_{ij}(x,t)$, $\partial_k g_{ij}(x,t)$ and $\partial_{ks} g_{ij}(x,t)$ are continuous and bounded for every (i,j,k,s) and every $(x,t) \in \Re^{n+1}$.*

Then the moment equations (8.22) and (8.23) yield the Fokker-Planck equation

$$\partial_t p(x,t) = -\sum_{i=1}^{n} \partial_i (f_i p) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \partial_{ij} (g_{ij} p), \tag{8.24}$$

as a consequence, and, conversely, the Fokker-Planck equation provides these moment equations. ■

PROPOSITION 8.3. *Assume that the following conditions are satisfied:*

(B1) *Conditions of proposition (8.1);*

(B2) *for every i $f_i(x,t)$ is a polynomial of the form*

$$f_i(x,t) = \sum_{|k|=0}^{K_i} f_{ik}(t) \varpi^k(x), \tag{8.25}$$

where $f_{ik}(t)$ is bounded, $|f_{ik}(t)| \leq M \lhd +\infty, 0 \leq t \leq +\infty$;

(B3) *for every (i,j), $g_{ij}(x,t)$ is a polynomial in the form*

$$g_{ij}(x,t) = \sum_{|k|=0}^{K_{ij}} g_{ijk}(t)\varpi^k(x), \tag{8.26}$$

where the $g_{ijk}(t)$ are bounded.

Then the moment equations (8.22) and (8.23) yield the Fokker-Planck equation as a consequence, and, conversely, the Fokker-Planck equation provides these moment equations. ■

Indications of the proof. These results are by now classical and there are a large number of different ways of obtaining them. For instance, in [8.6] we first obtained the state moment equations by using the moments of the increments, and then we used them to derive the Fokker-Planck equation via the property of the functions x^k, $k = 0,1,2,3,...$ in terms of the complete basis of functions.

These results show that under some mathematical assumptions which are not restrictive at all there is a complete equivalence beween the Fokker-Planck equation (8.24) on the one hand, and the moment equations (8.22) and (8.23) on the other, in such a way that we can use the latter for solving the former. To this end we shall refer to the maxiumum path entropy principle.

8.5.2. Maximum path entropy principle

Our purpose is to use the maxent principle to solve the Fokker-Planck equation by using the state moments equations (8.22) and (8.23) as constraints. But since these constraints involve the derivatives of the moments, that is to say the partial derivatives of the probability density $p(x,t)$, we shall first generalize the maxent principle as follows.

MAXIMUM PATH ENTROPY PRINCIPLE. *Assume that all we know about a stochastic process x(t) is a set of constraints on its probability density p(x,t). Then as an estimate of p(x,t) on the time interval $[T_1,T_2]$, we should select that density which satisfies these constraints and maximizes the path entropy*

$$H(X;T_1,T_2) := -\int_{T_1}^{T_2}\int_{\Re^n} p(x,t)\ln p(x,t)dxdt, \tag{8.27}$$

$$= \int_{T_1}^{T_2} H(X,t)dt, \tag{8.28}$$

for any arbitrary $[T_1,T_2]$. ■

The practical significance of this statement is quite understandable. For a given fixed interval $[T_1,T_2]$ we [8.8] have shown that $H(X;T_1,T_2)/(T_2-T_{1)}$ can be considered as the informational entropy of the portion of stochastic trajectory generated by x(t) on the time interval $[T_1,T_2]$; in such a manner that $H(X,T_1,T_2)$ can be thought of as the path entropy of the trajectory on $[T_1,T_2]$.

8.5.3. Application to the Fokker-Planck equation

(i) In order to solve the Fokker-Planck equation by means of the maximum path entropy principle with the state moments as constraints we shall rewrite the moment equations (8.22) and (8.23) in the form

$$\int_{\Re^n}\left[x_i\partial_t p(x,t)-f_i(x.t)p(x,t)\right]dx \;=\; 0, \tag{8.29}$$

$$\int_{\Re^n}\left[\omega^k(x)\partial_t p(x,t)-P_k(x,t)p(x,t)\right]dx \;=\; 0 \quad , \quad 2\le|k|\le K. \tag{8.30}$$

where K is the order of the approximation.

(ii) Next, we estimate the solution of the partial differential equation (8.24), and for a given initial $p(x,t_o)$, this solution is completely well defined at the instant T_1 and T_2. As a result we shall add the supplementary condition that $p(x,T_1)$ and $p(x,T_2)$ have given fixed values.

Given these prerequisites, we shall proceed as follows.

Step 1: By using the Lagrange multipliers $\lambda_o'(t),\lambda_i(t)$ and $\mu_k(t)$, we shall maximize the augmented path entropy

$$L \;:=\; \int_{\Re^n}\int_{T_1}^{T_2}\Bigg[-p\ln p+\lambda_o' p+\sum_{i=1}^{n}\lambda_i(t)(x_i\partial_t p-f_i p) + \sum_{|k|=2}^{K}\mu_k(t)\left(\varpi^k(x)\partial_t p-P_k p\right)\Bigg]dtdx, \tag{8.31}$$

which provides the variational condition of optimization

$$-\int_{\Re^n}\int_{T_1}^{T_2}\left[\ln p+1+\lambda_0'(t)+\sum_i\lambda_i(t)f_i+\sum_k\mu_k(t)P_k\right]\delta p(x,t)dtdx+$$

$$+\int_{\Re^n}\int_{T_1}^{T_2}\left[\sum_i \lambda_i(t)x_i + \sum_k \mu_k(t)\varpi^k(x)\right]\delta\big(\partial_t p(x,t)\big)dtdx \;=\; 0\cdot \tag{8.32}$$

Step 2: In order to manipulate the variation $\delta(\partial_t p(x,t))$, we shall write, for instance

$$\begin{aligned}
&\int_{\Re^n}\int_{T_1}^{T_2}\mu_k(t)\varpi^k(x)\delta\big(\partial_t p(x,t)\big)dtdx \\
&\quad= \int_{\Re^n}\varpi^k(x)\Big\{\big[\mu_k(t)\delta p(x,t)\big]_{T_1}^{T_2} - \int_{T_1}^{T_2}\dot{\mu}_k(t)\delta p(x,t)\Big\}dxdt \\
&\quad= -\int_{\Re^n}\int_{T_1}^{T_2}\dot{\mu}_k(t)\varpi^k(x)\delta p(x,t)dxdt\cdot
\end{aligned} \tag{8.33}$$

Step 3: Substituting this result into (8.32), we obtain the general form of the estimate $\hat{p}_K(x,t)$ in the form

$$\begin{aligned}
\hat{p}_K(x,t) \;=\; \exp\Bigg\{-\Bigg[&\lambda_0(t) + \sum_{i=1}^{n}\lambda_i(t)f_i(x,t) + \sum_{|k|=2}^{K}\mu_k(t)P_k(x,t) \\
&+\sum_{i=1}^{n}\dot{\lambda}(t)x_i + \sum_{|k|=2}^{K}\dot{\mu}_k(t)\varpi^k(x)\Bigg]\Bigg\},
\end{aligned} \tag{8.34}$$

where $\lambda_o(t)$ holds for

$$\lambda_0(t) \;:=\; 1+\lambda_o'(t). \tag{8.35}$$

8.5.4. Determination of the Lagrange multipliers

Step 1: In order to determine the explicit expression of the vectors

$$\lambda^T(t) \;:=\; \big(\lambda_o(t),\lambda_1(t),...,\lambda_n(t)\big)$$

and

$$\mu^T(t) \;:=\; \big(\mu_{k_1}(t),\mu_{k_2}(t),...\big),$$

we shall substitute (8.34) into the constraints (8.29) and (8.30) to obtain a nonlinear vector differential equation in the form

$$A(t,\lambda,\mu)\ddot{\lambda}+B(t,\lambda,\mu)\dot{\lambda}+C(t,\lambda,\mu)\lambda+D(t,\lambda,\mu)\ddot{\mu} + E(t,\lambda,\mu)\dot{\mu}+F(t,\lambda,\mu) = b(t,\lambda,\mu), \tag{8.36}$$

where the matrices A,B,C,D,E,F and the vector b are defined in terms of the expected values of functions which depend upon both x and t. For instance one has

$$b^T(t,\lambda,\mu) := \left(...,\langle f_i(x,t)\rangle,...,\langle P_k(x,t)\rangle,...\right).$$

Step 2. The determination of the solution of (8.36) is essentially a numerical computation problem, and for instance, one can use the iterative procedure defined by the equation

$$A(t,\lambda_n,\mu_n)\ddot{\lambda}_{n+1}+B(t,\lambda_n,\mu_n)\dot{\lambda}_{n+1}+C(t,\lambda_n,\mu_n)\lambda_n+D(t,\lambda_n,\mu_n)\ddot{\mu}_{n+1} +E(t,\lambda_n,\mu_n)\dot{\mu}_{n+1}+F(t,\lambda_n,\mu_n)\mu_{n+1} = b(t,\lambda_n,\mu_n). \tag{8.37}$$

Step 3. Initial conditions for $\lambda(t)$ and $\mu(t)$. In order to determine the values of the initial conditions on $\lambda(t)$ and $\mu(t)$, we shall proceed as follows.

(i) Refer to the initial conditions $p_o(x) = p(x,t_o)$ and assume that one has

$$p_o(x) = \exp\left\{-\sum_{i=0}^{n}\alpha_i x_i-\sum_{|k|=2}^{K}\beta_k\varpi^k(x)\right\}, \tag{8.38}$$

then comparing with (8.34) directly yields

$$\lambda_0(0) = \alpha_{0.}, \tag{8.39}$$

$$\lambda_i(0) = \mu_k(0) = 0 \ , \ i\geq 1 \ , \ |k|\geq 2, \tag{8.40}$$

and

$$\dot{\lambda}_i(0) = \alpha_i \ , \ i=1,2,...,n, \tag{8.41}$$

$$\dot{\mu}_k(0) = \beta_k \ , \ 2\leq|k|\leq K. \tag{8.42}$$

(ii) Assume now that $p_o(x)$ is not given in the special form (8.38), then by using a Galerking approximation, for instance, one can determine α_i and β_k so as to minimize the criterion

$$\min_{\alpha_i,\beta_k}\int_{\mathbb{R}^n}\left[\ln p_0(x)-\left(\sum_i \alpha_i x_i+\sum_k \beta_k \varpi^k(x)\right)\right]^2 p_o(x)dx, \tag{8.43}$$

and one will use the conditions (8.39) to (8.42).

Illustrative example. In order to illustrate the procedure, in the following, we shall outline the method for the well known special case defined by the one-dimensional Fokker-Planck equation

$$\partial_t p = -\partial_x(xf(t)p)+\frac{1}{2}\partial_{xx}(g(t)p), \tag{8.44}$$

with the initial condition

$$p(x,0) = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{(x-x_o)^2}{2\sigma^2}\right\}. \tag{8.45}$$

It is well known that the solution of (8.44) is exactly

$$p(x,t) = \frac{1}{(2\pi v(t))^{1/2}}\exp\left\{-\frac{(x-m_1(t))^2}{2v(t)}\right\}, \tag{8.46}$$

with

$$\begin{aligned} v(t) &:= m_2(t)-m_1^2(t), \\ \dot{m}_1(t) &= f(t)m_1(t) \quad , \quad m_1(0)=x_o, \\ \dot{m}_2(t) &= 2f(t)m_2(t)+g(t) \quad , \quad m_2(0) = \sigma^2+x_o^2. \end{aligned}$$

Let us seek the solution of (8.44) by using the maximum path entropy principle, and let us consider the approximations fo the second moment.

(i) One has successively

$$\begin{aligned} f(x,t) &= xf(t), \\ g(x,t) &= g(t), \\ P_k(x,t) &= kf(t)x^k+\frac{1}{2}k(k-1)g(t)x^{k-2}, \end{aligned}$$

and equation (8.34) yields the estimate

$$\begin{aligned}\hat{p}_2(x,t) &= \exp\left\{-\left[\lambda_o(t)+\mu_2(t)g(t)\right]-x\left[\lambda_1(t)f(t)+\dot{\lambda}_1(t)\right]\right. \\ &\quad \left.-x^2\left[2\mu_2(t)f(t)+\dot{\mu}_2(t)\right]\right\}, \qquad (8.47)\\ &= \exp\left\{-\alpha(t)-x\beta(t)-x^2\gamma(t)\right\}, \qquad (8.48)\end{aligned}$$

where the definition of $\alpha(t)$, $\beta(t)$ and $\gamma(t)$ is obvious.

(ii) We now substitute (8.48) into the normalizing condition

$$\int_{\Re^2}\hat{p}_2(x,t)dx = 1, \qquad (8.49)$$

together with the conditions (8.29) and (8.30), to have the differential equations

$$\dot{\alpha}+\dot{\beta}\langle x\rangle+\dot{\gamma}\langle x^2\rangle = 0, \qquad (8.50)$$

$$\dot{\alpha}\langle x\rangle+\dot{\beta}\langle x^2\rangle+\dot{\gamma}\langle x^3\rangle = -f(t)\langle x\rangle, \qquad (8.51)$$

$$\dot{\alpha}\langle x^2\rangle+\dot{\beta}\langle x^3\rangle+\dot{\gamma}\langle x^4\rangle = -2f(t)\langle x^2\rangle-g(t), \qquad (8.52)$$

where (8.50) is obtained by deriving (8.49) with respect to time.

(iii) The initial conditions on $\lambda_i(t)$ and $\mu_2(t)$ are (see Equs (8.39) to (8.42))

$$\lambda_o(0) = -\frac{1}{2}\ln(2\pi\sigma^2)-(x_o^2/2\sigma^2), \qquad (8.53)$$

$$\lambda_1(0) = 0\ ,\ \dot{\lambda}_1(0) = x_o/\sigma^2, \qquad (8.54)$$

$$\mu_2(0) = 0\ ,\ \dot{\mu}_2(0) = -(1/2\sigma^2). \qquad (8.55)$$

(iv) In order to calculate the respective values of α, β, γ, we need estimates of $\langle x_i\rangle$, $i = 1,..,4$; and to this end we shall refer to the equality $p(x,t) = p(x,0)$ therefore

$$\langle x\rangle_o = x_o, \qquad (8.56)$$

$$\langle x^2\rangle_o = x_0^2+\sigma^2, \qquad (8.57)$$

$$\langle x^3\rangle_0 = x_0^3+3\sigma^2x_0, \qquad (8.58)$$

$$\langle x^4\rangle_0 = x_0^4+6x_0^2\sigma^2+3\sigma^4. \qquad (8.59)$$

(v) Substituting (8.56)-(8.59) into (8.50)-(8.52) yields (α,β,γ) therefore $\lambda_0(t)$, $\lambda_1(t)$ and $\mu_2(t)$.

(vi) We now substitute $(\lambda_0(\tau)$, $\lambda_1(t)$ and $\mu_2(t)$ into the expression (8.47) of $\hat{p}_2(x,t)$ and then we can calculate $\left(\langle x\rangle, \langle x^2\rangle, \langle x^3\rangle, \langle x^4\rangle\right)$, therefore $(\alpha,\beta,\gamma)_2$ and hence $(\lambda_0(t),\lambda_1(t),\mu_2(2))_2$. And so on.

Further remarks and comments.

As is evident, here we have restricted ourselves to the solution of the Fokker-Planck equation, but it is clear that the method can be easily extended to the Fokker-Planck equation associated with Brownian motion of order n, since we have at hand the corresponding moment equations.

8.6. Maximum entropy and complex-valued fractional Brownian motion

8.6.1. Statement of the problem

It is well known that if one applies the maxent principle with the values of the first and second moments as constraints, one obtains the probability density of the normal law, which is the state probability density of the Brownian motion. Our purpose in the present section is to examine whether we have a similar result with $C\text{-}(fBm)_n$. What kind of results may one expect to obtain when one uses the maxent principle to determine the probability density of the $C\text{-}(fBm)_n$? The general expression of the density provided by this approach is in the form $p(x,t) = \exp\{V(\lambda,x)\}$, where λ is the Lagrange parameter vector; and the problem which then appears is that of determining the value of λ, what can be made by using a non-linear programming technique.

Here, in order to circumvent this pitfall, we shall consider a special set of random variables defined in a space derived from $\Re^{1/n}$ by means of the complex roots of the unity.In this way we shall convert the initial optimization with n constraints into a maximization with only one constraint, and thus simpler. To begin with, let us set the problem in the standard maxent approach.

Maxent approach to $C\text{-}(fBm)_n$

(i) Our purpose is to determine the probability $p(x,y)$ of a two-dimensional random variable (X,Y) which, in addition to the normalizing condition, satisfies the moment condition

$$\int_{\Re^2} z^j p(x,y)dxdy \;=\; 2\delta_{jn}\sigma^n, \tag{8.60}$$

where δ_{jn} is the Kronecker delta, and where the coefficient 2 is introduced to take account of the fact that we are dealing with the pair *(X,Y)*.

(ii) For the sake of consistency, we shall re-write condition (8.60) in the form

$$\int_{\Re^2} z^j p(z,z^*)dxdy \;=\; 2\delta_{jn}\sigma^n, \tag{8.61}$$

where $p(z,z^*)$ is defined by the equality

$$p(z,z^*) \;\equiv\; p(x,y),$$

with

$$z^* \;=\; x-iy.$$

(iii) With this notation one can neaningfully consider the entropy

$$H(z) \;:=\; -\int_{\Re^2} p(z,z^*)\ln p(z,z^*)dxdy, \tag{8.62}$$

and applying the maxent principle on (8.62) with the constraints (8.61) provides the estimate

$$\hat{p}(z) \;=\; \exp\left\{\lambda_0 + \sum_{i=1}^{n}\lambda_i z^i\right\}. \tag{8.63}$$

As we mentioned it above, the determination of the values of the Lagrange parameters is not so easy (it can be converted in a non-linear programming problem), and to obviate this difficulty we shall look for an alternative as follows.

8.6.2. Alternative to the formulation to the maxent principle

Statement of the problem revisited.

(i) The definition 6.6, equation (6.58) related to complex-valued Brownian motion of order *n*, suggests to apply the maxent principle in the subset

$$C^{1/n} \;:=\; \left\{z \middle| z = x_1\omega_1(n) + x_2\omega_2(n) + ... + x_n\omega_n(n)\right\}, \tag{8.64}$$

where $x_1, x_2, ..., x_n \in \Re$ are random variables which are mutually independent, and which have the same probability density $p(x)$, in such a manner that the problem now is to use the maxent principle for determining $p(x)$.

(ii) The equalities

$$E\left\{\left[x_1\omega_1 + x_2\omega_2 + ... + x_n\omega(n)\right]^j\right\} = 0 \quad , \quad j = 1,2,...,n-1, \tag{8.65}$$

are then automatically satisfied, in such a manner that we shall no longer need to use them as constraints when applying the maxent principle.

(iii) The only constraint which remains will be defined by the n-th moment

$$E\left\{\left(x_1\omega_1 + x_2\omega_2 + ... + x_n\omega_n\right)^n\right\} = n\sigma^n, \tag{8.66}$$

or what amounts to the same

$$\int_0^\infty x^n p(x)dx = \sigma^n. \tag{8.67}$$

Application of the maxent principle. Maximizing the entropy

$$H(X) = -\int_0^\infty p(x)\ln p(x)dx, \tag{8.68}$$

with the constraint (8.67) yields

$$p(x) = \frac{1}{n^{1/n}\sigma\Gamma\left(1+\frac{1}{n}\right)}\exp\left\{-\frac{x^n}{n\sigma^n}\right\}. \tag{8.69}$$

Generalized thermodynamics of order n. Identifying the state of the system with the n-tuple $(x_1, x_2, ..., x_n)$, its probability density is

$$p(x_1, x_2, ..., x_n) = \frac{1}{n\sigma^n\Gamma^n\left(1+\frac{1}{n}\right)}\exp\left\{-\frac{1}{n\sigma^n}\sum_{i=1}^n x_i^n\right\}; \tag{8.70}$$

and in quite a natural way, we are led to define the corresponding absolute temperature T_n by the equality

$$n\sigma^n = (k_B T_n)^n, \tag{8.71}$$

where k_B is the so called Boltzman constant.

The probability density (8.69) has already been obtained by some authors [8.5] who used the maximum entropy principle, but, as they themselves said, they were unable to ascribe a physical meaning to the result so obtained, and, moreover, they added, "We do not claim that this model is realistic. We discussed it only because of its mathematical simplicity and rigorous solvability". Here we show that this density is quite meaningful in terms of $C\text{-}(fBm)_n$. But a new question is then the following one? When is $C\text{-}(fBm)_n$ physically meaningful in systems modelling? The following section proposes possible answers.

8.7. On the practical meaning of the model

8.7.1. Modelling of collisions between n particles

Historically, Brownian motion appeared in statistical physics as follows. When small solid particles (about 10^{-4} cm in size) are introduced into a drop of liquid one observes that they do not remain at rest, but that they are constantly moving in a highly irregular way.

Following Einstein, a solid particle is subject to a fluctuating net force which results from the many random collisions of the particle with the molecules of the liquid. Given that the particle is small, the number of molecules with which it collides per unit time is small and thus fluctuates appreciably. In addition, since the mass of the particle also is small, any collision will have a sensible effect on it.

The basic assumption is that at each instant the particle will collide with one only molecule at a time (since it is small) and that the resulting impulse is +1 or -1, that is to say, the step vectors of the random walk which generates the Brownian motion. The idea is that, in so far as the size of the particle is small with respect to the mean distance between molecules, the former will cross through and will meet one molecule only.

Assume now that the particle can collide simultaneously with more than one molecules, say with n, $n \geq 2$, molecules. This can occur, for instance, when the size of the particle compares with that of the molecules. In such a case the angle of the

momentum of the particle after colliding will take on n possible values, and assuming that there is a complete symmetry, we shall assume that these angles are defined by the complex roots of unity, therefore in a quite natural way we are led to introduce the C-$(fBm)_n$.

It is well known that, up to now all attempts which have been made to apply statistical thermodynamics to human systems has met little, very little, sucess. We think that one possible reason for this defect is exactly that, in human systems, collisions between more than two individuals can occur and very often can initiate drastic changes in the dynamics of the system.

Clearly, the C-$(fBm)_n$ so introduced, should find application in humans ystems, as a way to take account of mutual interaction between more than two individuals.

8.7.2. Observation with measurement errors

In engineering mathematics Gaussian white noise appear either as disturbing additive signals or as measurement errors. Given a one-dimensional input $x(t)$ it is received in the final form

$$x'(t) \;:=\; x(t)+w(t), \tag{8.72}$$

or again, given a signal $x(t)$ under observation, it is measured in the form $x'(t)$ above.

For a 2D-signal (x,y), one is used to assume that the observation is in the form

$$x'(t) \;=\; x(t)+a_{11}(t)w_1(t)+a_{12}(t)w_2(t), \tag{8.73}$$
$$y'(t) \;=\; y(t)+a_{21}(t)w_1(t)+a_{22}(t)w_2(t), \tag{8.74}$$

where $w_1(t)$ and $w_2(t)$ are two Gaussian white noises which are not necessarily independent.

This model of observation is the most popular, and practically the only one which has been used in electrical engineering. Why? Probably because, on the surface, we have mathematical theories which yields deep results on this kind of equation. Nothing more.

Nevertheless, other models could be quite satisfactory as well. For instance one can use the multiplicative scheme

$$x'(t) \;=\; k(t)x(t) \quad , \quad x \in \Re, \tag{8.75}$$

or in the (x,y)-plane

$$(x' \quad y')^T = K(t)(x \quad y)^T, \tag{8.76}$$

where $K(t)$ is a 2×2 matrix and the superscript T denotes the transpose. This has been investigated by Kovanic [8.9] who selects $K(t)$ in the form

$$K_1(\theta(t)) = \begin{pmatrix} \cos\theta(t) & \sin\theta(t) \\ -\sin\theta(t) & \cos\theta(t) \end{pmatrix}, \tag{8.77}$$

or

$$K_2(\theta(t)) = \begin{pmatrix} \cosh\theta(t) & \sinh\theta(t) \\ \sinh\theta(t) & \cosh\theta(t) \end{pmatrix} \tag{8.78}$$

where $\theta(t)$ is a function parameter.

By using some arguments based on the mutual relations between syntax and semantics in natural languages, we had previously obtained exactly the same model as a basis for a possible theory of relative information [8.8]. In the following we shall see how this relative information theory is related to the present model of fractional Brownian motion, but before doing so we shall comment on the so called Heisenberg's principle.

8.8. Heisenberg's principle and the quantization principle

8.8.1. Review of the basic assumptions

For fixing our thoughts, assume that we are dealing with the dynamical system driven by the stochastic differential equation of order n,

$$dz(t) = f(z,t)dt + g(z,t)db(t,n) \quad , \quad z \in C \ , \tag{8.79}$$

where $b(t,n)$ is *a C-(fBm)$_n$* with moment $\sigma_n^n(t)$. Given that this equation is selected to describe a real physical system, how can we explain, or, again, which kinds of arguments could support the use of $b(t,n)$ instead of the usual $b(t,2)$?

In the following we shall characterize a model which is general enough to describe a broad class of practical real systems.

(H_1) Principle of complementary variables . We assume that the error term $db(t,n)$ results from the observation or the measurement of pairs (xj,yj), $j = 1,2,..,N$ of real valued variables, which are referred to as *complementary variables.* ■

Each pair is observed one at a time, and the corresponding results are aggregated in a summation to provide *db(t,n).*

(H2) Principle of invariance of uncertainty. We asssume that x_j and y_j are observed (or measured) with errors Δx_j and Δy_j which satisfy the invariance condition

$$(\Delta x_j)^2 + (\Delta y_j)^2 = \varepsilon^2, \tag{8.80}$$

where ε is a positive number of small magnitudes. ■

(H3) Quantization principle. We assume that Δx_j and Δy_j are quantized by the observation (or measurement) process, and we assume that this quantization can be modelled by the equation

$$\Delta x_j + i\Delta y_j = \varepsilon \exp\{2i\xi\pi / n\}, \tag{8.81}$$

where ξ is a random variable which takes on the integer values 1,2,..,n-1 with the respective probabilities 1/n, and n define the accuracy of the measurement. ■

8.8.2. Relation with Heisenberg's uncertainty principle

According to Heisenberg's uncertainty principle, for so-called pairs of conjugate variables *(x,y)* (such variables are defined by the special values of their commutator), for instance x and y may be respectively the position and the momentum of a particle, the corresponding measurement uncertainties cannot be reduced at will with an arbitrarily high accuracy.

If Δx and Δy denote the root mean square errors on x and y respectively, then one has necessarily the relation

$$\Delta x \Delta y \geq h / 2\pi, \tag{8.82}$$

where h is Planck's constant. This property is a consequence of Schwarz' inequality in linear algebra, and it is perhaps the main deviation of quantum mechanics from classical mechanics.

(i) Assume now that we can achieve the maximum accuracy possible to write the equality

$$\Delta x \Delta y = h / 2\pi; \tag{8.83}$$

then on taking the logarithm of both sides, we obtain the new equality

$$\ln \Delta x + \ln \Delta y \quad = \quad \ln h/2\pi \ . \tag{8.84}$$

(iii) Because of the smallness of the error terms on the one hand and of h on the other, these logarithms are negative, in such a manner that we can rewrite (8.84) in the form

$$\left[(-\ln \Delta x)^{1/2}\right]^2 + \left[(-\ln \Delta y)^{1/2}\right]^2 \quad = \quad \left[(-\ln(h/2\pi))^{1/2}\right]^2 \ , \tag{8.85}$$

which is exactly the equation (8.80).

Of course, the absolute values of these magnitudes are large, but it is always possible to introduce a scaling factor so as to have a small right side in equation (8.85).

In other words, if the Brownian motion is the result of an observation process, or of a measurement process which combines Heisenberg's principle and the quantization principle, then in a quite natural way we should arrive at complex-valued fractional Brownian motion.

References

8.1 ATLAN, H.; *Biological Organization and Information Theory* (in french), Hermann, Paris, 1972

8.2 BRILLOUIN, L.; *Science and Information Theory,* Academic Press, New York, 1956

8.3 GLANSDROFF, P. and PRIGOGINE, I.; *Thermodynamic Theory of Structure, Stability and Fluctuation*, Interscience Publications, New York, 1991

8.4 HAKEN, H.; *Information and Self-Organization: A Macroscopic Approach to Complex Systems,* Springer Verlag, Berlin, 1990

8.5 INGARDEN, R.S., KOSSAKOWSKI, A. and OHYA, M.; *Information Dynamics and Open Systems, Classical and Quantum Approach*, Kluwer, Dordrecht, 1997

8.6 JUMARIE, G.; A practical approach to nonlinear estimation by using the maximum entropy principle, *ASME Trans., J. Dynamic Systems Measurement and Control,* Vol 108, Series G, No 1, pp 49-56, 1986

8.7 JUMARIE, G.; Solution of the multivariate Fokker-Planck equation by using the maximum entropy principle, *J. Math. Phys.* , Vol 31, N0 10, pp 2389-2392, 1990

8.8 JUMARIE, G.; *Relative Information, Theories and Applications,* Springer Verlag, Berlin, 1990

8.9 KOVANIC, P.; A new theoretic and algorithmic basis for estimation, identification and control, *Automatica,* Vol 22, No 6, pp 657-674, 1986

8.10 LANDSBERG, P.T.; *Thermodynamics with Quantum Mechanical Illustrations,* Oxford University Press, Oxford, 1978

8.11 NICOLIS, G.; and PRIGOGINE, I.; *Self-organization in Non-equilibrium Systems*, Wiley and Sons, New York, NY, 1977

8.12 PRIGOGINE, I.; *Introduction to the Thermodynamics of Irreversible Processes* (in french) Dunod, Paris, 1968
8.13 SHANNON, C.E. and WEAVER, W.; *A Mathematical Theory of Communication*, University of Illinois Press, Chicago, 1949
8.14 SCHRODINGER, E.; *What is Life?* Cambridge University Press, Canbridge, 1946
8.15 Von FOERSTER, in Yovits and Cameron (Eds) , *On Self-Organizing Systems and Their Environments,* Pergamon Press, London, 1950

Chapter 9

Fractals, Path Entropy, and Fractional Fokker-Planck Equation

This long road which returns backward takes for ever, and this long road which goes forward also takes for ever.
Nietzsche (Zarathustra(III,2))

9.1. Introduction. Preliminary background.

9.1.1. Topological entropy of dynamical systems

Up to now we have defined path entropy of non-random functions and we have constructed a new model of complex-valued fractional Brownian motion of order n based on a random walk in the complex plane. One of the frequent features of these results is that they can be thought of as some consequences of the maximum entropy principle, therefore having direct relations with information thermodynamics. Our purpose in the present chapter is to go a step further and to look for other characteristics common between these two concepts, and in this way we shall put in evidence some connections between these models and fractals, and more especially with Hausdorff dimension and Liapunov exponent. In addition we shall show that there is a striking identity between the generalized heat equation of order $2n$ and the fractional Fokker-Planck equation of order $1/n$, in the special case when the mean value of the process is zero. We shall use this property to infer some results on the Fokker-Planck equation of fractal processes (or processes of fractional order) with non-zero mean values.. But before doing so, for the convenience of the reader, we shall call to mind some preliminary backgrounds.

Consider an ensemble of trajectory sections of duration T of a given dynamical system, with the property that they keep at a distance ε from each other. Clearly, any section of duration T of an arbitrary trajectory on the attractor lies in the ε-neighbourhood of at least one of the considered sections. Let $N(T,\varepsilon)$ denote the number of sections in the ensemble. Then the *topological entropy* H_{top} of these trajectories is defined by the expression (see for instance [9.2]

$$H_{top} := \lim_{\varepsilon \downarrow 0} \overline{\lim_{T \uparrow \infty}} \frac{\ln N(T,\varepsilon)}{T}. \tag{9.1}$$

G. Jumarie, *Maximum Entropy, Information Without Probability and Complex Fractals*
© Springer Science+Business Media Dordrecht 2000

Notice that the quantity $N(T,\varepsilon)$ is exactly the Shannon entropy of the random variable defined by the uniform probability $1/N(T,\varepsilon)$.

9.1.2. Kolmogorov entropy.

The idea is to generalize the Shannon entropy of random variables in order to deal with the uncertainty involved in the trajectory or path $x(t)$ generated by a dynamical system, in the phase plane, which can be finite-dimensional or infinite-dimensional.

Basically, the dynamical system under consideration is deterministic, in such a way that, in order to be fully consistent with the definition of Shannon entropy of random variables, it is necessary to randomize the problem in some sense, and to define the corresponding random events. This can be done as follows.

Consider the trajectory $x(t) = \left[x_1(t), x_2(t), ..., x_d(t)\right]$ of a dynamical system *close to a strange attractor* (this point is of importance) and assume that the d-dimensional phase space is partitionned into boxes b_i of size ε^d, whilst we consider the state of the system at intervals of time τ. We so define a sequence of instants of observations $t_0, t_1, ..., t_n; t_{i+1} - t_i = \tau$; and we are thus led to define the entropy of the finite $(n+1)d$-dimensional vector $\left[x(t_0), x(t_1), ..., x(t_n)\right]$.

To this end, assume that the subscript i is the index vector $i = (\alpha_1, \alpha_2, ..., \alpha_n)$; and consider the event $x(t_i) \in b_i$. Since we are close to a strange attractor each box is visited with a certain frequency, and we can define the probability $p(b_{i_0}, b_{i_1}, ..., b_{i_n})$ so that $x(t_0) \in b_{i_0}$, $x(t_1) \in b_{i_1}$, ..., and $x(t_n) \in b_{i_n}$. Then, following Shannon, we can introduce the entropy

$$K_n := -\sum_{i_0,...,i_n} p(b_{i_0},...,b_{i_n}) \ln p(b_{i_0},...,b_{i_n}). \tag{9.2}$$

We then arrive at the Kolmogorov entropy as follows. Since we consider the trajectory close to a strange attractor, we can assume that the system exhibits the property of ergodicity, in accordance with which the time average is equal to the mathematical expectation, and we can therefore meaningfully refer to the following definition.

Consider the trajectory $x(t)$ of a dynamical system close to a strange attractor A; then its *Kolmogorov entropy* is defined by the expression (see for instance [9.11]

$$K := -\lim_{\tau \downarrow 0} \lim_{\varepsilon \downarrow 0} \lim_{N \uparrow \infty} \frac{1}{N\tau} \sum_{i_0,...,i_N} p(b_{i_0},..,b_{i_N}) \ln p(b_{i_0},..,b_{i_N}), \tag{9.3}$$

where the limit on ε has to be taken after the limit on N.

By computing the difference K_{n+1} - K_n, one can rewrite K in the form

$$K = \lim_{\tau \downarrow 0} \lim_{\varepsilon \downarrow 0} \lim_{N \uparrow \infty} \frac{1}{N\tau} \sum_{n=0}^{N-1} (K_{n+1} - K_n), \tag{9.4}$$

which clearly shows that it defines an average loss of information. For a discrete trajectory, with discrete time step τ=1, the limit $\tau \rightarrow 0$ is omitted.

K = 0 for a periodic/quasi-periodic system; for a chaotic dynamics one has $0 \lhd K \lhd +\infty$, and $K \uparrow +\infty$ for a random system. K is positive (or zero), and this is because, basically, it can be identified with a transinformation or mutual information in the Shannon sense.

9.1.3. Hausdorff dimension

Let A be a subset of a metric space and let $n(\varepsilon)$ denote the number of sets of diameter ε needed to cover A. The *Hausdorff dimension* $D(A)$ of A is then defined by the expression

$$D(A) := \lim_{\varepsilon \downarrow 0} - \frac{\ln n(\varepsilon)}{\ln \varepsilon} . \tag{9.5}$$

In the information-theoretic framework this expression can be thought of as the quotient of a discrete entropy by a continuous entropy. Indeed, let $C(\varepsilon)$ denote the discrete random variable with the uniform probability $p_i = 1/n(\varepsilon)$, and in a like manner let $X(\varepsilon)$ denote the continuous random variable defined on the diameter of the circle, with the uniform density $1/\varepsilon$. Then one has the entropies

$$H(C(\varepsilon)) = \ln n(\varepsilon), \tag{9.6}$$

and

$$H(X(\varepsilon)) = \ln \varepsilon, \tag{9.7}$$

in such a manner that $D(A)$ can be rewritten in the form

$$D(A) = \lim_{\varepsilon \downarrow 0} - \frac{H(C(\varepsilon))}{H(X(\varepsilon))}. \tag{9.8}$$

9.1.4. Liapunov exponent

Let us consider the one-dimensional map

$$x_{n+1} = f(x_n) \quad , \quad n = 0,1,2,.., \tag{9.9}$$

where *f(x)* is a differentiable function; then its *Liapunov exponent* $\lambda(x_0)$ is defined by the expression (see for instance [9.11])

$$\lambda(x_0) := \lim_{N\uparrow\infty} \frac{1}{N} \sum_{i=0}^{N-1} \ln \left| f'(x_i) \right| . \tag{9.10}$$

Usually, the practical significance of $\lambda(x_o)$ is displayed as follows. Assume that the dynamics starts from two different initial states x_o and $x_o+\varepsilon$, then the deviation of the corresponding states $x_n(x_o)$ and $x_n(x_o+\varepsilon)$ will satisfy the condition

$$\left| x_n(x_o + \varepsilon) - x_n(x_o) \right| \cong \varepsilon \exp\{n\lambda(x_0)\}, \tag{9.11}$$

for n large enough. Clearly the Liapunov exponent is quite relevant whenever one analyzes a dynamical system close to a given nominal reference trajectory.

9.1.5. Monkey model of image entropy

General definition

Analogously with the entropy of random variables, it has been suggested to define the entropy of an image as follows (see for instance [9.6]

Let be given an image P (P for picture) defined on the domain $(x,y) \in \Omega \subset R^2$ and characterized by the brightness $b(x_i, y_j)$ for all *(i,j)*. Its image entropy is then defined by the expression

$$H(P,b) := - \sum_{i,j} b(x_i, y_j) \ln b(x_i, y_j) \quad , \quad (x,y) \in \Omega . \tag{9.12}$$

In the continuous case one will set

$$H(P,b) = - \int_{\Omega} b(x,y) \ln b(x,y) dxdy . \tag{9.13}$$

This definition appeals to a few comments. Indeed, given a random discrete scalar-valued random variable X with the probability distribution p_i, $i = 1,...,n$; its entropy is given by the expression

$$H(X) = -\sum_{i=1}^{n} p_i \ln p_i ; \tag{9.14}$$

but if we consider the entropy defined by the incomplete distribution $(p_j,..,p_{j+m})$, then its expression is

$$H(X; x_j \le X \le x_{j+m}) = -\sum_{i=j}^{j+m} p_i \ln p_i \Big/ \sum_{i=j}^{j+m} p_i . \tag{9.15}$$

Similarly, for a continuous random variable with the probability density $p(x)$, the corresponding mean value of uncertainty on the interval (a,b) is

$$H(X; a \le X \le b) = -\int_a^b p(x) \ln p(x) dx \Big/ \int_a^b p(x) dx . \tag{9.16}$$

As a result, it seems that a definition of $H(P,b)$ which would be fully consistent with the entropy of incomplete probability distribution (equ. (3.4) should be the average

$$. \ \overline{H}(P,b) := -\sum_{i,j \in \Omega} b(x_i, y_j) \ln b(x_i, y_j) \Big/ \sum_{i,j} b(x_i, y_j) \tag{9.17}$$

$\overline{H}(P,b)$ is the mean value of uncertainty involved in the image, while $H(P,b)$ is the total uncertainty in this image.

Photon / Unit Grey Level Allocation Model

The expression (9.12) can be obtained by viewing a digital image as a probability distribution with each pixel's grey level representing the number (proportional to the probability) of photons reaching this point [9.1]. We can assume that we have a number B of photons and an initial image in the form of an empty grid comprising n cells (pixels). The B photons are allocated one at a time among the n cells with uniform spatial probability [9.1]. This model is sometimes referred to as the monkey model, in analogy with a group of monkeys which randomly throw B balls into a two-dimensional array of boxes to form the image [9.3,9.4,9.12]. Each ball represents a unit grey level and each box a pixel, see Fig. 1 for the one-dimensional case

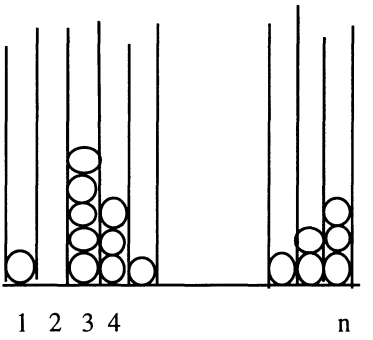


Fig.1 Photon / unit grey-level allocation model

The number of ways W that an image $(b_1,b_2,...,b_n)$ can be formed is

$$W(b_1,...,b_n) = B! \ / \ b_1!...b_n! \quad , \tag{9.18}$$

and in the classical way, using Stirling's approximation $n! = n^n e^{-n}(2\pi n)^{1/2}$, one finds that

$$\ln W \cong H(P) \tag{9.19}$$

9.2. Path entropy and Hausdorff dimension

Assume that f(.) is a continuous mapping $\Re \to \Re$ which satisfies the increment condition

$$|\Delta f(x)| = \rho(x)(\Delta x)^{\alpha}, \quad 0 \lhd \alpha \lhd 1, \tag{9.20}$$

$\Delta x \rhd 0$, where $\rho(x)$ is a positive bounded function, $0 \lhd \rho(x) \lhd \rho_M$. Let the stepwise approximation function $\hat{f}$ be defined by the span Δx, that is to say,

$$\hat{f}_j := f(x_0 + j\Delta x). \tag{9.21}$$

With these preliminary notations we can meaningfully refer to the entropy of non-random piece wise continuous functions (see Section 3.10), and consider the quotient

$$\hat{D}_c(\Delta x) \;:=\; \frac{H_c(\hat{f}(.);1,N)}{\ln \Delta x}. \tag{9.22}$$

which provides the relation

$$\lim \;\; \hat{D}_c(\Delta x) \;=\; \alpha \;\; as \;\; \Delta x \downarrow 0. \tag{9.23}$$

Indeed, according to the definition of H_c there is a mean value $\bar{x}$ such that

$$\begin{aligned} H_c(\hat{f}(.);1,N) \;&\cong\; \ln\left[\rho(\bar{x})(\Delta x)^{\alpha}\right], \\ &=\; \ln \rho(\bar{x}) + \alpha \ln \Delta x, \end{aligned} \tag{9.24}$$

therefore (9.23).

Now, we notice that α is related to the fractal dimension D by the equation [9.8]

$$\alpha \;=\; 2 - D$$

and we then have the equality

$$\lim \;\; \hat{D}_c(\Delta x) \;=\; 2 - D \;\; as \;\; \Delta x \downarrow 0, \tag{9.25}$$

which exhibits a connection between the path entropy of non-random functions and fractal dimension.

9.3. Path entropy and Liapunov exponent

Preliminary remarks on the informational meaning of $\lambda(x_o)$.

We refer to the subsection 9.1.4. related to the Liapunov exponent, we introduce the deviation

$$y_n \;:=\; x_n(x_o + \varepsilon) - x_n(x_0), \tag{9.26}$$

and we consider the linearized dynamics

$$y_{n+1} = f'(x_n)y_n, \tag{9.27}$$

where $f'(x_n)$ holds for $f'(x_n(x_0))$. Assume that, owing to some uncertainty in its exact value, y_o can be thought of as a random variable. It follows that y_n also can be considered as a random variable, and on taking the Shannon entropy of both sides of equation (9.27) we obtain the equality

$$H(y_{n+1}) = H(y_n) + \ln|f'(x_n)|, \tag{9.28}$$

therefore

$$H(y_{n+1}) = H(y_o) + \sum_{j=0}^{n} \ln|f'(x_j)|, \tag{9.29}$$

$$= H(y_o) + (n+1)\lambda(x_0). \tag{9.30}$$

One can arrive at another significance for the Liapunov exponent as follows. If we calculate the mean value,

$$\overline{H}(y_j; 0 \le j \le N) := \frac{1}{N+1}\sum_{j=0}^{N} H(y_j), \tag{9.31}$$

then one easily obtains the equality

$$\overline{H}(y_j; 0 \le j \le N) = H(y_0) + \frac{N}{2}\lambda(x_o), \tag{9.32}$$

therefore

$$\lim_{N\uparrow\infty} \frac{1}{N}\overline{H}(y_j; 0 \le j \le N) = \frac{1}{2}\lambda(x_o). \tag{9.33}$$

Application to entropy of non random functions

(i) As a first application we shall consider the path entropy of order zero, $c=0$, and in order to obtain simple formulae we shall consider the linear equation

$$y_{n+1} = \rho y_n, \tag{9.34}$$

which yields

$$y_n = \rho^n y_0, \tag{9.35}$$

and

$$\Delta y_n = \rho^n(\rho-1)y_0\ . \tag{9.36}$$

The path entropy of order zero ($c = 0$) of the trajectory is

$$H_o(y(.);0,N) = \frac{1}{N+1}\sum_{j=0}^{N}\ln\left|\rho^j(\rho-1)y_0\right|, \tag{9.37}$$

$$= \ln\left|(\rho-1)y_0\right| + \frac{1}{N+1}\sum_{j=0}^{N} j\ln|\rho|$$

$$= \ln\left|(\rho-1)y_0\right| + \frac{N}{2}\ln|\rho|. \tag{9.38}$$

(ii) We now consider the path entropy of order one, which reads

$$H_1(y(.);0,N) = \frac{\sum_{j=0}^{N}\left|\rho^j(\rho-1)y_0\right|\ln\left|\rho^j(\rho-1)y_0\right|}{\sum_{j=0}^{N}\left|\rho^j(\rho-1)y_0\right|}, \tag{9.39}$$

$$= \ln\left|(\rho-1)y_0\right| + \ln|\rho|\frac{\sum_{j=0}^{N} j|\rho|^j}{\sum_{j=0}^{N}|\rho|^j}, \tag{9.40}$$

$$= \ln\left|(\rho-1)y_0\right| + \frac{|\rho|\ln|\rho|}{1-|\rho|} - (N+1)\frac{|\rho|^{N+1}\ln|\rho|}{1-|\rho|^{N+1}},$$

therefore the limit

$$\ln|(\rho-1)y_0| + \frac{|\rho|\ln|\rho|}{1-|\rho|} \quad , \quad |\rho| \lhd 1, \tag{9.41}$$

$$\lim_{N\uparrow\infty} H_1(y(.);0,N) \quad = \quad +\infty \quad , \quad |\rho \rhd 1|. \tag{9.42}$$

Relation with Liapunov exponent.

(i) In order to explicitly introduce the Liapunov exponent we shall set $\rho = \exp\{\lambda\}$, here λ is the Liapunov exponent, in such a manner that equation (9.38) can be rewritten in the form

$$H_o(y(.);0,N) \quad = \quad \ln|y_0| + \ln|\rho - 1| + \frac{N}{2}\lambda. \tag{9.43}$$

As a result, if we refer to the equation (9.32) and make the identification

$$H(y_0) \quad \equiv \quad \ln|y_o| \tag{9.44}$$

we then obtain the equality

$$H_0(y(.);0,N) \quad = \quad \overline{H}(y_j;0 \le j \le N) + \ln|\rho - 1| \tag{9.45}$$

which provides

$$\lim_{N\uparrow\infty} \frac{1}{N} H_0(y(.);0,N) \quad = \quad \lim_{N\uparrow\infty} \frac{1}{N} \overline{H}(y_j;0 \le j \le N)$$
$$= \quad \lambda. \tag{9.46}$$

The entropies H_o and $\overline{H}$ vary in the same way: both ones increase with N when $|\rho| \rhd 1$, and they decrease when $|\rho| \lhd 1$.

(ii) In contrast, $H_1(y(.);0,N)$ is always increasing with N, but it is bounded when $|\rho| \lhd 1$ and unbounded when $|\rho| \rhd 1$.

(iii) H_0 and H_1 have different behaviours depending upon the stability of the system. H_0 decreases indefinitely with increasing N (stability) or increases indefinitely with N (instability) whilst H_1 either has a limited value (stability) or increases indefinitely (instability).

(iv) H_1 is a mean value of information, such that, according to its definition (9.39) the total amount of uncertainty involved in the trajectory generated by the dynamics (9.34), when it is stable (i.e., $|\rho| \lhd 1$) is

$$\begin{aligned} H_{total}(y(.);0,\infty) &= H_1(y(.);0,\infty)\sum_{j=0}^{\infty}\left|\rho^j(1-\rho)y_0\right|, \\ &= |y_0|\frac{|1-\rho|}{1-|\rho|}\left[\ln\left|(\rho-1)y_0\right| + \frac{|\rho|\ln|\rho|}{1-|\rho|}\right]. \end{aligned} \tag{9.47}$$

9.4. Path entropy of random functions and fractals

9.4.1. Application to Brownian motion

Consider the Brownian motion $b(t,2) = b(t)$ with zero mean and the variance $\sigma^2(t)$. Formally, the derivative of $b(t)$ is the Gaussian white noise $w(t) \approx N(0,\sigma^2(t))$.

(i) A simple calculation provides

$$E\{|w(t)|\} = 2(2\pi)^{-1/2}\sigma(t), \tag{9.48}$$

$$E\{|w(t)|\ln|w(t)|\} = (2\pi)^{-1/2}\sigma(t)[4\ln 2\sigma(t) - 2\gamma], \tag{9.49}$$

where γ is the Eulerian constant; therefore the approximation (see subsection 4.2.2 for the definition of the average entropy)

$$\left\langle H_1^1(b(.);0,T)\right\rangle \cong 2\frac{\int_0^T \sigma(t)\ln\sigma(t)dt}{\int_0^T \sigma(t)dt} + 2(\ln\ 2 - \gamma). \tag{9.50}$$

In the special case when $\sigma(t) = \sigma =$ constant, one has

$$\left\langle H_1^1(b(.);0,T)\right\rangle \cong 2\ln\sigma + 2(\ln\ 2 - \gamma). \tag{9.51}$$

(ii) The average entropy of order 2 is

$$\left\langle H_1^2(b(.);0,T)\right\rangle \cong \frac{\int_0^T \sigma(t)\ln\sigma(t)dt}{\int_0^T \sigma(t)dt}, \tag{9.52}$$

which, in the special case when $\sigma(t) = \sigma =$ constant, turns out to be

$$\left\langle H_1^2(b(.);0,T)\right\rangle \cong \ln \sigma. \tag{9.53}$$

(iii) We can now compare these results with those provided by Shannon entropy. Indeed we have (H_{sh} emphasizes that we are dealing with the Shannon entropy of random variables)

$$H_{sh}(w(.);0,T) = \int_0^T \ln(\sigma(t)\sqrt{2\pi})dt, \tag{9.54}$$

$$= T\ln\sqrt{2\pi} + \int_0^T \ln\sigma(t)dt \quad , \tag{9.55}$$

and this results is quite consistent with (9.51) and (9.53) when the variance is constant.

9.4.2. Application to real-valued fractional Brownian motion

We now consider a fractional real-valued Brownian motion with zero mean and the variance

$$E\left\{db_a^2(t)\right\} = \sigma^2(dt)^{2a}. \tag{9.56}$$

Here we cannot use the formal derivative of $b_a(t)$ to calculate the corresponding path entropy, since it is not meaningful even from an engineering mathematics point of view, and to circumvent this pitfall we shall use a discrete approximation with a small discretizing span τ, to obtain

$$\left\langle H_1^2(b_a(.);0,T)\right\rangle \cong \frac{\sum_{i=1}^{i=n}\tau\sigma(t_i)\ln\sigma(t_i)}{\sum_{i=1}^{i=n}\tau\sigma(t_i)} + \left(a-\frac{1}{2}\right)\ln\tau \ , \tag{9.57}$$

$$\cong \left\langle H_1^2(b(.,1/2);0,T\right\rangle + \left(a-\frac{1}{2}\right)\ln\tau. \tag{9.58}$$

We then have the limiting relation

$$\lim_{\tau\downarrow 0} \frac{\left\langle H_1^2(b_a(.);0,T,\tau)\right\rangle}{\ln\tau} = a-\frac{1}{2}, \tag{9.59}$$

which looks like the equation (9.5) defining Hausdorff dimension.

In the following we shall see how these considerations apply to $C\text{-}(fBm)_n$, but before doing so, we shall once more examine the problem of defining the path entropy of a trajectory in the complex plane.

9.5. Path entropy of time functions in the complex plane

In the chapter 3 we have defined the entropy of holomorphic functions in the complex plane, and here we shall more especially consider the entropy of C-valued function of time. This will be made by using the rationale which allowed us to defined the entropy of non-random functions, and in order to emphasize this slight discrepancy we shall sometimes refer to *path entropy.*

Here again, in computer programming, the notations refer to their local environment, and $z(t)$ will represent the state of a system in the complex plane (it is not the distributed parameter of Chapter 4!)

Our purpose in the present subsection is to define the path entropy of the complex function $z(t) = x(t) + iy(t)$, where t denotes the time. Strictly speaking, $z(t)$ is a vector (we cannot consider $z(t)$ as the pair of functions $(x(t),y(t))$, and in order to obtain the expression of its path entropy, we have to duplicate the rationale which provides the entropy of non random $\Re \rightarrow \Re$ functions.

To this end we shall read the first-order Taylor expansion

$$z(t+dt) = z(t)+\dot{z}(t)dt+o(dt), \tag{9.60}$$

as the vector equation

$$M(t+dt) = M(t)+\dot{M}(t)dt+o(dt),$$

or

$$dM = \dot{M}(t)dt. \tag{9.61}$$

For a given dt the uncertainty involved in the length of dM is $\ln\left|\dot{M}(t)\right|$,

$$\ln\left|\dot{M}(t)\right| = \ln|\dot{z}(t)| = \ln\left|\frac{ds}{dt}\right|, \tag{9.62}$$

where $s(t)$ is the length of the path generated by $(x(t),y(t))$. Thus we arrive at the following:

DEFINITION 9.1 A measure of the amount of uncertainty involved in the $\Re \to C$ complex valued function $z(t) = x(t)+iy(t)$ on the time interval $(0,T)$, and which is fully consistent with Shannon entropy of non-random variable and entropy of non random functions, is given by the expression

$$H_c(z(.);0,T) = \frac{\int_0^T |\dot{z}(t)|^c \ln|\dot{z}(t)|dt}{\int_0^T |\dot{z}(t)|^c dt}, \tag{9.63}$$

In terms of the arc length we shall have

$$H_c(z(.);0,T) = \frac{\int_0^T (\dot{s}(t))^c \ln \dot{s}(t)dt}{\int_0^T (\dot{s}(t))^c dt}. \blacksquare \tag{9.64}$$

In the following we shall extend and adapt this definition to stochastic processes.

9.6. Path entropy of complex-valued stochastic processes

The term path entropy refers to the entropy of the stochastic process considered as a function of time; in other words it is the entropy of the stochastic process on a given time interval, in contrast to Shannon entropy which is the entropy at a given instant t.

Consider the trajectory generated by the stochastic process $z(t) \in C$ in the time interval $[0,T]$. Basically, the amount of uncertainty it involves will depend upon the way in which it is observed, and we shall have to consider the observation of one trajectory alone, the serial observation of a family of trajectories and the parallel observation of a family of trajectories (see Chapter 4 for the corresponding definitions). Therefore the following entropies:

Path sample entropy

In this case, as mentioned above, we shall consider the single (recorded) trajectory under consideration as a deterministic function, to write directly

$$H_c(z(.);0,T) := \frac{\int_0^T |\dot{z}(t)|^c \ln |\dot{z}(t)|dt}{\int_0^T |\dot{z}(t)|^c dt}. \tag{9.65}$$

Serial path entropy

The above path sample entropy is a random variable, and thus we shall consider the amount of uncertainty it defines on the family of paths as the mathematical expectation,

$$H_{c,s}(z(.);0,T) := E\{H_c(z(.);0,T)\}. \tag{9.66}$$

In many applications it is not easy to calculate this average in such a way that one will look for an approximation which can be suficient in practice. For instance, one can refer to the expression

$$\hat{H}_{c,s}(z(.);0,T) := \frac{E\left\{\int_0^T |\dot{z}(t)|^c \ln|\dot{z}(t)|dt\right\}}{E\left\{\int_0^T |\dot{z}(t)|^c dt\right\}}. \tag{9.67}$$

Parallel path entropy

Here the density of entropy is then $E\{\ln|\dot{x}(t)|\}$, and we have

$$H_{c,p}(z(.);0,T) := \frac{\int_0^T e^{cE\{\ln|\dot{z}(t)|\}} E\{\ln|\dot{z}(t)|\}dt}{\int_0^T e^{cE\{\ln|\dot{z}(t)|\}} dt}. \tag{9.68}$$

Here again, this expression is not very easy to use in practical problems, so it will be of interest to look for some approximation. For instance, one can define the average moment

$$\langle \dot{m}_k(t)\rangle := E\left\{|\dot{z}(t)|^k\right\}^{1/k}, \quad k = 1,2,3,.., \tag{9.69}$$

and consider the estimate

$$\hat{H}^k_{c,p}(u(.);0,T) := \frac{\int_0^T |\langle \dot{m}_k(t)\rangle|^c \ln|\langle \dot{m}_k(t)\rangle| dt}{\int_0^T |\langle \dot{m}_k(t)\rangle|^c dt}. \tag{9.70}$$

9.7. Path entropy of fractional stochastic processes

9.7.1. Path entropy of fractional Brownian motion

In order to see how the results of the preceding section can be extended to fractional stochastic processes, we shall first consider the process $z(t,n)$ defined by the equation

$$dz(t,n) = f_n(z,t)w(t,n)(dt)^{1/n}, \tag{9.71}$$

(i) We firstly recall that $E\{(dz(t,n))^j\} = 0$ when $j = 1,2,...,n-1$; in such a manner that if we want to define a mean value for $dz(t,n)$, we have to go up to the n-th moment and to consider the quantity

$$\begin{aligned} \lhd dz(t,n) \rhd &= E^{1/n}\left\{(dz(t,n))^n\right\}, \\ &= E^{1/n}\left\{f_n^n(z,t)w^n(t,n)\right\}, \\ &= \left|\sigma_n(t)f_n(w,t)\right|(dt)^{1/n}. \end{aligned} \tag{9.72}$$

(ii) In this expression, the term $(dt)^{1/n}$ will be merely considered as the scaling span of time, in such a manner that the density of uncertainty involved in the average of $dz(t,n)$ is $\ln\left|\sigma_n(t)f_n(z,t)\right|$. We then arrive at an estimate of the parallel path entropy in the form

$$\hat{H}_{c,p}(u(.n);0,T) = \frac{\int_0^T \left|\sigma_n f_n\right|^c \ln\left|\sigma_n f_n\right| dt}{\int_0^T \left|\sigma_n f_n\right|^c dt}. \tag{9.73}$$

9.7.2. Path entropy of fractional stochastic dynamics

Preliminary remarks

For fixing our thoughts, let us consider the scalar equation

$$dx = f_1(t)dt + f_2(t)w(t,2)(dt)^{1/2}. \tag{9.74}$$

(i) In this special case, one can split dx into the form

$$dx = dx_1 + dx_2, \tag{9.75}$$

$$dx_1 = f_1(t)dt, \tag{9.76}$$

$$dx_2 = f_2(t)w(t,2)(dt)^{1/2}, \tag{9.77}$$

and we are then led to consider the following two entropies

$$H_c(x_1(.);0,T) = \frac{\int_0^T |f_1(t)|^c \ln |f_1(t)| dt}{\int_0^T |f_1(t)| dt}, \tag{9.78}$$

and

$$H_c(x_2(.);0,T) = +\frac{\int_0^T |\sigma_2 f_2|^c \ln |\sigma_2 f_2| dt}{\int_0^T |\sigma_2 f_2|^c dt}. \tag{9.79}$$

$H_C(x_1(.))$ defines the amount of uncertainty involved in the average trajectory of the process, whilst $H_C(x_2(.))$ measures the uncertainty involved in the deviation from this mean trajectory.

A little bit more formally, one way to support this approach is to remark that one has the following equalities

$$E\{dx\} = f_1(t)dt, \tag{9.80}$$

$$E\{dx^2\} = f_2^2(t)\sigma_2^2(t)dt + o((dt)^2), \tag{9.81}$$

therefore the expressions (9.78) and (9.79).

Extension to nonlinear stochastic dynamics

The remark above can be directly extended to non-linear fractional stochastic dynamics, and in order to illustrate how this can be done we shall consider the process

$$\begin{aligned} dz &= f_1(z,t)dt + f_2(z,t)w(t,2)(dt)^{1/2} + f_3(t,3)w(t,3)(dt)^{1/3}, \\ &= dz_1(t) + dz_2(t) + dz_3(t). \end{aligned} \tag{9.82}$$

A simple calculation provides the conditional expectations

$$E\{dz|z\} = f_1(z,t)dt, \tag{9.83}$$

$$\begin{aligned} E\{dz^2|z\} &= f_1^2(dt)^2 + (\sigma_2 f_2)^2 dt \\ &= (\sigma_2 f_2)^2 dt + o((dt)^2), \end{aligned} \tag{9.84}$$

$$E\{dz^3|z\} = f_1^3(dt)^3 + f_1(f_2\sigma_2)^2(dt)^2 + (f_3\sigma_3)^3 dt$$
$$= (\sigma_3 f_3)^3 dt + o((dt)^2), \tag{9.85}$$

therefore the entropies

$$H_c(z_1(.);0,T) = \frac{\int_0^T |f_1(z,t)|^c \ln |f_1(z,t)| dt}{\int_0^T |f_1(z,t)|^c dt}, \tag{9.86}$$

and

$$H_c(z_j(.);0,T) = \frac{\int_0^T |\sigma_j f_j(z,t)|^c \ln |\sigma_j f_j(z,t)| dt}{\int_0^T |\sigma_j f_j(z,t)|^c dt}, \quad j = 2,3. \tag{9.87}$$

9.8. Hausdorff dimension and path entropy of fractional white noise

Refer to the process $z(t,n)$ defined by the equation (9.71), and assume that we approximate its trajectory by the stepwise function $\hat{z}(t,n)$ defined by the span Δt, clearly

$$\hat{z}(t_j,n) = z(t_0 + j\Delta t, n). \tag{9.88}$$

The path entropy of this stepwise function is (see Chapter 3)

$$H_c(\hat{z}(.,n);1,N) = \frac{\sum_{j=1}^{N-1} \left|\langle \Delta z(t_j,n)\rangle\right|^c \ln \left|\langle \Delta z(t_j,n)\rangle\right|}{\sum_{j=1}^{N-1} \left|\langle \Delta z(t_j,n)\rangle\right|^c}, \tag{9.89}$$

and if we consider the quotient

$$\hat{D}_c(z,\Delta t) := \frac{H_c(\hat{z}(.,n);1,N)}{\ln \Delta t}, \tag{9.90}$$

then, according to equation (9.72) we have the relation

$$\lim \quad \hat{D}_c(z,\Delta t) \quad = \quad 1/n \qquad as \qquad \Delta t \downarrow 0. \tag{9.91}$$

If we bear in mind that the limit $1/n$ is related to the fractal dimension D of the path by the equation

$$\frac{1}{n} \quad = \quad 2 - D, \tag{9.92}$$

one then has again the equality (see equation (9,25))

$$\lim \quad \tilde{D}_c(z,\Delta t) \quad = \quad 2 - D \qquad as \qquad \Delta t \downarrow 0. \tag{9.93}$$

which relates path entropy and fractal dimension in the complex plane.

9.9. Fractional Fokker-Planck equation and difference of fractional order

9.9.1. Fractional derivatives and fractional difference

DEFINITION 9.2. Let f: $\Re \to \Re$, $x \to f(x)$ denote a continuous function, then its fractional derivative $f^{(\alpha)}$ of order α , $\alpha \in R$, $\alpha \notin N$ is defined by the following expression (see for instance [9.10]

$$f^{(\alpha)}(x) \quad := \quad \frac{1}{\Gamma(-\alpha)} \int_0^x (x-u)^{-\alpha-1} f(u)du \quad , \quad \alpha \lhd 0 \tag{9.94}$$

$$\left[f^{(-n+\alpha)}(x) \right]^{(n)} \quad , \quad \alpha \rhd 0, \tag{9.95}$$

where n denotes the smallest integer such that $n > \alpha$. ■

DEFINITION 9.3. Consider the function $f(x)$ of the Definition 9.2, and let $h \rangle 0$ denote a discretizing span; then the fractional difference $\Delta^{\alpha} f(x)$ of order α of $f(x)$ is defined by the expression

$$\Delta^{\alpha} f(x) \quad := \quad \sum_{j=0}^{\infty} (-1)^j \binom{\alpha}{j} f[x + (\alpha - j)h]. \tag{9.96}$$

An alternative to the definition 9.2. of fractional derivative is the limit

$$f^{(\alpha)}(x) = \lim \frac{\Delta^{\alpha} f(x)}{h^{\alpha}} \quad as \quad h \downarrow 0. \tag{9.97}$$

It is clear that the fractional derivative defined by (9.97) involves values of *f(.)* on *[-∞,0]*, in such a manner that, to be fully consistent with (9.96), the expression (9.94) should be modified in the form

$$f^{(\alpha)}(x) = \frac{1}{\Gamma(-\alpha)} \int_{-\infty}^{x} (x-u)^{-\alpha-1} f(u) du, \tag{9.98}$$

as suggested in [9.7], mainly to lessen the effect of the value of f(0).

So if we describe a dynamical system by the fractional differential equation

$$x^{(\alpha)}(t) = f(x,t) \tag{9.99}$$

it is clear that according to (9.96) we *implicitly assume that x(t) is not Markovian*, in the sense that $x(t + dt)$ does not depend upon $x(t)$ only.

9.9.2. Fokker-Planck of fractional order

A few years ago [9.5], we introduced a fractional Fokker-Planck equation (FPE in the following) in the form

$$\frac{\partial^{\alpha} p}{\partial t^{\alpha}}(x,t) = -\frac{\partial}{\partial x}(fp) + \frac{1}{2}(-1)^{\alpha+1} \frac{\partial^2}{\partial x^2}(gp) \quad , \quad \alpha = \frac{1}{n} \tag{9.100}$$

where $f(x,t)$ and $g(x,t)$ are functions which have suitable mathematical continuity properties in order that the solution $p(x,t)$ may exist. This equation defines the probability density of a stochastic process satisfying the transition conditions

$$E\{d^{\alpha} x | x,t\} = f(x,t) dt^{\alpha}, \tag{9.101}$$

$$E\{(d^{\alpha} x)^2 | x,t\} = (-1)^{\alpha+1} g(x,t) dt^{\alpha}. \tag{9.102}$$

where $d^{\alpha}x$ is the increment of order α (see Equation (9.96)).

Multiplying both sides of (9.100) by x and x^2 respectively and integrating by parts, we obtain the dynamical equation of the mathematical expectations $m_1(t) := E\{x(t)\}$ and $m_2(t) := E\{x^2(t)\}$ in the form of the fractional differential equations

$$m_1^{(\alpha)}(t) = \int_R f(x,t)p(x,t)dx, \tag{9.103}$$

$$m_2^{(\alpha)}(t) = \int_R \left[2xf(x,t)+(-1)^{\alpha+1}g(x,t)\right]p(x,t)dx. \tag{9.104}$$

Important remark It is important to point out that equations (9.101) and (9.102) cannot be replaced by

$$E\{dx|x,t\} = f(x,t)dt^{\alpha}$$

and

$$E\{dx^2|x,t\} = (-1)^{\alpha+1}g(x,t)dt^{\alpha}.$$

Indeed, in such a case the process would be fully Markovian, and as a result we would be entitled to apply the Chapman-Kolmogorov equation

$$p(x',t+\tau) = \int_R p(x',t+\tau|x,t)p(x,t)dx,$$

to obtain

$$p(x',t+\tau) = p(x',t)+\alpha\tau^{\alpha}\left[-\partial_{x'}(fp)+\frac{1}{2}\partial_{x'x'}(gp)\right].$$

But the difference $\Delta p/\tau^{\alpha}$ does not tends to the fractional derivative $\partial^{\alpha}p/\partial t^{\alpha}$.

In other words, the fractional FPE describes a stochastic process of which the increments $x(tj+\tau)$-$x(tj)$ are not mutually independent.

The question which then arises is the following. Is there any common feature between the fractional stochastic process described by the fractional FPE (9.100) and the *C-(fBm)*$_n$ associated with the heat equation of order n? We are examining this problem in the following section.

9.10. On a relation between Fokker-Planck equation of fractional order and heat equation of order $2n$

We consider the special case $f(x,t) \equiv 0$ in (9.100) to obtain the fractional FPE

$$\begin{aligned} \frac{\partial^\alpha}{\partial t^\alpha} p(x,t) &= (-1)^{\alpha+1}\sigma^2 \frac{\partial^2}{\partial x^2} p(x,t) \quad , \quad x \in R \quad , \quad t \in R, \\ p(x.0) &= p_o(x) \quad , \quad \alpha = 1/n \quad , \quad \alpha \in N^* \end{aligned} \tag{9.105}$$

which yields the probability density of a fractional Brownian motion defined in terms of fractional derivative. We have the following

LEMMA 9.1. *The solution of the fractional Fokker-Planck equation (9.105) is similar to the solution of the heat equation*

$$\frac{\partial p}{\partial t}(x,t) = (-1)^{n+1} \frac{\partial^{2n} p}{\partial x^{2n}}(x,t) \quad , \quad (x,t) \in \Re \times \Re . \blacksquare \tag{9.106}$$

Proof Define the Fourier transform

$$\hat{p}(\xi,t) := \frac{1}{2\pi} \int_R e^{ix\xi} p(x,t) dx; \tag{9.107}$$

then taking the Fourier transform of (9.106) yields the equation

$$\frac{\partial^\alpha}{\partial t^\alpha} \hat{p}(\xi,t) = (-1)^{\alpha+2} \sigma^2 \xi^2 \hat{p}(x,t),$$

of which the solution is

$$\hat{p}(\xi,t) = \hat{p}_o(\xi) \exp\left\{ (-1)^{\frac{\alpha+2}{\alpha}} (\sigma\xi)^{\frac{2}{\alpha}} \right\}. \tag{9.108}$$

The inversion of (9.108) provides

$$p(x,t) = \frac{1}{2\pi} \int_R q(x-y,t,\alpha) p_o(y) dy, \tag{9.109}$$

with

$$q(x,t,\alpha) = \left[\exp\left\{(-1)^{\frac{\alpha+2}{\alpha}}(\sigma\xi)^{\frac{2}{\alpha}}t\right\}\right]^{\wedge}$$

$$= \frac{1}{2\pi}\int_R e^{ix\xi}\exp\left\{-(\sigma\xi)^{2n}t\right\}. \tag{9.110}$$

And this expression (9.110) is exactly the solution of the even order heat equation (9.106)!

Formal proof. This equivalence can be formally derived as follows, at least to get a hint.

Define the partial derivative operators D_t and D_x with respect t and x respectively, to rewrite the heat equation (9.106) in the form

$$\left(D_t - (-1)^{n+1}D_x^{2n}\right)p(x,t) = 0. \tag{9.111}$$

We then have the operator equality

$$D_t = (-1)^{n+1}D_x^{2n};$$

and on taking the $(1/n)$-th power of both sides, we obtain the equality

$$D_t^{1/n} = (-1)^{1+\frac{1}{n}}D_x^2, \tag{9.112}$$

which defines the heat equation

$$\frac{\partial^\alpha}{\partial t^\alpha}p(x,t) = (-1)^{1+\alpha}\frac{\partial^2}{\partial x^2}p(x,t) \quad , \quad \alpha = \frac{1}{n}. \tag{9.113}$$

In the next section we shall try to use the above identification between heat equation and the fractional Fokker-Planck equation with zero mathematical expectation to infer some results in the general case when the mean value of the process in not zero. Our problem is the following.

9.11. A new approach to fractional Fokker-Planck equation

Problem. How can we extend the formulation above, clearly the identification between the heat equation and the Fokker-Planck equation, in order to deal with processes with non zero mean value? ■

At the present stage we are guessing a possible solution, and to this end we shall proceed as follows:

Step (i) Assume that the mean value of the process is defined by the differential equation

$$\dot{m}(t) = f(t); \tag{9.114}$$

then we shall write the corresponding heat equation of order $2n$ in the form

$$\frac{\partial p}{\partial t}(x,t) = -f(t)\frac{\partial p}{\partial x}(x,t) + (-1)^{n+1} g^{2n}(t)\frac{\partial^{2n} p}{\partial x^{2n}}(x,t), \tag{9.115}$$

where the term $g^{2n}(t)$ is introduced in order to generalize the result as much as possible.

Indeed, multiplying both sides of (9.115) by x and integrating with respect to x over $\Re$ will yield (9.114).

Step (ii) Introducing the operators D_t and D_x we rewrite equation (9.115) in the form

$$D_t p(x,t) = \left[-D_x f(t) + (-1)^{n+1} D_x^{2n} g^{2n}(t)\right] p(x,t)$$

hence the equality

$$D_t + f(t)D_x = (-1)^{n+1} g^{2n}(t) D_x^{2n}. \tag{9.116}$$

Step (iii) Taking the power $1/n = \alpha$ of both sides yields

$$\left(D_t + D_x f(t)\right)^{\alpha} = (-1)^{1+\alpha} D_x^2 g^2(t). \tag{9.117}$$

Step (iv) Formally, we then have the equality

$$\sum_{j=0}^{\infty}\binom{\alpha}{j}\left(D_x f\right)^j D_t^{\alpha-j} \quad = \quad (-1)^{1+\alpha} D_x^2 g^2 \cdot$$

Step (v) Hence the fractional Fokker-Planck equation

$$\frac{\partial^\alpha p}{\partial t^\alpha}(x,t) \quad = \quad -\sum_{j=1}^{\infty}\binom{\alpha}{j} f^j(t)\frac{\partial^\alpha p}{\partial x^j \partial t^{\alpha-j}}(x,t) + (-1)^{1+\alpha} g^2(t)\frac{\partial^2 p}{\partial p^2}(x,t)\cdot \tag{9.118}$$

Step (vi) When $f(x,t)$ and $g(x,t)$ depend explicitly upon x and t, we would then have the fractional equation]

$$\frac{\partial^\alpha p}{\partial t^\alpha}(x,t) \quad = \quad -\sum_{j=1}^{\infty}\binom{\alpha}{j}\frac{\partial^\alpha (fp)}{\partial x^j \partial t^{\alpha-j}}(x,t) + (-1)^{1+\alpha}\frac{\partial^2 (g^2 p)}{\partial x^2}(x,t)\cdot \tag{9.119}$$

References

9.1. FRIEDEN, B.R.; Restoring with maximum likelihood and maximum entropy, *J. Optical Soc. of America*, Vol 64, No 4, pp 511-518, 1972

9.2. GAPONOV, A.V. and RABINOVICH, M.I. *Nonlinearities in Action*, Springer, Berlin, 1992

9.3. GULL, S.F. and DANIELL, G.J.; Image reconstruction from incomplete and noisy data, *Nature*, Vol 272, pp 686-690, 1976

9.4. JAYNES, E.T.; Monkey, kangaroos and N, in *Maximum Entropy and Bayesian Methods in Applied Statistics,* J.H. Justice (ed.), Cambridge University Press, Cambridge, pp 26-58, 1986

9.5. JUMARIE, G.; A Fokker-Planck equation of fractional order with respect to time, *J. Math. Physics*, Vol 33, No 10, pp 3536-3542, 1992

9.6. JUMARIE, G.; Kolmogorov entropy, image entropy and entropy of non random functions, A comparison, *Int. J. General Systems.;* to appear

9.7. MANDELBROT, B.B. and Van NESS, J.W.; Fractional Brownian motion, fractional noises and applications, *SIAM Review*, Vol 10, No 4, pp 422-437, 1968

9.8. MANDELBROT, B.B.; *The Fractal Geometry of Nature,* W.H. Freeman, New York, 1977, reprint, Plenum Publishing, New York, 1992

9.9. MAYER-KRESS, G. (Ed.)*Dimension and Entropies in Chaotic Systems*, Springer, Berlin, 1987

9.10. OLDHAM, K.B. and SPANIER, J.; *The Fractional Calculus*, Academic Press, New York, 1974

9.11. SCHUSTER, H.G.; *Deterministic Chaos*, Physik Verlag, Weinheim, 1984

9.12. SKILLING, J.; Theory of maximum entropy image reconstruction, in *Maximum Entropy and Bayesian Methods in Applied Statistics,* J.H. Justice (ed.), Cambridge University Press, Cambridge, pp 156-178, 1986

Chapter 10

Outline of Applications

Physical models are as different from the world as a geographical map is from the surface of the earth.
Léon Brillouin

The ultimate purpose of any theory is to provide new approaches to practical problems, and in the present chapter we shall outline some applications of the above results. Analysis of the stochastic stability of dynamical systems, the theory of time series of fractional order, fractals in image processing, solution of the master equation, definition of the master equation of fractional order, optimal control of dynamical systems subject to complex-valued fractional Brownian motion motion, fractals in human systems, the relation between dynamics of information and fractals.

10.1. Background on stochastic stability

10.1.1. Stochastic Liapunov function

The idea which is behind the concept of the so called Liapunov function of deterministic dynamics is well known. Given the system driven by the vector equation

$$\dot{x}(t) = f(x) \quad , \quad f(0) = 0 \quad , \quad x \in \Re^n \, , \tag{10.1}$$

let us consider a function $V(x)$ which is continuous in a certain region D comprising the origin of coordinates. If $V(x)$ is positive definite on D ($V(x) > 0$, $x \in D$, $x \neq 0$), ($V(0) = 0$) and if $dV(x(t)/dt \lhd 0$, then the equilibrium position zero is asymptotically stable. Indeed, in such a case $V(x(t))$ tends necessarily to zero, therefore $V(\lim x(t)) = 0$ and $\lim x(t) = 0$ as t increases.

Kushner [10.10] generalized this approach in order to deal with Markovian systems. By using the so called Dynkin formula [10.4] which provides a technique for associating a super-martingale with a Markovian process, he constructed such a supermartingale expressed in terms of the Liapunov function $V(x)$, and then applied Kolmogorov super-martingale inequality to conclude that $V(x)$ converges with probability one as time increases. This approach is very attractive but unfortunately it works only for

G. Jumarie, *Maximum Entropy, Information Without Probability and Complex Fractals*
© Springer Science+Business Media Dordrecht 2000

very simple (low-dimensional) cases, and it is not easy to use as soon as the number of state parameters takes some importance.

This is the reason why, in the following, we shall still consider the use of the Liapunov function, but in a simpler stochastic framework. Our goal is to have a formulation which would be tractable in a practical implementation on the one hand, and which could be easily extended to dynamical systems in the phase plane on the other hand; and to this end we shall work by using the well known technique of moment equations [10.6,10.9]. But before doing so, for the convenience of the reader we shall call to mind the essentials of Liapunov stability.

10.1.2. Liapunov stability via Chebyshev's inequality

(i) Consider the one-dimensional stochastic dynamics

$$dx = f_1(x,t)dt + f_2(x,t)db(t,2). \tag{10.2}$$

DEFINITION 10.1. Let there be given a stochastic process $x(t)$ which has only one equilibrium position at $x = 0$, and consider the (candidate) Liapunov function $V^2(x)$. Then $x(t)$ is said to have the property of second-order stability in theLiapunov sense at $x = 0$, if the following conditions are satisfied:

$$E\{d(V^2)\} \leq 0 \quad as \quad t \uparrow +\infty, \tag{10.3}$$

$$E\{(dV^2)^2\} \downarrow 0 \quad as \quad t \uparrow +\infty \cdot \blacksquare \tag{10.4}$$

This definition can be supported as follows. Assume that (10.3) and (10.4) are satisfied, then according to Chebyshev's inequality one will have the inequality

$$\Pr\{dV^2(t) \triangleright \varepsilon\} \leq E\{(dV^2)^2 / \varepsilon^2\}, \tag{10.5}$$

for any positive ε, therefore $dV^2(t) \to 0$ with probability one. In other words $V^2(x(t))$ converges to a constant as t increases, and this constant is necessarily defined by $x = 0$.

(i) In practical applications, we shall calculate the derivatives of the function $\phi := V^2$ which are

$$(V^2)_x = 2V(x)V_x(x),$$

$$(V^2)_{xx} = 2(V_x)^2 + 2VV_{xx},$$

and Itô's lemma will yield

$$d(V^2) = \left[2VV_x f_1 + (V_x^2 + VV_{xx}) f_2^2 \sigma^2\right] dt + 2VV_x f_2 db(t,2). \tag{10.6}$$

(ii) If $p(x,t)$ denote the probability density of $x(t)$, then one has the mathematical expectations

$$E\{dV^2\} = dt \int_R \left[2VV_x f_1 + (V_x^2 + VV_{xx}) f_2^2 \sigma^2\right] p(x,t) dx, \tag{10.7}$$

and

$$E\{(dV^2)^2\} = 4\sigma^2(t) dt \int_R (VV_x f_2)^2 p(x,t) dx, \tag{10.8}$$

and then, in order to obtain stability conditions, in terms of structural paremeters we have only to verify that the conditions (10.3) and (10.4) are satisfied.

10.1.3 Practical implementation

(i) Using (10.7), condition (10.3) reads

$$\int_R \left[2VV_x f_1 + (V_x^2 + VV_{xx}) f_2^2 \sigma^2\right] p(x,t) dx \leq 0, \tag{10.9}$$

and in practical applications it will be simpler to check that the coresponding integrand is negative.

(ii) There remains now to write that the derivative of (10.8) is negative to have

$$\frac{d}{dt}(\sigma^2(t)) I + \sigma^2 \dot{I}(t) \lhd 0 \tag{10.10}$$

with

$$\begin{aligned} I(t) &:= \int_R (VV_x f_2)^2 p(x,t) dx \\ &=: \int_R \phi^2(x,t) p(x,t) dx. \end{aligned} \tag{10.11}$$

To this end, we refer to the Fokker-Planck equation

$$\partial_t p(x,t) = -\partial_x (f_1 p) + (1/2) \partial_{xx} (f_2^2 \sigma^2 p), \tag{10.12}$$

which provides the equality

$$\dot{I}(t) = \int_R (\partial_t \phi^2 + f_1 \partial_x \phi^2 + (1/2) f_2^2 \sigma^2 \partial_{xx} \phi^2) p dx, \tag{10.13}$$

and on substituting into (10.9) we obtain the sufficient stability condition

$$\int_R \left[\phi d_t(\sigma^2) + \sigma^2 \left(\partial_t \phi^2 + f_1 \partial_x \phi^2 + (1/2) f_2^2 \sigma^2 \partial_{xx} \phi^2\right)\right] p(x,t) dx \leq 0. \tag{10.14}$$

If we drop the integral we obtain the following sufficient stability conditions, in the form

$$2VV_x f_1 + (V_x^2 + VV_{xx}) f_2^2 \sigma^2 \leq 0, \tag{10.15}$$

$$\phi d_t(\sigma^2) + \sigma^2 (\partial_t \phi^2 + f_1 \partial_x \phi^2 + \frac{1}{2} f_2^2 \sigma^2 \partial_{xx} \phi^2) \leq 0. \tag{10.16}$$

Example 1. As an illustrative example consider the stochastic differential equation

$$dx = a x dt + x db(t,2), \tag{10.17}$$

with time independent σ. We choose $V(x) := x^2$. Condition (10.15) yields

$$4ax^4 + 6\sigma^2 x^4 \leq 0$$

or

$$2a + 3\sigma^2 \leq 0. \tag{10.18}$$

And condition (10.16) provides the inequality

$$8ax^4 + 12\sigma^2 x^4 \leq 0$$

which amounts to the condition (10.18).

The same example has been considered by Kushner [10.10] who finds the stability condition

$$2a + \sigma^2 \leq 0, \tag{10.19}$$

which is less restrictive than (10.18) (clearly (10.18) implies (10.19)). This discrepancy which is quite understandable can be explained as follows:

(i) We have used the approximate condition ((10.15),(10.16)) instead of (10.9) and (10.14) which are averaged by the probability density $p(x,t)$, and one may expect that the latter should provide more accurate conditions that the former.

(ii) Next, Kushner's approach refers to Kolmogorov's inequality for martingales, and it is well known that the latter is finer (less restrictive) than Chebyshev's.

In the next section we shall examine how we can extend this approach to stochastic fractional dynamics, but before doing so we shall try to put in evidence the kind of problems which then occur in this case.

10.2 Stability of fractional stochastic dynamics

10.2.1. Introductory remarks and comments

On the stability of fractional dynamics via Liapunov function

A small example will be helpful for our purpose. Consider the complex-valued stochastic dynamics (here z is the complex state $x + iy$)

$$dz = f_1(z,t)dt + f_3(z,t)db(t,3) , \tag{10.20}$$

which yields (dz^2 is to understand as $(dz)^2$)

$$dz^2 = f_3^2(z,t)db^2(t,3) \tag{10.21}$$

and

$$dz^3 = f_3^3(z,t)\sigma_3^3(t)dt . \tag{10.22}$$

Let us select a function $V(z)$ such that

$$V(z) = 0 \Rightarrow z = 0 . \tag{10.23}$$

A simple calculation yields

$$dV = (V_z f_1 + \frac{1}{6}V_{zzz}f_3^3\sigma_3^3)dt + V_z f_3 db(t,3) + \frac{1}{2}V_{zz}f_3^2 db^2(t,3), \tag{10.24}$$

therefore (V^* is complex conjugate of V)

$$\begin{aligned} d(VV^*) &= V^* dV + V dV^* \\ &= 2\Re\left[V^* (V_z f_1 + \frac{1}{6} V_{zzz} f_3^3 \sigma_3^3) \right] dt \\ &\quad + 2\Re\left[V^* V_z f_3 \, db(t,3) + \frac{1}{2} V^* V_{zz} f_3^2 \, db^2(t,3) \right], \end{aligned} \tag{10.25}$$

and

$$\left[d(VV^*)^2 \right] = 2\Re\left[(V^* dV)^2 \right] + 2|V dV^*|^2 . \tag{10.26}$$

With these results the parallels of equations (10.3) and (10.4) provide the stability conditions

$$E\{d(VV^*\} \leq 0 \quad as \quad t \uparrow +\infty \tag{10.27}$$

$$E\left\{ \left[d(VV^*)^2 \right] \right\} \downarrow 0 \quad as \quad t \uparrow +\infty , \tag{10.28}$$

that is to say (see equations (10.25) and (10.26)),

$$\int_{R^2} \Re\left[V^* (V_z f_1 + \frac{1}{6} \sigma_3^3 V_{zzz} f_3^3 \right] p(z,t) dx dy \leq 0, \tag{10.29}$$

$$\int_{R^2} \left(\Re\left[(V^* dV)^2 \right] + |V dV^*|^2 \right) p(z,t) dx dy \downarrow 0 \quad as \quad t \uparrow +\infty . \tag{10.30}$$

Further remarks and comments

The conditions (10.29) and (10.30) are expressed in terms of $p(z,t)$ (which is defined by the generalized Fokker-Planck equation) and it would be of interest to use them in order to derive other conditions which do not involve $p(z,t)$ and thus would be easier to apply. Unfortunately, here, we cannot work with equations similar to (10.15) and (10.16), for the very reason that $p(z,t)$ may be positive or negative.

In order to circumvent this drawback we shall analyze the stability by using the dynamical equations of the state moments.

10.2.2. Stochastic stability via moment equations

Statement of the method (i) Refer to the stochastic fractional differential equation

$$dz = f_1(z,t)dt + \sum_{j=2}^{n} f_j(z,t)w(t,j)(dt)^{1/j}, \tag{10.31}$$

or

$$dz = f_1(z,t)dt + \sum_{j=2}^{n} f_j(z,t)db(t,j), \tag{10.32}$$

and assume that the functions $f_j(z,t)$, $j = 1,2,..,n,$ are some polynomials w.r.t. z. Then by using the generalized Fokker-Planck equation , one can obtain, in a very simple way, a set of linear differential equations which determine the moments $m_j(t)$, $j = 1,2,..,n$.

(ii) This being so, in the framework of system identification, strictly speaking, the dynamical system defined by the equation (10.31) or (10.32) is completely described by the its first n moments $m_1(t)$, $m_2(t),...,m_n(t)$.

In other words, the dynamical system may be defined either by the set $\{f_1, \sigma_2 f_2,..., \sigma_n f_n\}$ or by the set $\{m_1(t), m_2(t),..., m_n(t)\}$; which amounts to saying that the set $\{.., \sigma_j f_j,..\}$ is completely determined by the set $\{..,m_j,..\}$.

(iii) Assume that the system starts from the initial position u_0, one then has that

$$m_j(0) = u_0^j \quad , \quad j = 1,2,...,n.$$

(iv) We shall say that the x-system is stable if the dynamical system with the state vector is $(m_1(t),m_2(t),...,m_n(t))$ is itself stable.

Example 2. In order to show how to proceed, we consider the stochastic fractional differential equation (10.20) which provides the Foker-Planck equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial z}(f_1 p) - \frac{1}{3!}\sigma_3^3(t)\frac{\partial^3}{\partial u^3}(f_3^3(z,t)p). \tag{10.33}$$

(i) Multiplying both sides of (10.33) by z^j, $j = 1,2,...$ and integrating by part, we obtain the dynamical equations

$$\dot{m}_1(t) = \int_{R^2} f_1(z,t)p(z,t)dxdy \quad , \quad m_1(0) = z_0, \tag{10.34}$$

$$\dot{m}_2(t) = 2\int_{R^2} z f_1(z,t)p(z,t)dxdy \quad , \quad m_2(0) = z_0^2, \tag{10.35}$$

$$\dot{m}_j(t) = \int_{R^2} j\left[z^{j-1}f_1 + \frac{1}{3!}(j-1)(j-2)z^{j-3}f_3^3\sigma_3^3\right]p(z,t)dxdy \, , \, j \geq 3. \tag{10.36}$$

(ii) Assume now that we are dealing with the special case

$$dz = a\,z\,dt + z\,b(t,3) \quad , \quad z(0) = z_0, \tag{10.37}$$

with σ_3 independent of time. We then have the following dynamical equations

$$\dot{m}_1 = a m_1 \quad , \quad m_1(0) = u_0, \tag{10.38}$$
$$\dot{m}_2 = 2a m_2 \quad , \quad m_2(0) = u_0^2, \tag{10.39}$$
$$\dot{m}_3 = (3a + \sigma_3^3) m_3 \quad , \quad m_3(0) = z_0^3. \tag{10.40}$$

Then zero is asymptotically stable if $m_j(t)$ tends to zero as t increases, for $j = 1,2,3$ that is to say if

$$3a \leq -\sigma_3^3 \quad .$$

Further remarks and comments. Assume that we apply the same technique to the Itô stochastic differential equation

$$\dot{x} = a\,x + x\,w(t,2) \quad , \quad x(0) = x_0.$$

It is sufficient to consider the dynamical equations of the moments $m_1(t)$ and $m_2(t)$, which are

$$\dot{m}_1 = a m_1 \quad , \quad m_1(0) = z_0,$$
$$\dot{m}_2 = (2a + \sigma_2^2) m_2 \quad , \quad m_2(0) = z_0^2,$$

and zero is asymptotically stable if one has the condition $2a \lhd -\sigma_2^2$. It is exactly the condition which has been obtained by Kushner [10.10] in his approach via stochastic Liapunov function.

In the following, we shall show how complex-valued fractional Gaussian white noises provides an approach to time series of fractional order.

10.3. Fractional time series in the complex plane

In accordance with the literature on time series, we introduce the backward shift operator B defined by the relation $Bz_t := z_{t-1}$, (the symbol := means that the left hand side is defined by the right hand side) and the backward difference operator ∇,

$$\begin{aligned} \nabla z_t &:= z_t - z_{t-1} \\ &= (1-B)z_t. \end{aligned}$$

Analogously with the definitions of the various models of time series (see for instance Box et al [10.1], Brockwell et al [10.2]) we shall introduce the following classes of fractional time series.

Stationary complex-valued fractional linear filter model

Define the transfer function

$$\psi(B) := 1 + \psi_1 B + \psi_2 B^2 + \ldots,$$

where $\psi_1, \psi_2, .. \in C$ is a sequence which may be finite or infinite. Then the complex-valued fractional filter C-$(FF)_n$ model of order n is written as

$$z_t(n) - \mu = \psi(B) w_t(n), \tag{10.41}$$

where $\mu \in C$ is the constant mean of the process and $w_t(n)$ is a $C\text{-}(GWN)_n$ of order n with constant moments.

Stationary complex-valued fractional auto-regressive model

Define the auto-regressive operator

$$\phi(B) := 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p,$$

$\phi_i \in C$, $i = 1,..,p$; then the complex-valued fractional auto-regressive $C\text{-}(FAR)_n$ model of order n is written as

$$\phi(B)\left[z_t(n) - \mu\right] = w_t(n) \tag{10.42}$$

where μ and $w_t(n)$ are defined as in equation (10.41).

Stationary complex-valued fractional auto-regressive moving average model

With the notations above, the complex-valued fractional auto-regressive moving average *C-(FARMA)$_n$* model of order n is defined by the equation

$$\phi(B)\left[z_t(n)-\mu\right] = \theta(B)w_t(n). \tag{10.43}$$

Non stationary complex-valued fractional model

The nonstationary complex-valued fractional auto-regressive integrated moving average *C-(FARIMA)$_n$* is defined by the equation

$$\phi(B)(1-B)^d z_t(n) = \theta(B)w_t(n), \tag{10.44}$$

where $d = 0,1,2,\ldots$.

As is well known, the basic tool for analyzing time series is the so called auto-covariance function which provides a useful method for identifying the various coefficients involved in the definition of the process. And in this way we are led to examine what happens with this concept in the case of C-valued time series.

10.4. Autocovariance and C-fractional time series

When z_t is a complex-valued process it is customary to define its autocovariance function $\sigma_{zz}(.,.)$ by the relation

$$\sigma_{zz}(t,t+\tau) = E\left\{z_t z_{t+\tau}^*\right\} - E\left\{z_t\right\}E\left\{z_{t+\tau}^*\right\}, \tag{10.45}$$

and the process is said to be stationary if $E\left\{\left|z_t\right|^2\right\} < \infty$, $E\{z_t\}$ is independent of t and $\sigma_{zz}(t,t+\tau)$ is independent of t, in which case it is customary to write $\sigma_{zz}(\tau)=\sigma_{zz}(t,t+\tau)$.

Here, in our approach to complex-valued time series, a first approach would be to set the following:

DEFINITION 10.2. The auto-covariance function $\gamma_{zz}(t,t+\tau)$ of the complex-valued stochastic process z_t is

$$\begin{aligned}\gamma_{zz}(t,t+\tau) &= E\left\{\left[z_t - Ez_t\right]\left[z_{t+\tau} - Ez_{t+\tau}\right]\right\}, \\ &= E\left\{z_t z_{t+\tau}\right\} - E\left\{z_t\right\}E\left\{z_{t+\tau}\right\}. \blacksquare\end{aligned} \tag{10.46}$$

But unfortunately, defined in this way, the C-valued autocorrelation function is useless for analyzing fractional time series, since it is then zero. Indeed, one has the following:

LEMMA 10.1. *The fractional stochastic processes z_t defined by the models (10.41), (10.42), (10.43) and (10.44) have zero covariance: $\gamma_{zz}(t,t+\tau) = 0$.* ■

Proof. (i) Refer to z_t in the $(FF)_n$ model (10.41); then one has

$$\begin{aligned} \gamma(t,t+\tau) &= E\{(\psi(B)w_t(n))(\psi(B)w_{t+\tau}(n))\}, \\ &= \sum \psi_{kj} E\{w_{t-k\tau}(n)w_{t-j\tau}(n)\}, \\ &= 0 \ . \end{aligned} \tag{10.47}$$

The same calculation applies to the $(FMA)_n$ model (10.43).

(ii) Assume now that z_t is defined by the $(FAR)_n$ model (10.42); then one has the equality

$$z_t(n) - \mu = \phi^{-1}(B)w_t(n) \tag{10.48}$$

which can be rewritten in the form

$$z_t - \mu = \sum \phi_j w_{t-j\tau}(n) \quad , \tag{10.49}$$

therefore the result.

(iii) The $(FARMA)_n$ model provides

$$\begin{aligned} z_t - \mu &= \theta(B)\phi^{-1}(B)w_t(n) \\ &= \sum \eta_k w_{t-k\pi}(n), \end{aligned} \tag{10.50}$$

therefore the result.

This completes the proof of the lemma.

As is evident, this property requires a slight modification of the concept of stationarity for fractional complex-valued time series, and to this end we shall use an autocovariance function of order n.

10.5. Stationarity of order n of C-valued fractional time series

10.5.1. Autocovariance of order n

DEFINITION 10.3. (Autocovariance function of order n). Let z_t denote a C-valued fractional stochastic process of order n, which satisfies the condition

$$\left|E\left\{(z_t - Ez_t)^n\right\}\right| \lhd \infty \ ; \tag{10.51}$$

then, its autocovariance of order n is defined by the expression

$$\gamma_{z^n}(t_1,t_2,..,t_n) \ = \ E\left\{(z_{t_1} - Ez_{t_1})(z_{t_2} - Ez_{t_2})...(z_{t_n} - Ez_{t_n})\right\}, \tag{10.52}$$

where, in order to shorten the notation, Ez_t stands for the mathematical expectation. ■

As already pointed out, this concept is of interest mainly in the analysis of C-valued fractional processes of order n, for which the autocovarianace of order j is zero, when $j=2,3,..,n-1$.

DEFINITION 10.4. (Stationarity of order n). The C-valued fractional time series of order n, z_t with index $Z = \{0, \pm 1, \pm 2, \pm 3, ...\}$, is said to be stationary whenever one has the following equalities

$$E\left\{z_t^n\right\} \lhd \infty \qquad \textit{for all } t \in Z \ , \tag{10.53}$$

$$E\left\{z_t^j\right\} \ = \ \mu_j \ , \ j = 1,2,...,n-1 \ , \ t \in Z \ , \tag{10.54}$$

and

$$\gamma_{z^n}(t,t+\tau_1,....,t+\tau_n) \ = \ \gamma_{z^n}(\tau_1,\tau_2,...,\tau_n) \ . \ \blacksquare \tag{10.55}$$

Similarly to real-valued processes, this concept of stationarity can be thought of as weak stationarity, but it will be largely sufficient for our purpose.

10.5.2. More about complex-valued correlation (of order 2)

The following remarks can be of help in obtaining more insight into the practical significance of the complex-valued covariance function.

To this end we shall write $z(t)$ in the form

$$z(t) \ = \ x(t) + iy(t) \qquad ,$$

and we shall assume that $x(t)$ and $y(t)$ have zero mean and that $z(t)$ is stationary of order two. A simple calculation then yields

$$\gamma_{zz}(\tau) \ = \ \gamma_{xx}(\tau) - \gamma_{yy}(\tau) + i\left[\gamma_{xy}(\tau) + \gamma_{yx}(\tau)\right]. \tag{10.56}$$

In other words, a correlation which is equal to zero means that $x(t)$ and $y(t)$ have the same variance on the one hand, and that the following condition

$$E\{x(t)[y(t+\tau)+y(t-\tau]\} = 0$$

holds on the other hand.

(i) A simple calculation shows that

$$\gamma_{zz}(\tau) = \gamma_{zz}(-\tau). \tag{10.57}$$

(ii) Moreover, one has the equality

$$\begin{aligned} \gamma_{zz}(0) &= \gamma_{xx}(0)-\gamma_{yy}(0)+2i\gamma_{xy}(0) \\ &= \sigma_x^2(0)+\sigma_y^2(0)+2i\sigma_{xy}(0) \end{aligned} \tag{10.58}$$

In other words the equality $\gamma_{ZZ}(0) = 0$ means that $x(0)$ and $y(0)$ are independent, and that in addition they have the same variance.

(ii) The complex variance of $z(t)$ is

$$\begin{aligned} \sigma_z^2(t) &= E\{(z-Ez)^2\} \\ &= \sigma_x^2(t)+\sigma_y^2(t)+2i\sigma_{xy}(t) \end{aligned} \tag{10.59}$$

The C-valued correlation function can then be defined by the relation

$$\rho_{zz}(\tau) = \frac{\gamma_{zz}(t,t+\tau)}{\sigma_{zz}(t)\sigma_{zz}(t+\tau)} \quad , \tag{10.60}$$

and here all we can claim is that one has the inequality

$$|\rho_{zz}(\tau)| \leq 1 \quad . \tag{10.61}$$

10.6. Auto-covariance of order *n* of linear fractional time series

For pedagogical reasons, and mainly to shorten the text, we shall illustrate our purpose by means of simple illustrative examples, but the reader will generalize easily by himself.

10.6.1. Auto-covariance of order n of (FMA)$_n$

Let us consider the time series defined by a fractional Gaussian white noise of order three

$$\begin{aligned} z_t &= w_t(3) - \theta w_{t-1}(3), \\ E\{z_t\} &= 0, \end{aligned} \tag{10.62}$$

(which amounts to considering the difference z_t - Ez_t). The task is to find two equations in terms of covariance of order three, which will allow us to determine the coefficient θ and the third moment σ^3 of the noise $w_t(3)$.

(i) A first equation is obtained by computing $E\{z_t^3\} = \gamma_{z^3}(0,0)$ to have

$$\gamma_{z^3}(0,0) = (1-\theta^3)\sigma^3, \tag{10.63}$$

(ii) and a second equation is

$$\gamma_{z^3}(0,1) = E\{z_t z_t z_{t+\tau}\}, \tag{10.64}$$

$$= -2\theta^3\sigma^3 . \tag{10.65}$$

Remark that equation (10.64) is not the only one that we can select, for instance we could calculate $\gamma_{z^3}(1,1)$, to obtain

$$\begin{aligned} \gamma_{z^3}(1,1) &= E\{(w_t - \theta w_{t-1})(w_{t+1} - \theta w_t)^2\}, \\ &= \theta^3\sigma^3 . \end{aligned} \tag{10.66}$$

10.6.2. Auto-covariance of order n of (FAR)$_n$

Here again, we shall consider a very simple example to illustrate the model. This model is defined by the equation

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + w_t(3), \tag{10.67}$$

$$E\{z_t\} = 0.$$

(i) Multiplying both sides by $z_{t-r}z_{t-s}$, and taking the mathematical expectation of the result so obtained yields the difference equation

$$\gamma_{z^3}(r,s) = \phi_1\gamma_{r-1,s-1} + \phi_2\gamma_{r-2,s-2} \quad , \quad r \triangleright 0 \quad , \quad s \triangleright 0. \tag{10.68}$$

(ii) Dividing by $\gamma_{z^3}(0,0)$ we get the correlation function in the form

$$\rho_{z^3}(r,s) = \phi_1\rho_{r-1,s-1} + \phi_2\rho_{r-2,s-2} \quad , \quad r \triangleright 0 \quad , \quad s \triangleright 0. \tag{10.69}$$

Whence the parallel of the so called Yule-Walker equations,

$$\rho_{z^3}(1,1) = \phi_1\rho_{z^3}(0,0) + \phi_2\rho_{z^3}(1,1) \quad , \quad \rho_{z^3}(0,0) = 1, \tag{10.70}$$

$$\rho_{z^3}(2,2) = \phi_1\rho_{z^3}(1,1) + \phi_2\rho_{z^3}(0,0) \tag{10.71}$$

which provides the coefficients

$$\phi_1 = \frac{\rho_{z^3}(1,1)(1-\rho_{z^3}(2,2))}{1-\rho^2_{zz^n}}, \tag{10.72}$$

$$\phi_2 = \frac{\rho_{z^3}(2,2)-\rho^2_{z^3}(1,1)}{1-\rho^2_{z^3}}. \tag{10.73}$$

(iii) The general solution of the equation (10.69) is

$$\rho(r,s) = C_1e^{\lambda_1(r+s)} + C_2e^{\lambda_2(r+s)}, \tag{10.74}$$

where λ_1 and λ_2 are the roots of the characteristic equation

$$1-\phi_1e^{-2\lambda} - \phi_2e^{-4\lambda} = 0. \tag{10.75}$$

These considerations can be generalized in a straightforward manner to more complex systems.

Yule-Walker equation of order n

Let us consider the time series

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + ... + \phi_p z_{t-p} + w_t(3), \tag{10.76}$$

Multiplying both sides by $z_{t-r}z_{t-s}$ and taking the mathematical expectation of the result so obtained, we obtain the equation

$$\begin{aligned} \gamma_{z^3}(r,s) &= \phi_1\gamma_{z^3}(r-1,s-1) + \phi_2\gamma_{z^3}(r-2,s-2) + ... \\ &+ \phi_p\gamma_{z^3}(r-p,s-p), \quad , \quad r \triangleright 0 \quad , \quad s \triangleright 0 \end{aligned} \tag{10.77}$$

Dividing both sides by $\gamma_{z^3}(0,0)$, we have the correlation equation

$$\rho_{z^3}(r,s) = \phi_1\rho_{z^3}(r-1,s-1) + ... + \phi_p\rho_{z^3}(r-p,s-p) \quad . \tag{10.78}$$

On substituting $(r,s) = (1,1), (2,2),$ into (10.78), we obtain the set of linear equations

$$\rho_{z^3}(1,1) = \phi_1 + \phi_2\rho_{z^3}(1,1) + ... + \phi_p\rho_{z^3}(p-1,p-1), \tag{10.79}$$

$$\rho_{z^3}(2,2) = \phi_1\rho_{z^3}(1,1) + \phi_2 + ... + \phi_p\rho_{z^3}(p-2,p-2), \tag{10.80}$$

.....

$$\rho_{z^3}(k,k) = \phi_1\rho_{z^3}(k-1,k-1) + \phi_2\rho_{z^3}(k-2,k-2) + ... + \phi_p, \tag{10.81}$$

which provides a practical approach to the determination of the coefficients $\phi_1, \phi_2, ..., \phi_p$.

10.6.3. Auto-covariance of order n of (FARMA)$_n$

Let us now consider the time series

$$z_t - \phi_1 z_{t-1} = w_t(3) - \theta_1 w_{t-1}(3) \tag{10.82}$$

We need three equations to determine the values of the parameters ϕ_1, θ_1 and σ, and they can be obtained as follows.

(i) Calculating the third power of equation (10.82), that is to say

$$\left(z_t - \phi_1 z_{t-1}\right)^3 = \left[w_t(3) - \theta_1 w_{t-1}(3)\right]^3,$$

and then on taking the mathematical expectation, we have

$$\gamma_{z^3}(0,0) - 3\phi_1\gamma_{z^3}(0,1) + 3\phi_1^2\gamma_{z^3}(1,1) = (1+\theta_1^3)\sigma^3. \tag{10.83}$$

(ii) Multiplying both sides of (10.82) by $z_{t-r}z_{t-s}$ and taking the mathematical expectation yields

$$\gamma_{z^3}(r,s) - \phi_1\gamma_{z^3}(r-1,s-1) = \gamma_{wz^2}(-r,-s) - \theta_1\gamma_{wz^2}(-r+1,-s+1). \tag{10.84}$$

Setting $r = 0$ and $s = 0$, we obtain

$$\gamma_{z^3}(0,0) - \phi_1\gamma_{z^3}(1,1) = -\theta_1\gamma_{wz^2}(1,1). \tag{10.85}$$

In order to obtain the value of $\gamma_{wz2}(1,1)$, we multiply both sides of (10.82) by z_tz_t, we take the mathematical expectation to have

$$\gamma_{z^3}(o,o) - \phi_1\gamma_{z^3}(1,1) = \gamma_{wz^2}(0,0) - \theta_1\gamma_{wz^2}(1,1). \tag{10.86}$$

According to (10.82) one has the equation

$$z_t = \phi_1 z_{t-1} + w_t(3) - \theta_1 w_{t-1}(3),$$

and on squaring, we obtain

$$\begin{aligned} z_t^2 = {} & \phi_1^2 z_{t-1}^2 + w_t^2(3) + \theta_1^2 w_{t-2}^2(3) + 2\phi_1 z_{t-1} w_t(3) \\ & -2\phi_1\theta_1 z_{t-1} w_{t-1}(3) - 2\theta_1 w_t(3) w_{t-1}(3). \end{aligned}$$

Multiplying by w_t and taking the mathematical expectation, it comes

$$\gamma_{wz^2}(0,0) = \sigma^3. \tag{10.87}$$

10.7. On the invertibility of complex-valued fractional time series

On the surface, the theory of invertibility as known for (usual) time series should be easily extended to C-fractional time series.

For instance the *(FARMA(p,q)n)* process defined by the equation

$$\phi(B)z_t = \theta(B)w_t(n) \tag{10.88}$$

is said to be invertible if there exists a sequence of C-constants $\{\psi_j\}$ such that

$$\sum_{j=0}^{\infty}[\psi_j] \lhd \infty, \tag{10.89}$$

and

$$z_t = \sum_{j=0}^{\infty}\psi_{jj}w_{t-j}(n). \tag{10.90}$$

Formally, equation (10.88) provides the equality

$$z_t = \frac{\theta(B)}{\phi(B)}w_t(n), \tag{10.91}$$

and the search of the invertibilty condition for this process amounts to find the conditions for expanding $\theta(B)/\phi(B)$ in series of powers of B

10.8. Complex white noise in image processing

In the present section we shall show that the Gaussian white noise so defined is quite relevant in image processing.

10.8.1. Fractals in dynamic vision

Our main assumption is the following one:

Conjecture. As a consequence of the basic definition of an image in terms of pixels, a meaningful alternative to the Gaussian white noise $w(t)$ in image processing, is the complex-valued fractional Gaussian white noise of order 4, $w(t,4)$, defined by the expression

$$db(t,4) =: w(t,4)(dt)^{1/4}, \tag{10.92}$$

or what amounts to the same,

$$b(t,4) = \int_0^t w(\tau,4)(d\tau)^{1/4}. \tag{10.93}$$

Support for this conjecture.

Assume that the image is defined in the (x_1,x_2)-plane and let $z := x_1+ix_2$. The displacement of objects can be described by the equation

$$z(t+1) = R(t)(z(t)-T(t)), \tag{10.94}$$

where $T(t)$ and $R(t)$ define a translation and a rotation respectively. In a continuous time model, we shall write

$$\dot{z}(t) = f_1(t)z(t)+f_2(t), \tag{10.95}$$

again with the observation

$$z'(t) = z(t)+w_z(t). \tag{10.96}$$

Here, in our computer vision problem, $z(t)$ is the co-ordinate of a pixel of the image, and one may safely assume that z' is a neighbouring pixel of $z(t)$: this is the modelling which probably involves the smallest numbers of assumptions.

In other words, $w(t)$ can be thought of as a C-valued random variable which takes on the values $\Delta\rho\ exp\{2ik\pi/4\}$, $k = 0,1,2,3$ with the uniform probability 1/4, what amounts to replacing equation (10.94) by

$$z'(t) = z(t)+\tilde{w}(t,4), \tag{10.97}$$

where $\tilde{w}(t,4)$ is the random walk of order *4*.

The continuous version of this observation equation would be

$$dz'(t) = z(t)dt+db(t,4). \tag{10.98}$$

10.8.2. Further remarks and comments on the significance of the model

Consider the set of pixels pictured on the lattice in the Fig.1, and assume that the point under

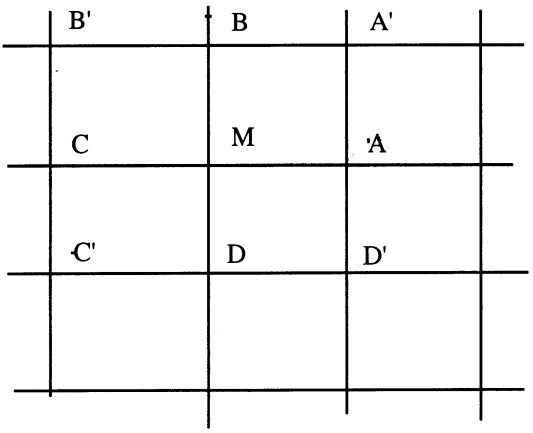


Fig.1 Lattice of pixels in images

consideration is the one referred to as M.

When M is observed with some measurement errors, then in a quite natural way it will be identified (or confused) with one of its neighbouring points. At first glance this observed point might be an element either of the set $\{A,B,C,D\}$ or of the set $\{A',B',C',D'\}$. A model of this error term Δz_j at the discrete time j can be written in the form

$$\Delta z_j = \beta_j R_j(4) w_j + \sqrt{2}(1-\beta_j) R_j(4) w_j e^{i\pi/4} , \qquad (10.99)$$

where β_j is a Bernouilli random variable (β_j = 1 or 0 with the respective probabilities p and q, $p+q = 1$), $R_j(4)$ is the Rademacher random variable of order 4, and w_j is a Gaussian random variable with zero mean.

The presence of the Bernouilli variable contributes some difficulties of theoretical nature, but simplification can occur in some special cases of interest.

(i) If we restrict ourselves to the nearest points $\{A,B,C,D\}$, one then has the equality

$$\Delta z_j = R_j(4) w_j \qquad (10.100)$$

which yields the Gaussian white noise of order 4, $w(t,4)$

(ii) Assume now that one has the equality $p = q = 1/2$. In this case one can replace (10.99) by the equation

$$\Delta z_j = \lambda(1+\sqrt{2}\, e^{i\pi/4})R_j(4)w_j, \tag{10.101}$$

where λ is a real-valued parameter, whereby we obtain a *C-(fBm)*$_4$ with the characteristic function

$$\varphi_b(v,t) = \exp\left\{\lambda(1+\sqrt{2}\, e^{i\pi/4})\frac{(iv)^n}{n!}\sigma^n t\right\}. \tag{10.102}$$

(iii) A third *approximate* model consists in working as if all the points A,A',B,B',C,C' were on the same circle, in which case one will write

$$\Delta z_j = R_j(8)w_j, \tag{10.103}$$

to obtain a Gaussian white noise of order 8.

(iv) A simplified version of (10.99) is

$$\Delta z_j = \lambda_1 R_j(4)w_j + \lambda_2\sqrt{2}\, e^{i\pi/4}R_j(4)w_j, \tag{10.104}$$

where λ_1 and λ_2 denote two real valued weighting coefficients.

10.9. Brightness dynamics and fractional white noise

In the following we shall put in evidence one of the effects of complex fractional white noise on the dynamics of brightness functions.

10.9.1. Definition of the problem

In continuous time modelling we assume that the motion of the considered image is defined by the linear diferential equation (10.95) which we rewrite for convenience as

$$\dot{z}(t) = f_1(t)z(t) + f_2(t). \tag{10.105}$$

Unfortunately, $z(t)$ cannot be directly observed and is subject to some measurement errors. As a model, we shall assume that the observed variation is not dz as defined by the equation

$$dz = (f_1 z + f_2)dt, \tag{10.106}$$

but rather is $d_e z$ such that

$$d_e z = (f_1 z + f_2)dt + db(t,n), \tag{10.107}$$

where $b(t,n)$ is a C-$(fBm)_n$ with

$$E\{(db(t,n))^n\} = \sigma^n dt \cdot \tag{10.108}$$

Given this observation framework, what is the value of the mathematical expectation of the *observed* brightness function variation ?

10.9.2. Mean value of the observed brightness function

In the complex plane the brightness function $g(x,y)$ will be defined by the equality $g(x+iy) = g(z)$, and we shall introduce the observed variation $d_e g(z)$,

$$d_e g(z) = g(z + d_e z) - g(z). \tag{10.109}$$

One has the following

LEMMA 10.2. *Assume that $g(z)$ is differentiable up to the order n. Then the conditional expectation of $d_e g(z)$, given z, is provided by the expression*

$$E\{d_e g(z)|z\} = g'(z)(f_1 z + f_2)dt + g^{(n)}(z)\sigma^n dt \ . \blacksquare \tag{10.110}$$

Proof. As usual, Taylor's expansion yields

$$g(z + d_e z) = g(z) + \sum_{j=1}^{n-1} \frac{1}{j!} g^{(j)}(d_e z)^j + \frac{1}{n!} g^{(n)}(z)(d_e z)^n + o(dz)^{n+1} \quad , \tag{10.111}$$

with

$$E\{d_e z|z\} = (f_1 z + f_2)dt, \tag{10.112}$$

$$E\{(d_e z)^j|z\} = 0 \quad , \quad j = 2,...,n-1, \tag{10.113}$$

$$E\{(d_e z)^n|z\} = \sigma^n dt \quad , \tag{10.114}$$

from which we obtain

$$E\{g(z + d_e z)|z\} = g(z) + g'(z)(f_1 z + f_2)dt + g^{(n)}(z)\sigma^n dt + o(dt^2), \tag{10.115}$$

whence the result.

This expression involves the additional term $g^{(n)}(z)\sigma^n dt$ of which the magnitude depends upon the n-th derivative of $g(z)$. As mentioned, in image processing n will be equal to four.

10.10. Solution of the master equation

10.10.1. Preliminary background

We consider a stochastic system of which the real-valued state $x(t)$ takes on values in a discrete space which we define to be the space of integers. The simplest model is the so called *birth - and - death* process. In this process, transition may only occur from state x to states x+1 or x-1; transition probabilities are assumed to be differentiable with respect to time and their derivatives (λ_x the birth rate, and μ_x death rate) depend only on x (in the homogeneous case). If $p_{xy}(t)$ denotes the probability of going from x to y in time t, we easily obtain the so called *backward* and *forward equations*,

$$\dot{p}_{xy}(t) = \lambda_x p_{x+1,y}(t) + \mu_x p_{x-1,y}(t) - (\lambda_x + \mu_x)p_{xy}(t), \tag{10.116}$$

$$\dot{p}_{xy} = p_{x,y-1}(t)\lambda_{y-1} + p_{x,y+1}(t)\mu_{y+1} - (\lambda_x + \mu_x)p_{xy}(t). \tag{10.117}$$

From equation (10.117), saturating with respect to initial distribution, and making the substitution $y \leftarrow x$ in the result so obtained, we then arrive at the *master equation*

$$\dot{p}(x,t) = \lambda_{x-1}p(x-1,t) + \mu_{x+1}p(x+1,t) - (\lambda_x + \mu_x)p(x,t), \tag{10.118}$$

where $p(x,t)$ is the probability that the system be in state x at time t.

Solving the master equation is a hard task, except for some very simple cases. In the following we shall show how $C\text{-}(fBm)_n$ provides a simulation technique to find this solution.

10.10.2. Simulation technique for the master equation

Basic assumption. We shall assume that x is large enough, in such a waythat applying Taylor's expansion to $\lambda_{x-1}p(x\text{-}1,t)$ and $\lambda_{x+1}p(x+1,t)$ yields convergent series.

Transformation of the master equation.

Introducing the forward and backward finite difference

$$\Delta g(x) := g(x+1)-g(x), \tag{10.119}$$
$$\nabla g(x) := g(x)-g(x-1), \tag{10.120}$$

one can write,

$$\lambda_{x-1}p(x-1,t) = \lambda_x p(x,t)-\nabla(\lambda_x p(x,t)), \tag{10.121}$$
$$\mu_{x+1}p(x+1,t) = \mu_x p(x,t)+\Delta(\mu_x p(x,t)); \tag{10.122}$$

and substituting into the master equation (10.118), we have

$$\dot{p}(x,t) = \Delta(\mu_x p(x,t))-\nabla(\lambda_x p(x,t)). \tag{10.123}$$

This being so, D(.) being shorthand for d(.)/dt, one has the equalities

$$\Delta = D+(1/2!)D^2+\ldots+(1/n!)D^n+\ldots\ ,$$
$$\nabla = D-(1/2!)D^2+\ldots+(-1)^{n+1}(1/n!)D^n+\ldots$$

which, when applied to master equation (10.118) , yields

$$\frac{\partial p}{\partial t}(x,t) = -\sum_{j=0}^{\infty}\frac{1}{(2j+1)!}\frac{\partial^{2j+1}}{\partial x^{2j+1}}\left[(\lambda_x-\mu_x)p\right] + \sum_{j=0}^{\infty}\frac{1}{(2j)!}\frac{\partial^{2j}}{\partial x^{2j}}\left[(\lambda_x+\mu_x)p\right]. \tag{10.124}$$

Stochastic differential equation associated with the master equation

Our purpose now is to derive a stochastic differential equation of which the Fokker-Planck equation would be equivalent to the master equation, and to this end it might be helpful to bear in mind the following result.

LEMMA 10.3. *Consider the Itô stochastic differential equation*

$$dz(t) = f_1(z,t)dt + \sum_{j=2}^{\infty} f_j(z,t)db(t,j) \quad , \quad z \in C \, , \tag{10.125}$$

where b(t,j), j = 2,3,... are normalized independent C-(fBm)$_H$'s, that is to say with unit moments. Then the corresponding state probability density p(z,t) is solution of the PDE

$$\frac{\partial p}{\partial t}(z,t) = -\frac{\partial}{\partial z}(f_1 p) + \sum_{j=2}^{\infty} \frac{(-1)^j}{j!} \frac{\partial^j}{\partial z^j}(f_j p) \, . \; \blacksquare \tag{10.126}$$

On applying this result to the equation (10.124), we can then consider the latter as defining the probability density of the stochastic system described by the stochastic differential equation

$$\begin{aligned} dz &= (\lambda_z - \mu_z)\left(dt + \sum_{j=1}^{\infty} db(t,2j+1) \right) \\ &\quad + (\lambda_z + \mu_z) \sum_{j=1}^{\infty} db(t.2j) \quad . \end{aligned} \tag{10.127}$$

Simulation technique

This result suggests, in a straightforward way, a Monte Carlo approach to solving the equation (10.124), that is to say, the master equation.

10.11. Master equation of fractional order

10.11.1. Preliminary remarks

A fractional master equation can be found in the literature [10.5] in the form

$$\frac{\partial^\alpha p}{\partial t^\alpha}(x,t) = \sum_{x'} \gamma(x-x')p(x',t) \quad , \quad 0 \lhd \alpha \lhd 1 \, , \tag{10.128}$$

where $\partial^\alpha p / \partial t^\alpha$ is the fractional derivative in the Liouville sense. Loosely speaking, we start from the master equation, and we replace the partial derivative with respect to time by a partial derivative of fractional order. For instance, the master equation (10.118) would yield the fractional master equation

$$\frac{\partial^\alpha p}{\partial t^\alpha} = \lambda_{x-1} p(x-1,t) + \mu_{x+1} p(x+1,t) - (\lambda_x + \mu_x) p(x,t). \tag{10.129}$$

This equation is exactly the parallel of the fractional Fokker-Planck equation

$$\frac{\partial^\alpha p}{\partial t^\alpha}(x,t) = -\frac{\partial}{\partial p}[f(x,t)p(x,t)] + \frac{1}{2}\frac{\partial}{\partial x}[g(x,t)p(x,t)], \tag{10.130}$$

and to the best of our understanding, both of them can be thought of as related to stochastic processes with long-range dependence.

The question which then arises in quite a natural way is the following: Is the master equation (10.128) still meaningful in terms of $C\text{-}(fBm)_n$? If it is not, how can we obtain the suitable corresponding equation? ■

In order to obtain some hints about the answer we shall refer to the analogy which appears between the heat equation and the Fokker-Planck equation, and we shall assume that this analogy holds for any order n. The reader might be convinced by the following two examples.

(i) Consider the master equation of the Poisson process,

$$\frac{\partial p}{\partial t}(x,t) = -\lambda[p(x,t) - p(x-1,t)]. \tag{10.131}$$

Clearly, it can be thought of as the discrete version or approximation of the Fokker-Planck equation

$$\frac{\partial p}{\partial t}(x,t) = -\lambda \frac{\partial p}{\partial x}(x,t). \tag{10.132}$$

(ii) Assume that $\lambda = \mu =$ constant in the master equation (10.118), then we can rewrite it in the form

$$\frac{\partial p}{\partial t} = \lambda[p(x+1,t)-2p(x,t)+p(x-1,t)], \tag{10.133}$$

which suggests the Fokker-Planck equation

$$\frac{\partial p}{\partial t}(x,t) = \lambda\frac{\partial^2 p}{\partial x^2}(x,t). \tag{10.134}$$

10.11.2. How to guess master equation of fractional order?

We shall apply the analogy above to the heat equation

$$\frac{\partial p}{\partial t}(x,t) = (-1)^{n+1}\frac{\partial^{2n} p}{\partial x^{2n}}(x,t), \tag{10.135}$$

and to the fractional Fokker-Planck equation

$$\frac{\partial^\alpha p}{\partial t^\alpha}(x,t) = (-1)^{1+\alpha}\frac{\partial^2 p}{\partial x^2}(x,t). \tag{10.136}$$

As we showed previously, in the framework of complex-valued fractional Brownian motion with independent increment (in contrast to long range dependent processes) there is a complete equivalence between these two equations and either can be used to describe stochastic processes of fractional order. We shall assume that this is also true for the master equation.

(i) Given this prerequisite, let us refer to the backward shift operator B defined in Section 10.3, $B^k.p(x,t) = p(x-k,t)$. The approximation of the heat equation (10.135) yields the master equation

$$\frac{\partial p}{\partial t}(x,t) = (-1)^{n+1}(B-1)^{2n}p(x,t), \tag{10.137}$$

or, in a more detailed form,

$$\frac{\partial p}{\partial t}(x,t) = (-1)^{n+1}\sum_{k=0}^{2n}(-1)^k\binom{2n}{k}p(x-k,t)\cdot \tag{10.138}$$

This equation should be thought of as a basic equation by means of which one could infer more general models. For instance, one could write,

$$\frac{\partial p}{\partial t}(x,t) = \sum_{k=0}^{n} \gamma(k) p(x-k,t) \cdot \tag{10.139}$$

(ii) The second model of fractional master equation is provided by the equation (10.136) which direct yields

$$\frac{\partial^{\alpha} p}{\partial t^{\alpha}}(x,t) = (-1)^{1+\alpha}\left[p(x+1,t) - 2p(x,t) + p(x-1,t)\right], \; 0 \lhd \alpha \lhd 1. \tag{10.140}$$

A moodel which could be generalized in the form (10.128), for instance.

10.12. Stochastic optimal control of order *n*

10.12.1. Statement of the problem

Warning about the notation. In this section, in order to comply with the literature on optimal control, locally $z(t) := x(t) + iy(t)$ will denote the complex time-dependent state of the system, and $u(t)$ will be the complex-valued control. Here again, the notations are local notations.

Statement of the Problem. With this notation the dynamical system under consideration is defined by Itô's stochastic differeential of order n,

$$dz(t) = f_1(z,u,t)dt + f_n(z,u,t)db(t,n) \quad , \quad z(0) = z_0 \; , \tag{10.141}$$

where $f_1(.)$ and $f_n(.)$ are two functions which are suitably defined in order that $z(t)$ exist (loosely speaking, these conditions are similar to those related to Itô's stochastic differential equations of order two), and where $b(t,n)$ is a $C\text{-}(fBm)_n$ with the characteristic function $exp\{(iv)^n \sigma^n(t)/n!\}$.

The purpose is to determine the (optimal) control $u^*(t)$ which minimizes the cost function (the star here does not refer to complex conjugate !)

$$G := E\left\{h(z(T),T)) + \int_0^T g[z(t),u(t),t]dt\right\}, \tag{10.142}$$

where T has a given fixed value.

Obviously it is assumed that this optimization makes sense.

It is well known that, since there is no single optimal trajectory there is no stochastic equivalent to the Euler-Lagrange equations, and that consequently it is necessary to work via the principle of optimality, which will be done in the following.

10.12.2. Stochastic principle of optimality of order n

We consider the stochastic value function

$$V^*(t) = E\left\{h(z^*(T),T)) - \int_T^t g\left[z^*(\tau),u^*(\tau),\tau\right]d\tau\right\}, \tag{10.143}$$

and we calculate its total time derivative in two different ways.

(i) Firstly, one has

$$\frac{dV^*(t)}{dt} = -g(z^*(t),u^*(t),t)) \ . \tag{10.144}$$

The mathematical expectation $E\{.\}$is dropped since $z(t)$ and $u(t)$ can be known without error at time t.

(ii) Secondly, we use the Taylor lemma of order n to obtain dV^*. We express dV^* in series expansion, and retaining the n-th degree term the incremental change in V^* can be written as follows:

$$\begin{aligned} dV^* &= E\left\{\frac{\partial V^*}{\partial t}dt + \sum_{j=1}^{n}\frac{1}{j!}\frac{\partial^j V^*}{\partial z^j}(dz)^j\right\} \\ &= E\left\{V_t^* dt + \sum_{j=1}^{n}\frac{1}{j!}(V^*)^{(j)}\left[f_1 dt + f_n db(t,n)\right]^j\right\}. \end{aligned} \tag{10.145}$$

Functions of $z(t)$ equal their own expectations. $E\{db^j(t,n)\} = 0,\ j = 1,\ldots,n\text{-}1$. Dividing by dt and replacing the n-th term by its time derivative, we obtain

$$dV^* = V_t^* dt + V_z^t f_1 dt + \frac{1}{n!}(V^*)^{(n)} f_n{}^n \sigma_n{}^n dt,$$

therefore

$$\frac{dV^*}{dt} = V_t^* + V_z^* f_1 + \frac{1}{n!}(V^*)^{(n)} f_n{}^n \sigma_n{}^n . \tag{10.146}$$

(iii) The stochastic principle of optimality is obtained by combining equs (10.144) and (10.146), to yield

$$V_t^*(t) = -\min_u \left\{ \begin{array}{l} g\left(z^*(t),u(t),t\right) + V_z^* f_1\left(z^*(t),u(t),t\right) \\ +\dfrac{1}{n!}\dfrac{\partial^n V^*}{\partial z^n} f_n{}^n\left(z^*(t),t\right)\sigma_n{}^n(t) \end{array} \right\}. \tag{10.147}$$

10.12.3. Application to linear systems with quadratic cost

It is well known that in the special case of dynamics subject to Gaussian white noise simplifications occur for linear systems with quadratic cost [10.13]. Let us examine what happens here with the stochastic dynamics of order *n*.

Problem. We consider the dynamics

$$dz = \left[f_{11}(t)z + f_{12}(t)u\right]dt + f_n(t)db(t,n), \tag{10.148}$$

and we select the cost function in the form

$$G := E\left\{\frac{1}{n!}s(T)z^n(T) + \frac{1}{2}\int_0^T \left[a(t)z^2 + 2b(t)zu + c(z)u^2\right]dt\right\}. \tag{10.149}$$

As usual, the quadratic term can be thought of either as a distance or as an energy, while the term $z^n(T)$ is introduced for convenience in order to have a suitable derivative of the value function approach.

Solution . Analogously to the solution of the linear quadratic problem with Gaussian white noise we shall assume that the value function is in the form

$$V(t) = \frac{1}{n!}s(T)z^n(T) + \frac{1}{n!}\int_t^T s(\tau)f_n{}^n(\tau)\sigma_n{}^n(\tau)d\tau. \tag{10.150}$$

The term in $z^n(T)$ is the *certainty-equivalent value function* and the integral is the *stochastic-value function increment.*

The optimality stochastic partial differential equation (10.147) becomes here

$$V_t^* = -\min_u \left\{ \begin{array}{l} \dfrac{1}{2}\left[az^2 + 2bzu + cu^2\right] + \dfrac{s(t)}{(n-1)!}z^{n-1}(f_{11}z + f_{12}u) \\ +\dfrac{1}{n!}s(t)f_n{}^n(t)\sigma_n{}^n(t) \end{array} \right\}. \tag{10.151}$$

Differentiating with respect to $u(t)$ and equating to zero, we obtain the solution

$$u^*(t) = -\frac{1}{c}\left[bz + \frac{1}{(n-1)!} f_{12} s(t) z^{n-1}\right]. \tag{10.152}$$

10.12.4. A Variational approach via moment equations

From a practical standpoint it is well known that the difficulty in the dynamic programming approach is that the computation of the solution is cumbersome and time consuming. In addition, apart from the special case of linear dynamics with quadratic cost, generally one cannot obtain a closed form or at least an approximate closed form, for the solution, and this is a serious drawback for the engineer who need to analyze in as a simple manner as possible the effects of the variation of the model parameters.

Fortunately, there is a family of problems which can be solved by means of Lagrange's variational approach. They are defined by dynamics and cost densities which both are polynomials with respect to the state z and the control u. In this case, one can then convert the stochastic problem into a deterministic problem by using the state moments as new state vector, and the control moments as a new control vector.

This case occurs in the theory of neighbouring optimal control and can be broken down into the following step.

(i) Determine the nominal (deterministic) solution of the problem without noise, which will be considered as the reference trajectory of the stochastic neighbouring optimal approach.

(ii) By using a Taylor expansion around the nominal solution, specify the cost function. This expansion may be of any order, depending upon the accuracy sought.

(iii) By using a Taylor expansion around the nominal trajectory specify the dynamic equations (or constraint). Here again, this expansion may be of any order, and the only requirement is that the order selected in (ii) and the order in (iii) be mutually consistent.

(iv) Take the mathematical expectation of these expressions. The mathematical expectation of the cost will involve state moments up to the order k; and as a result we shall have to consider the corresponding k-th state moment equation .

(v) By using the Euler-Lagrange variational technique, determine the optimal control of the deterministic problem so defined by the expectation of the cost function and the moment equations.

(vi) In many realistic instances, this control will be obtained in the form

$$\bar{u} = \sum_j u_j(t) E\{z^j\}. \tag{10.153}$$

Then, when it is the case, the on-line control u(t) will be obtained by making the substituton $E\{z^j\} \leftarrow z^j$, to yield

$$u(t) = \sum_j u_j(t) z^j \ . \tag{10.154}$$

In the reference [10.7] the reader will find the detailed application of this approach to the optimal control of stochastic dynamics subject to Gaussian white noise, and the reference [10.8] displays all the mathematical formulae we need to extend the method to the analysis of stochastic systems defined in terms of *C-(fBm)*$_n$.

10.13. On the presence of fractals in natural science

10.13.1. Fractals in phycics

As a first general comment, we shall point out that the conditions which are necessary to obtain a *C-(fBm)*$_n$ are so simple and so general, that *one should expect to find fractals in a large number of natural science areas.*

In Chapter 8, related to thermodynamics of order n, we have shown that the basic assumptions of quantum mechanics are quite similar, or at least are quite consistent with those which generate *C-(fBm)*$_n$, in such a manner that, at first glance, we should not be surprized at all to find fractals in quantum mechanics.

There are four main fundamental competing theories in quantum mechanics. Currently, most theoretical physicists are working on super-strings and super-membranes. A little bit older is Penrose's theory of twistors. Newer than these two theories is Connes' theory [10.3] of non-commutative geometry. And recently Nottale [10.12] suggested the theory of fractal space-time.

In an approach to Nottale's idea, we should consider the complex variable $z = x + it$, but then we need *another* time τ, for instance a slow time or a fast time, to scale the dynamics of z. As an example, assume that t and τ are related by the equation $t=\gamma\tau$, where γ is a real valued constant, then we would have to consider the complex variable

$$z = x + i\gamma\tau$$

with

$$dz^2 = dx^2 - \gamma^2 d\tau^2$$

and

$$|dz^2| = dx^2 + \gamma^2 d\tau^2.$$

10.13.2. Fractals in human systems

In the Chapter 8 we alluded to the use of *C-(fBm)*$_n$ in the thermodynamic analysis of human systems, and our claim is that this approach should be of significant prospect.

(i) The law of the *C-(fBm)*$_n$ is stable in Levy's sense.

(ii) This law can be obtained by a central limit theorem under very broad conditions.

(iii) And the meaning of the *C-(fBm)*$_n$ in terms of collisions between n particles is quite relevant.

We think that this new model should find some applications in finance (see for instance [10,11]). There are cycles in finance and the *C-(fBm)*$_n$ is described in terms of cycles!!

But as is evident, the new problem which then appears is the following. Can any dynamics be randomized in terms of *C-(fBm)*$_n$? We shall comment on this point in the following.

(i) Randomization of one-dimensional system

Formally, modelling a system by means of Gaussian white noises (i.e. of order 2) is very easy and does not give rise to any difficulty, again from a formal standpoint. And one might even suspect that many authors refer to this model mainly and only because it is very simple from the theoretical standpoint. Loosely speaking, we are dealing with the deterministic-theoretic system described by the equation

$$dx = f_1(x,t) \quad , \quad x \in \Re, \tag{10.155}$$

but it appears that in practice it is subject to external random disturbances, and as a result we select the corresponding model in the form

$$dx = f_1(x,t)dt + f_2(x,t)db(t,2). \tag{10.156}$$

The matter is slightly different with the complex-valued fractional white noise. Given the real-valued one-dimensional system (10.155), how can we meaningfully

construct its disturbed companion dynamics in the complex plane by using *C-(fBm)$_n$?* Given a two-dimensional (x,y)--system, is it always possible to write its disturbed companion dynamics by using a stochastic differential equation subject to *C-(fBm)$_n$?*

(ii) Two-dimensional systems and C-(fBm)$_n$

For 2D-systems, the remark is rather obvious. The system

$$\dot{x}(t) = \varphi(x,y,t) \quad , \tag{10.157a}$$
$$\dot{y}(t) = \psi(x,y,t) \tag{10.157b}$$

can be written in the form

$$\dot{z}(t) = f(z,t) \quad , \quad z \in C \tag{10.158}$$

when and only when one has the equalities

$$\varphi(x,y,t) = \Re[f(z,t)], \tag{10.159a}$$
$$\psi(x,y,t) = \Im[f(z,t)]. \tag{10.159b}$$

For instance, the dynamical system

$$\dot{x} = x^2 - y^2,$$
$$\dot{y} = 2xy$$

yields

$$\dot{z} = z^2.$$

Clearly, only some special 2D-systems will be candidates for a model via *C-(fBm)$_n$*. Or again, fractals via C-(fBm)$_n$ can occur only in 2D-systems which exhibit some special structural conditions. And this is neither surprizing nor disappointing! This is quite rightly so! For instance chaotic dynamics appear only in some special cases, and for 3D-systems only.

(iii) One-dimensional systems and C-(fBm)$_n$

Assume now that we are dealing with the one-dimensional system

$$\dot{x}(t) = f(x,t) \quad . \tag{10.160}$$

Unfortunately, here we cannot extend the problem by considering the model (10.158) for the very reason that, in doing so, we shall introduce the dummy state variable $y(t)$ which is absolutely meaningless on the physical standpoint.

As a matter of fact, we are lucky since the $C\text{-}(fBm)_n$ $b(t,n)$ can also be written in the form

$$w(t,n) = R(n)|w(t)|, \tag{10.161}$$

where R(n) is the Rademacher random variable. Using this result we shall write the disturbed version of (10.160) in the form

$$dx = f(x,t) + \Re\big(R(n)|w(t)|\big) \tag{10.162}$$

10.14. Information and fractal dynamics

Let us consider a one-dimensional system (x,t) with the state x, and define its tangential dynamics $(\Delta x, \Delta t)$. According to information of non-random functions, Δx and Δt involve the respective amounts of uncertainties $\ln|\Delta x|$ and $\ln|\Delta t|$ and we are going to examine the relation between these two data.

In terms of information transfer, the tangential dynamics $(\Delta x, \Delta t)$ will be thought of as transforming the informational input $\ln|\Delta t|$ into the informational output $\ln|\Delta x|$.

In the following, we shall restrict our analysis to those systems which evolve with information loss, that is to say such that $\ln|\Delta x| \lhd \ln|\Delta t|$, and we shall derive the general form of their dynamical equations.

(i) The first model is provided by the equality

$$\ln|\Delta x| = \ln|\Delta t| - \ln\rho, \tag{10.163}$$

with $\rho \in \Re$, $\rho \rhd 1$; therefore the dynamical equation (since Δt is positive)

$$\rho\dot{x}(t) = \varepsilon(t) \quad , \quad \varepsilon(t) = \pm 1, \tag{10.164}$$

which can be generalized in the form

$$\dot{x}(t) = 1/f(x) \quad , \quad |f(x)| \rhd 1. \tag{10.165}$$

(ii) The second model is defined by the equality

$$\ln|\Delta x| = \rho \ln|\Delta t| \quad , \quad 0 \lhd \rho \lhd 1, \tag{10.166}$$

therefore the equation

$$|\Delta x| = (\Delta t)^{\rho}, \tag{10.167}$$

which provides the fractal dynamics

$$|\Delta x| = (\Delta t)^{f(x)} \quad , \quad 0 \lhd |f(x)| \lhd 1. \tag{10.168}$$

The model (i) describes a system in which the information loss is constant, whilst in the model (ii) the information loss is proportional to the input.

References

10.1. BOX, G.E.P. and JENKINS, G.M. *Time Series Analysis. Forecasting and Control,* Holden-Day, San Francisco, 1970

10.2. BROCKVWELL, P.J. and DAVIS, R.A.; *Time Series. Theoriy and Method,* Springer, 1987, Berlin

10.3. CONNES, A.; *Noncommutative geometry*, Academic press, New York, 1994

10.4. DYNKIN, E.B.; Controlled random sequences, *Theory of Probability and its Applications,* Vol 10, pp 1-14, 1965

10.5. ELWAKIL, S.A. and ZAHRAN, M.A.; Fractional integral representation of master equations, *Chaos, Solitons and Fractals,* Vol 10, No 8, pp 1309-1313, 1999

10.6. JUMARIE, G.; A practical variational approach to stochastic optimal control via state moment equations, *Journal of Franklin Institute: Engineering and Applied Mathematics,* Vol 332 B, No 6, pp 761-772, 1995

10.7. JUMARIE, G.; Improvement of stochastic neighbouring optimal control, using non-linear Gaussian white noise terms in the Taylor expansions, *J. of Franklin Institute: Engineering and Applied Mathematics*, Vol 333 B, No 5, pp 773-789, 1966

10.8. JUMARIE, G.; Stochastic dynamics and fractional white noise in the complex plane. Stability, entropy and fractal dimension, *Systems Analysis, Modelling, Simulation,* Vol 134, pp 307-334, 1999

10.9. KATS. I.I. and KRASOVSKII, N.N.; On the stability of systems with random disturbances, *J. Applied Mathematics and Mechanics (PMM)*, Vol 24, pp 65-71, 1960

10.10. KUSHNER, H.J.; *Stochastic Stability and Control*, Academic Press, New-York, 1967

10.11. MANDELBROT, B.B.; *Fractals and Scaling in Finance. Discontinuity, Concentration, Risk,* Springer Verlag, New York, 1997

10.12. NOTTALE, L.; Fractal-Space Time and Microphysics, World Scientific, Singapour, 1993

10.13 STEINGEL, R.; *Stochastic Optimal Control,* Wiley, New York, 1986

Index