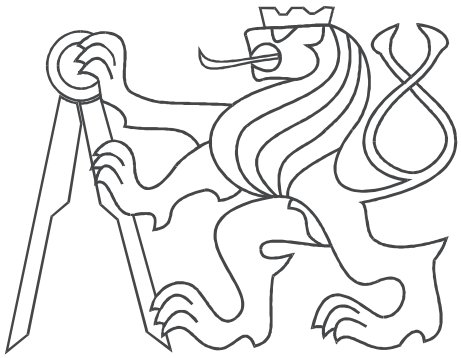
CZECH TECHNICAL UNIVERSITY IN PRAGUE

FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING



DOCTORAL THESIS

Applications of Generalized Statistics and Multifractals in Financial Markets and Thermodynamics

2015 JAN KORBEL

Bibliografický záznam

Autor:	Ing. Jan Korbel České vysoké učení technické v Praze Fakulta jaderná a fyzikálně inženýrská Katedra fyziky
Název práce:	Aplikace zobecněné statistiky a multifraktálů na finančních trzích a v termodynamice
Studijní program:	Matematické inženýrství
Studijní obor:	Matematická fyzika
Vedoucí práce:	Ing. Petr Jizba, Ph.D. České vysoké učení technické v Praze Fakulta jaderná a fyzikálně inženýrská Katedra fyziky
Akademický rok:	2015/2016
Počet stran:	111
Klíčová slova:	Ekonofyzika; Zobecněná statistika; Multifraktály; Neexenzivní termodynamika; Rényiho entropie

Bibliogaphic entry

Author:	Ing. Jan Korbel Czech Technical University in Prague Faculty of Nuclear Sciences and Physical Engineering Department of Physics
Title of Dissertation:	Applications of Generalized Statistics and Multifractals in Financial Markets and Thermodynamics
Degree Programme:	Mathematical Engineering
Field of Study:	Mathematical Physics
Supervisor:	Ing. Petr Jizba, Ph.D. Czech Technical University in Prague Faculty of Nuclear Sciences and Physical Engineering Department of Physics
Academic year:	2015/2016
Number of pages:	111
Keywords:	Econophysics; Multifractals; Generalized statistics; Nonextensive thermodynamics; Rényi entropy

Abstrakt

Tato práce se zabývá vybranými tématy z teorie komplexních systémů a ekonofyziky. Zaměřuje se zejména na multifraktální analýzu, anomální difuzi a teorii zobecněných entropií. Tyto modely jsou založeny na několika univerzálních konceptech - škálování, zobecněná statistika a extenzivita. Všechna tato témata jsou široce studována z teoretického hlediska. Podrobně jsou diskutovány nejdůležitější otázky každého z témat, jako například odhad škálovacích parametrů v případě multifraktální analýzy, modely s těžkými rameny a frakční modely v případě anomální difuze nebo speciální třídy zobecněných entropií. V návaznosti na to to jsou také navrženy a prezentovány aplikace výše uvedených modelů na finančních trzích a v termodynamice.

Abstract

This thesis deals with selected topics from the theory of complex systems and econophysics. It is mainly focused on multifractal analysis, anomalous diffusion and theory of generalized entropies. These models are based on several universal concepts - scaling, generalized statistics and extensivity. All of these topics are broadly studied from the theoretical point of view. Salient issues of each topic, such as the estimation of characteristic scaling exponents in the case of multifractal analysis, heavy-tailed and fractional models in the matter of anomalous diffusion, or special classes of generalized entropies are discussed in detail. Subsequently, applications of the aforementioned models in financial markets and thermodynamics are presented.

To Jana,
my beloved wife,
because she is the most important person in my life.

Acknowledgements

There are many people I would like to thank. First, I would like to thank my scientific collaborators, namely Dr. Mohammad Shefaat (FU Berlin), Dr. Xaver Sailer (Nomura Ldt.), Prof. Dalibor Štys and his research group (Institute of Complex Systems, University of South Bohemia), particularly Dr. Renata Rychtáriková and Dr. Štěpán Papáček. I would also like to thank my colleague Václav Zatloukal for many hours of interesting discussions. I also greatly appreciate the opportunity of spending several months at Freie Universität Berlin, where I had the chance to work under the supervision of Prof. Dr. Hagen Kleinert, Dr.h.c. mult. Above all, I would like to express my immense thanks to my supervisor, Dr. Petr Jizba. It has been more than six years since we first met and I am thankful for all advices, opportunities and motivation he gave me during my entire studies.

Of course, this work would not have been possible without the encouragement and the support of my family and my friends.

Contents

List of Symbols

symbol	name	definition
$\langle\cdot\rangle$	expectation value	$\langle f\rangle = \sum_i p_i f_i$
$\langle\cdot\rangle_q$	escort mean	$\langle f\rangle_q = \sum_i \rho_i(q) f_i$
$\lceil x\rceil$	ceiling function	lowest integer exceeding x
$\lfloor x\rfloor$	floor function	largest integer not exceeding x
card	number of elements	
${}_{x_0}\mathcal{D}^{\nu}_{x}$	Riemann-Liouville fractional derivative	Sect. 3.2.1
${}^{*}_{x_0}\mathcal{D}^{\nu}_{x}$	Caputo fractional derivative	Sect. 3.2.2
$\mathfrak{D}^{\nu}_{x}$	Riesz-Feller fractional derivative	Sect. 3.2.3
$\mathcal{D}_q(\mathcal{P})$	hybrid entropy	Sect. 4.3.3
$E_{\mu,\nu}$	Mittag-Leffler function	Appendix C
$f(\alpha)$	multifractal spectrum	(2.8)
$\mathcal{F}[f](p) = \overline{f}[p]$	Fourier transform	$\mathcal{F}[f](p) = \int_{\mathbb{R}} e^{ipx} f(x)\mathrm{d}x$
${}_{x_0}\mathfrak{I}^{\nu}_{x}$	fractional integral	(3.6)
$\mathcal{H}_{\alpha,\beta;\bar{x},\bar{\sigma}}(p)$	Stable Hamiltonian	(3.41), Appendix A
$i.i.d.$	independent, identically distributed	
$\mathcal{H}(\mathcal{P})$	Shannon entropy	(4.2)
$\mathcal{I}_q(\mathcal{P})$	Rényi entropy	(2.15), Sect. 4.3.1
$L_{\alpha,\beta}(x)$, $L_{\alpha,\beta}(t)$	Lévy distribution, Lévy process	Sect. 3.3.2
$\mathcal{L}[f](s) = \widehat{f}[s]$	two-sided Laplace transform	$\mathcal{L}[f](s) = \int_{\mathbb{R}} e^{-sx} f(x)\mathrm{d}x$
$\mathcal{M}[f](z)$	Mellin transform	Appendix B
ρ_i^q	escort distribution	$\rho_i(q) = p_i^q / \sum_j p_j^q$
$\mathcal{S}_q(\mathcal{P})$	Tsallis entropy	Sect. 4.3.2
$\tau(q)$	scaling function	(2.11)
$W(t)$	Wiener process	Sect. 3.1
$W_H(t)$	Fractional Brownian motion	Sect. 3.3.1
$Z(q), Z(q,s)$	partition function	(2.10)

List of Figures and Tables

List of Figures

List of Tables

Chapter 1

Introduction

A lot of systems observed in the nature - dynamical, biological, chemical, quantum, sociological or financial, just to name a few - exhibit a wide range of complex phenomena including non-linearity, phase transitions, regime switching, sudden changes and/or memory effects. Usually it is extremely hard to describe dynamics of such systems within the conventional framework represented by classical mechanics, equilibrium thermodynamics and the theory of diffusion. Nevertheless, these theories often serve as springboards for various generalizations and adaptations. The models which are based on some kind of universal, generally applicable principles belong to the most successful. In the thesis we particularly focus on models based on *self-similarity*, *scaling* and the concept of *generalized additivity*. It is universality which makes the models successful in many interdisciplinary areas including both theoretical works as well as practical applications. The amount of possible applications represents a strong motivation for rapid development of these areas and encourages looking for new interdisciplinary fields, in which the aforementioned ideas can improve effectiveness and predicability of the models.

Let us mention a few examples of areas in which the ideas known from theory of complex systems have helped to establish new disciplines. Application of methods commonly used in physics on financial markets, known as *econophysics* [1, 2], was established as a response to increased demand of realistic forecasting in finance. Indeed, financial markets are a very complex and complicated system and it is essential to use appropriate sophisticated models for successful trading. As an example of this kind of connection we can mention *multifractal analysis*. Multifractals were originally observed in dynamical systems but afterwards celebrated great success in financial markets. *Generalized statistics* [3] with generalized versions of *limit theorems* and *stable distributions* serve as another example. Additionally, *nonextensive thermodynamics* [4] have celebrated great success with the idea of replacing the Shannon entropy by generalized versions of entropy. As evolution of these research areas was sometimes rather precipitous and has brought many interesting moments, we briefly summarize some as-

pects of historical evolution of research fields related to the previously mentioned topics and point out some of the rudimental and pioneering works.

1.1 Historical Overview

This section provides an overview of historical evolution and thr current state of the art of complex systems and related topics such as multifractal analysis or the theory of generalized statistics. It is interesting that many model are based on very similar ideas, although they might seem different on the first sight. That is why these ideas have made their way through the theory of complex systems. We gradually go through some of the topics and present the most important works which have largely contributed to the particular topics.

Scaling and *self-similarity* belong to the most important properties of complex systems. They have been known for a very long time, since they are often associated with fractals. Fractal systems can be observed everywhere in real natural systems. In the era of differential calculus, i.e., in the times of Newton and Leibnitz, researchers believed that most of the processes observed in nature can be described in terms of derivatives and integrals. However, later, it turned out that many processes cannot be described in terms smooth trajectories. This was later confirmed by the theory of stochastic processes. These extremely rough processes are usually not differentiable, but they can be described by a specific scaling rule, or, in more realistic cases, by a set of scaling rules. If the system can be described by a single dominant scaling rule, we refer to it as a *unifractal*. On the other hand, if the system is described by a whole continuous set of local scaling rules with different intensities, we talk about *multifractals*. The first works related to the theory of scaling exponents were done by L. Hölder and particularly by H. .E Hurst, a British hydrologist who was the first one to study long-term dependence in hydrology [5]. Further important contributions were done e.g., by H. Hentschel and I. Proccacia [6, 7] and by the pioneer of multifractal analysis, a French and American mathematician B. B. Mandelbrot [8, 9]. Since that time, multifractals found wide application in chemistry [10] or in finance [11]. Nowadays, it is still a hot topic with an active community and many interesting open problems.

The theory of *generalized statistics* is also connected to the topic of scaling exponents. When a process is described by many independent, identically distributed (i.i.d.) increments, then the infinite sum of these increments is described by the normal distribution. This is the result of *Central limit theorem* under the assumption of finite variance. When we omit the assumption of finite variance, we obtain the whole class of *Lévy distribution* which are stable under the operation of convolution. The theory of *stable distributions* was broadly studied by B. V. Gnedenko and A. N. Kolmogorov [12]. Interestingly, these distributions are closely related to fractional calculus through *fractional diffusion equations*. Fractional calculus operates with generalizations of ordinary

derivatives and integrals for non-natural, real orders. These generalizations have been studied since the nineteenth century, but the first attempt on a systematic description is dated to the second half of the twentieth century. All these mathematical descriptions lead to processes which can describe systems with sudden jumps (also called "black swans" [13]) more accurately. These black swans are observed, for instance, in quantum systems [14] or financial markets [15].

It is interesting that similar ideas incorporating scaling and generalized statistics can also be found in thermodynamics. In statistical physics, which is a link between equilibrium thermodynamics and the theory of information, have been investigated systems, in which the ordinary extensivity of variables is disrupted because of openness of the system and/or information/energy flows. Such systems have to be described in the regime of non-equilibrium thermodynamics [16]. For some particular cases, it is nevertheless possible to recover some of the thermodynamical properties by using generalized statistics. The two most important examples of generalized information measures are the Rényi entropy discovered by a Hungarian mathematician A. Rényi [17] and Tsallis entropy (also called Tsallis-Havrda-Charvát entropy). Tsallis entropy was firstly discovered in connection with the theory of information divergences by Czech mathematicians J. Havrda and F. Charvát [18] and applied to physics by C. Tsallis [19]. These two entropies opened a new playground for description of systems with long-range correlations, open systems and multifractal systems, called *nonextensive thermodynamics*.

Generally, concepts based on general ideas which find their applications in several scientific fields open discussion about similarities of two or more different fields and bring new ideas adopted in other theories. That is one of their main benefits. Apart from the aforementioned examples, let us mention for example the concept of path integrals [20], which has found its applications in many fields including quantum mechanics, solid state physics or financial markets. One of the aims of this thesis to point out the existence of similar concepts which can be successfully applicable in several fields.

1.2 Aims of the Thesis

The thesis has several targets. As outlined in the previous sections, the thesis presents several general concepts. To the main concepts discussed in the thesis belong scaling, multifractals, generalized statistics, nonextensivity and Legendre structure. It is important to discuss their important theoretical aspects as well as to show the potential of practical applications. The thesis is mainly focused in applications in financial markets, because such applications represent a hot topic in the field of econophysics. Nevertheless, we also mention other possible applications, for instance applications in thermodynamics or in models of developed turbulence. Additionally, the second aim is to cover the topics which have been investigated during author's studies and to provide

a comprehensive overview. There is usually not enough space to present some broader perspective in the articles. All technical details or connections to related topics have to be omitted. Therefore, the thesis provides the optimal format to cover all of these interesting points, so that the reader gets a complete overview about the topic.

The thesis is based on several articles that have been published during author's studies or are currently in the submission process. The thesis connects all of these topics and provides an additional space for more general perspective. Namely, the results from Ref. [21], which discusses some important technical aspects of Diffusion entropy analysis, are presented in Sect. 2.4. Applications to financial series, done in several papers, e.g. in Ref. [22], are presented in Sect. 5.1. Ref. [23] shows the application of Double-fractional diffusion to the theory of option pricing. Theoretical aspects of Double fractional diffusion are discussed in Sect. 3.3.3 and estimation on the real data is presented in Sect. 5.2. Ref. [24] compares several important classes of nonextensive generalized entropies and presents a new class of hybrid entropies and corresponding MaxEnt distributions. The results can be found in Sect. 4.3.3.

The thesis is organized as follows: after this introductory chapter come three theoretical chapters. Namely, chapter 2 covers the multifractal analysis, chapter 3 presents several models of anomalous diffusion and chapter 4 discusses possible generalizations of Shannon entropy. Consequently, chapter 5 is dedicated to applications in finance. The last chapter is devoted to conclusions and perspectives. List of all author's publications published or submitted during the period the doctoral studies can be found at the end of the thesis.

Chapter 2

Multifractal analysis

Scaling and (multi)fractals belong to the most popular concepts in physics, chemistry, biology and many other complex systems. This chapter briefly reviews the existing mathematical framework and compares methods for estimation of multifractal scaling exponents. We particularly discuss some theoretical aspects of Diffusion entropy analysis. At the end of the chapter, we also presents some possible applications of multifractals in physics.

2.1 Fractals and Self-similarity

There exist many real systems with characteristic scaling properties and inner structure which is determined by the scaling rules. This is often connected with fractal properties of the system. Contrary to ordinary physical systems described by (systems of) differential equations with smooth trajectories, fractal systems are systems with rough, non-differentiable structure. When we define fractal dimension, one of the necessary conditions is that the fractal dimension of a smooth function is the same as its topological dimension. As a consequence, a simple rule for recognition of fractal systems can be formulated: if the fractal dimension differs from topological dimension, fractal structure is incorporated in the system.

Popular examples of fractals commonly emerging in the nature include snowflakes, fern leaves, mountain ranges, Romanesco broccoli, coastlines and many others. Moreover, fractals found their applications also in other scientific fields. Let us mention, e.g., astronomy and the rings of Saturn, electromagnetism an the structure of electric discharge or biology with the structure of blood vessel [25]. According to observations,it is necessary to distinguish several kinds of fractals. The most rigorous are *proper fractals*, which obey the scaling rule for all scales. *Natural fractals* are fractals which follows the scaling rule up to some particular scale determined usually by microstructure limitations or by measurement accuracy. The most general type of fractal are *statistical fractals*.

They fulfill the scaling rules only for some statistical quantities. In the real systems are usually observed the latter two types.

There exist several definitions of fractal dimension on different levels of mathematical rigor. We stick to the most illustrative one and sketch the other possibilities. The most familiar is the so-called *box-counting* dimension, which, as the name suggests, is based on counting of boxes in the embedding space. Let us have a set $F \subset \mathbb{R}^D$ and let us divide the space into non-overlapping boxes, l-mesh, of volume l^D. We count the number of boxes which have non-empty intersection with F and denote as $N_F(l)$. When F is a smooth curve, the number acquires the scaling rule $N_F(l) = Cl^{-1}$. We can clearly recognize the dimension as the exponent at $1/l$. Subsequently, we can generally consider $N_F(l)$ in the form

$$N_F(l) = c(l)\, l^{-d_F}\,, \tag{2.1}$$

where $c(l)$ is a slowly varying function of l, i.e.

$$\lim_{l\to\infty} \frac{c(al)}{c(l)} = 1 \quad \text{for all } a > 0\,. \tag{2.2}$$

We can easily extract d_F from previous equation, so

$$d_F = \lim_{l\to 0}\left(-\frac{\ln N_F(l)}{\ln l} + \frac{\ln c(l)}{\ln l}\right) = \lim_{l\to 0}\frac{\ln N_F(l)}{\ln 1/l}\,, \tag{2.3}$$

which is nothing else then the definition of the box-counting fractal dimension. We have to be aware that nothing guarantees the existence of the limit. Nonetheless, in practical applications, we are limited by the measurement precision. Execution of the limit is intractable. In these cases is the limit replaced by linear regression of $\ln N_F(l)$ versus $-\ln l$.

More rigorous approach provides so-called *Hausdorff dimension*, which is based on l-covers. We define l-cover as a countable cover. The elements of the cover are sets containing points which have their respective distance at most equal to l. This determines a class of measures defined as $\sum_i |U_i|^q$, where $\{U_i\}$ is the l-cover (compare with the definition of partition function in Sect. 2.2.) For certain values of $q \in [0, d_H)$ is the measure infinite in $l \to 0$ limit, while for $q \in (d_H, \infty]$ tends the sum to zero. The parameter d_H is therefore the Hausdorff dimension and the sum is nothing else than the generalization of D-dimensional volume for non-natural dimension. Indeed, when both fractal dimensions exist, they are both the same.

Many fractals can be generated trough self-similar transformations. The recursive procedure of fractal creation is a very popular technique and there exist dozens of methods based on simple recursive rules. Among others, Iterated function systems or L-systems [26] provide two examples. All these methods are based on *similarities*. Similarity S is a transformation which just rescales the set but preserves the shape. It holds that $\|S(x) - S(y)\| = c\|x - y\|$. A *self-similar* object is composed of similar copies of

itself, therefore, it can be expressed as $F = \bigcup_i S_i(F)$, where S_i are similarities. Fractal dimension d can be easily determined from equation

$$\sum_i c_i^d = 1\,, \tag{2.4}$$

where c_i are characteristic coefficients of S_i.

We can find self-similar properties not only in systems described geometrically, but also in probabilistic systems. Examples provide stochastic processes, random fields or unifractal cascades. We have to slightly generalize the concept of the fractal dimension in these case. Firstly, because of probabilistic nature of the embedding space, we should work with probabilistic measures. These measures are usually naturally available in the probability space, so it does not usually restrict our investigations. Secondly, in time-evolutionary systems, as e.g., stochastic systems, we have an additional structure given by the time evolution. We have to admit that the time coordinate is conceptually different and this should be reflected when calculating the dimension in space-time coordinate space (x-t space). Additionally, there is no natural measure in time-space coordinate space, i.e., it is not possible (in non-relativistic theories) to mix space coordinates with time and to measure the distances between (x,t)-points. Time is just the parameter of the system. It can be overcome by definition of so-called *affinity*, which, loosely speaking, imposes the implicit scale ratio between time and space coordinates which afterwards enables to define a distance on the space-time coordinate space. Consequently, this allows to define a concept of *self-affinity*, defined as self-similarity in space-time coordinate space with affinity.

At this point, we remind the basic fractal properties of some particular stochastic processes. The most popular stochastic process is the Wiener process, defined e.g., in [27]. The scaling properties can be treated via its conditional distribution

$$p(x,t|x_0,t_0)\mathrm{d}x = \frac{1}{\sqrt{2\pi D(t-t_0)}} \exp\left(-\frac{(x-x_0)^2}{2D(t-t_0)}\right)\mathrm{d}x\,. \tag{2.5}$$

The distribution has is invariant under the transform

$$\Delta x = \alpha \Delta x' \tag{2.6}$$
$$\Delta t = \alpha^2 \Delta t'\,. \tag{2.7}$$

Scaling parameter α cancels out and the distribution remains unchanged. Thus, we become the scaling property $|\Delta x| \propto (\Delta t)^{1/2}$. The exponent is called *Hurst exponent*. It is an important measure for estimation of (multi)-fractal properties and will be further investigated in Sect. 2.3.1. In the following overview are presented fractal dimensions of some familiar stochastic processes:

- sample paths of Wiener process in $\mathbb{R}^n$ ($n \geq 2$) have dimension 2,

- graphs of Wiener process in x-t space have dimension $\frac{3}{2}$,
- graphs of fractional Brownian motion $W_H(t)$ (see Sect. 3.3.1) in x-t space have dimension $2 - H$.
- sample paths of Lévy process $L_\alpha(t)$ (see Sect. 3.3.2) have dimension $\max\{1, \alpha\}$,
- graphs of Lévy process in x-t space $\max\{1, 2 - \frac{1}{\alpha}\}$,

These processes serve often as a springboard for more complex processes. However, many systems cannot be completely described by processes with one scaling exponent. In the real systems are usually present several scaling exponents or even the whole spectrum of scaling exponents. Therefore, we introduce a concept that enables description of processes with more scaling exponents.

2.2 Multifractal Analysis

For many systems are global scaling rules too restrictive. On the other hand, local scaling rules can be often observed. Systems described by more scaling exponents are called *multifractal systems*. These local scaling exponents are usually characteristically distributed for a given system and therefore can be used for classification. In multifractal analysis is assumed that the distribution of scaling exponents has also some typical spectrum of scaling exponents. This spectrum of scaling exponents is called *multifractal spectrum* and fully characterizes the multifractal properties of given system [7]. In this section we show an intuitive definition of multifractal scaling exponents. More rigorous definitions based on multifractal measures can be found e.g., in Ref. [9].

Let us divide the space into distinct regions $K_i(s)$ depending on the typical scale s. We suppose that there is defined a characteristic quantity in each region. Usually, it is the probability distribution p_i. We consider that the probability distribution scales as $p_i \propto s_i^\alpha$. In the limit of small s, we assume that the distribution of scaling exponents can be expressed as a smooth function of α, i.e., in the form

$$P(\alpha, s)\,\mathrm{d}\alpha = c(\alpha) s^{-f(\alpha)}\,\mathrm{d}\alpha\,, \tag{2.8}$$

where $c(\alpha)$ is a slowly varying function of α. Scaling exponent $f(\alpha)$ is called multifractal spectrum and is nothing else than the fractal dimension of subset which scales with exponent α. Hence, in multifractal analysis we assume that there are two probability distributions. Scaling exponents of these distributions determine the behavior of the system. It is also convenient to introduce another approach of multifractal exponents estimation. We introduce the partition function $Z(q, s)$, which is the analogue of its

thermodynamical counterpart (more about multifractal thermodynamics in Sect. 2.5.2). We consider that the partition function scales with the *scaling function* $\tau(q)$:

$$Z(q,s) = \sum_i p_i(s)^q = \langle \mathcal{P}^{q-1} \rangle \propto s^{\tau(q)} \,. \tag{2.9}$$

The relation to multifractal spectrum can be obtained by plugging into the definition of partition function:

$$Z(q,s) = \int \mathrm{d}\alpha P(\alpha,s) p(s,\alpha)^q = \int \mathrm{d}\alpha \, c(\alpha) s^{-f(\alpha)} s^{\alpha q} \propto s^{\tau(q)} \,. \tag{2.10}$$

In the limit of small s is possible to use the steepest descent approximation. Thus, the main term contributing to integral is the one with the smallest exponent. Finally, we get

$$\tau(q) = \inf_\alpha (\alpha q - f(\alpha)) = q\alpha(q) - f(\alpha(q)) \tag{2.11}$$

where $\alpha(q)$ is the exponent which minimizes previous expression. This transform is called Legendre-Frenchel transform or convex conjugation. The properties of the transform are summarized in Ref. [28]. Additionally, when we consider differentiability of scaling exponents, we end with classic Legendre relations, namely

$$\begin{aligned} \tau(q) &= q\alpha(q) - f(\alpha(q)) \,, & (2.12)\\ \frac{\mathrm{d}\tau(q)}{\mathrm{d}q} &= \alpha(q) \,, & (2.13)\\ q &= \frac{\mathrm{d}f(\alpha(q))}{\mathrm{d}q} \,. & (2.14) \end{aligned}$$

In this case, we can immediately write down analogous relations for scaling exponent α, because twice performed Legendre transform gives us back the original function.

The partition function is also closely related to Rényi entropy (which properties are extensively discussed in Sect. 4.3.1), because

$$\mathcal{I}_q(\mathcal{P}(s)) \equiv \mathcal{I}_q(s) = \frac{1}{q-1} \ln Z(q,s) \,. \tag{2.15}$$

The connection to Rényi entropy is important, because it enables us to collate multifractal exponents to so-called *generalized dimension*

$$D(q) = \lim_{s\to 0} \frac{1}{q-1} \frac{\ln \sum_i p_i(s)^q}{\ln s} = \frac{\tau(q)}{q-1} \,. \tag{2.16}$$

The generalized dimension is nothing else than scaling exponent of x^{q-1}-power mean, so

$$\sqrt[q-1]{\langle \mathcal{P}^{q-1}(s) \rangle} \propto s^{D(q)} \,. \tag{2.17}$$

Actually, it can be considered as a generalization of several dimension measures, as topological dimensions ($q = 0$), box-counting dimension ($q \to 1$) or correlation dimension ($q = 2$) [6]. This provides a nice interpretation of the scaling function $\tau(q)$, which is proportional to generalized dimension and therefore measures the distortion from monofractal behavior, which is represented by the curve $\tau_D(q) = D(q-1)$.

The main issue in multifractal analysis is the problem of scaling coefficients estimation. Strictly speaking, the exponents should be extracted from relations in the $s \to 0$ limit, which is in practical applications intractable, because we usually work with measured discrete data. The next section presents some methods of measuring the scaling exponents.

2.3 Estimation of Scaling Exponents

Real applications demand a different approach of scaling exponents estimation. As discussed in previous sections, the estimation based on small-scale limit is unthinkable, because the objects are usually not theoretical (multi)-fractals across all scales. They are rather natural-fractals, with scaling laws perceptible only up to some treshold. We also have to face to the problem of finite amount of data which also changes estimation of relevant quantities. In practical applications, we usually start with some finite set of elements $\{x_i\}_{i=1}^{N}$, which can be a time series, a sequence of measurements, etc. We need to extract the scaling elements only from this limited amount of data. Because we should make the estimation over at least a few scales, small datasets are generally not very suitable for such methods.

We gradually introduce some of the popular techniques for estimation of multifractal exponents and briefly compare their strong and weak aspects. We start with a monofractal technique called *Rescaled range analysis* (RSA). Main reason is that it was historically the first method based on the celebrated *Hurst exponent* and also because of its conceptual clearness. Subsequently, we discuss the multifractal version of HE called *Generalized Hurst exponent* (GHE). As next, we treat probably the most popular technique, called *Detrended fluctuation analysis* (DFA) based on calculation of fluctuations around local trends. Finally, we present the *Diffusion entropy analysis* (DEA) based on estimation of Rényi entropy. Apart from these methods, there have been developed many other methods, among others let us mention Multifractal wavelet analysis [29].

We extensively discuss the related problems. For instance, estimation of probability distributions as histograms or estimation of the optimal bin-width belong to the most important. All presented methods are demonstrated on one-dimensional datasets, however, generalizations to more dimensions are straightforward.

2.3.1 Rescaled Range Analysis

Rescaled range analysis is a simple method based on estimation of Hurst exponent in time series, introduced by H. E. Hurst [5, 30], a British hydrologist and pioneer of theory of scaling exponents. The method is simply based on the investigation of range measured on different scales. From knowledge of properties of stochastic processes (particularly fractional Brownian motion), we deduce that the estimated scaling parameter corresponds to the Hurst exponent. Let us have a series $\{x_i\}_{i=1}^{N}$. For each particular scale s, we divide the series to parts of length s, i.e., we have $X_1(s) = \{x_i\}_{i=1}^{s}$, $X_2(s) = \{x_i\}_{i=s+1}^{2s}$, etc. Similarly to other methods, we have to remove the localization and scale dependence. For this end, we transform the series by subtraction of local means

$$y_i = x_i - \bar{X}_j \tag{2.18}$$

where $\bar{X}_j$ is the corresponding mean, e.g. for $i \in \{1, \ldots, s\}$ we have $\bar{X}_1 = \frac{1}{s}\sum_{i=1}^{s} x_i$, etc. Analogously to previous notation, we have $Y_1(s) = \{y_i\}_{i=1}^{s}$, and so on. For each part $j = 1, \ldots \lfloor N/s \rfloor$, two quantities are calculated, namely *Range* of the series

$$R_j(s) = \max\{Y_j(s)\} - \min\{Y_j(s)\} \tag{2.19}$$

and *Standard deviation*

$$S_j(s) = \sqrt{\frac{Y_j(s) \cdot Y_j(s)}{s}} \tag{2.20}$$

where $a \cdot b$ denotes the scalar product. The ratio R/S is used for estimation of Hurst exponent. We average all local R/S ratios to obtain the global R/S ratio that scales as

$$R/S(s) = \frac{1}{\lfloor N/s \rfloor} \sum_{j=1}^{\lfloor N/s \rfloor} \frac{R_j(s)}{S_j(s)} \propto s^H \,. \tag{2.21}$$

Similarly to all other methods, we assume that the scaling dependence is not far from exact scaling, i.e. $R/S(s) = Ks^H$. Eventually, we can estimate the Hurst exponent from doubly-logarithmic linear regression. Despite its simplicity, which can sometimes cause improper estimations, Rescaled range analysis in the extremely popular method for detection of the characteristic scaling exponent.

2.3.2 Generalized Hurst Exponent

Morales et al. [31] introduced a method which enables to generalize Hurst exponent for multifractal systems. It was successfully applied e.g., in text analysis [32]. The method is slightly improved compared to the R/S analysis and provides the whole spectrum of exponents. The exponent is not based on estimation of Range, for which is necessary to work with large amount of data. Instead, the estimation is based on so-called *structure*

function, which also scales in time with some characteristic scaling exponent. It is defined as

$$K_q(s) = \frac{\langle |x_{i+s} - x_i|^q \rangle}{\langle |x_i|^q \rangle} \,. \tag{2.22}$$

The averaging is done over index i. The averaging method is called *moving time-window* averaging. We shall note, that for $q = 2$, the structure function is proportional to correlation function $C(s) = \langle x_{i+s} x_i \rangle$, which corresponds to the fact that the generalized dimension is for $q = 2$ equal to correlation dimension. We shall note that the denominator $\langle |x_i|^q \rangle$ is not depending on the lag s and therefore does not influence the scaling behavior. However, for large q's can the numerator lead to huge numbers and the denominator tends to normalize the structure function.

The *Generalized Hurst exponent* (GHE) is then defined as

$$K_q(s) \propto s^{qH(q)} \,. \tag{2.23}$$

The parameter $H(q)$ is constant for monofractal series and is equal to (classic) Hurst exponent. In case, when $H(q)$ is not constant, we obtain the Hurst exponent as $H(1) = H$, while the other values are connected with the rest of multifractal scaling exponents.

When investigating time series, it is also possible to use exponential smoothing method, which accentuates the most recent values and suppresses past values. The exponentially weighted average is defined as follows:

$$\langle x \rangle^w = \sum_{j=1}^{N} w_j x_{N-j} \tag{2.24}$$

where $w_j = w_0 \exp\left(-\frac{j}{\theta}\right)$. Parameter θ represents the characteristic time decay. The method represents an elegant and easy way to estimate scaling exponents. Following method shows an alternative way of exponent estimation based on calculation of fluctuations from local trends.

2.3.3 Detrended Fluctuation Analysis

Detrended fluctuation analysis (DFA) is a method based on estimation of local linear/ quadratic/ ... trends and measuring fluctuations from local trends. It was originally introduced in Refs. [33, 34]. Similarly to R/S analysis we begin with subtraction of mean. If we begin with noise-like series, i.e. the series of returns (or successive differences, so $\xi_i = x_{i+1} - x_i$), we have to create a aggregated series, so

$$y_i = \sum_{j=1}^{i} (\xi_j - \langle \xi \rangle) \,. \tag{2.25}$$

When we start with the series $\{x_i\}_{i=1}^N$, we create the series of successive differences and make the mean subtraction after that. We divide the series into parts of length s and estimate the local trends $\hat{y}^\nu$. The local fluctuation function is then defined as

$$F(\nu, s)^2 = \sum (y_{s(\nu-1)+i} - \hat{y}_i^\nu)^2 \,. \tag{2.26}$$

The total fluctuation function can be calculated (similarly to calculation of generalized dimension in Sect. 2.2) as a x^q-power mean, so

$$F(q, s) = \left\{ \frac{1}{N_s} \sum_{\nu=1}^{N_s} [F(\nu, s)^2]^{q/2} \right\}^{1/q} \,. \tag{2.27}$$

Let us assume that the Fluctuation function scales with exponent $h(q)$, i.e. we have $F(q, s) \propto s^{h(q)}$. Therefore, we obtain that

$$\sum_{\nu=1}^{N_s} [F(\nu, s)^2]^{q/2} \propto s^{qh(q)-1} \,. \tag{2.28}$$

When the series $\{x_i\}_{i=1}^N$ is stationary, normalized (successive differences have zero mean) and positive, it is possible to omit the detrending procedure, because the detrending procedure is in this case equivalent to subtraction of the mean value of returns. Correspondingly, it is convenient to rewrite the following sum as

$$\sum_{\nu=1}^{N_s} |F(\nu, s)|^q = \sum_{\nu=1}^{N_s} |y_{\nu s} - y_{\nu(s-1)}|^q \,, \tag{2.29}$$

where can be recognized estimated probabilities

$$p_s(\nu) = |y_{\nu s} - y_{\nu(s-1)}| = \sum_{j=\nu(s-1)+1}^{\nu s} \xi_j \,. \tag{2.30}$$

Consequently, the sum of local fluctuations is equal to the partition function

$$Z(q, s) = \sum_\nu p_s(\nu)^q \tag{2.31}$$

and therefore

$$\tau_{DFA}(q) = qh(q) - 1 \,. \tag{2.32}$$

The method was originally constructed in mono-fractal version for $q = 2$. For $q = 1$, the procedure is related to R/S-analysis and Hurst exponent.

Short discussion is necessary at this place. The validity of the previous relation is restricted by necessity of detrending procedure for general series. From mathematical

point of view, if the empirical probability was created from detrended series, so $p_s(\nu) = \sum_{j=\nu(s-1)+1}^{\nu s}(\xi_j - \langle\xi\rangle)$, the probability is properly defined, which means that it is not a proper measure on the probability space. More discussion is contained in Ref. [35]. Apart from that, the generalized dimension calculated from scaling function is equal to

$$D_{DFA}(q) = \frac{qh(q) - 1}{q - 1}\,. \tag{2.33}$$

If the generalized dimension is a finite number, we automatically obtain that $h(1) = 1$, but it does not have to be true for an arbitrary series. This is again connected with the detrending issue. Some authors, as e.g. [36] use an alternative approach in estimation of generalized dimension. It is based on definition of *cumulant generating function* $K(q)$

$$\tau(q) = D(q - 1) - K(q) \tag{2.34}$$

where D is topological dimension. When dividing the previous equation by $(q - 1)$ we can define the *codimension function*

$$D(q) = D - C(q)\,. \tag{2.35}$$

In case of monofractal series, i.e., when the codimension function is equal to Hurst exponent, we obtain the familiar relation between Hurst exponent and fractal dimension

$$D_F = D - H\,. \tag{2.36}$$

We shall note that similarly to multifractal spectrum, there exists a *codimension spectrum* associated with $K(q)$ through Legendre transform:

$$c(\gamma) = \sup_q(q\gamma - K(q))\,. \tag{2.37}$$

It is possible to show that working with codimension function can partially overcome the problems with estimation of fractal dimensions that are present in techniques as R/S analysis or DFA. Alternatively, we can directly estimate the probability distributions and therefore obtain less pathological estimation of generalized dimension. The approach is based on estimation of Rényi entropy. After an introduction, we briefly compare the method with other methods discussed in this chapter.

2.3.4 Diffusion Entropy Analysis

In previous sections were presented methods that are built on descriptive statistical measures as rescaled range, mean, variance or more generally q-cumulants. In the cases when the underlying model exhibits power-law decay in probability distribution due to presence of extreme events or long-term memory, the theoretical statistics can be

indeterminable or infinite and the empirical counterparts are not describing the theoretical model correctly. Moreover, as discussed in previous sections, the above described approaches may not work properly for every series. As a consequence, we turn our attention to another multifractal method based on estimation of Rényi entropy called *Diffusion entropy analysis*, introduced originally by Scafetta et al. [37], in monofractal version based on Shannon entropy, and further generalized by Huang et al. [38]. The method is based on estimation of Rényi entropy. The advantage of entropy-based approaches is that they manage working with distribution with scaling exponents. As an example, let us consider a probability distribution with a single scaling exponent δ. This distribution can be directly written in the form

$$p(x,t)\mathrm{d}x = \frac{1}{t^\delta} F\left(\frac{x}{t^\delta}\right) \mathrm{d}x\,. \tag{2.38}$$

To this class of distributions belong e.g. Gaussian distribution or Lévy-stable distributions (their definition and basic properties can be found in Appendix A). The scaling exponent can be detected by calculation of differential (or continuous) Shannon entropy which is defined as

$$\mathcal{H}(t) = -\int_{\mathbb{R}} \mathrm{d}x\, p(x,t) \ln[p(x,t)] \tag{2.39}$$

which is in the case of distribution with single scaling exponent equal to

$$\begin{aligned}\mathcal{H}(t) &= -\int_{\mathbb{R}} \mathrm{d}x\, \frac{1}{t^\delta}\, F\left(\frac{x}{t^\delta}\right)\, \ln\left[\frac{1}{t^\delta}\, F\left(\frac{x}{t^\delta}\right)\right] = \\ &= -\int_{\mathbb{R}} \mathrm{d}y\, F(y)\, \ln\left[\frac{1}{t^\delta}\, F(y)\right] = A + \delta \ln t\,.\end{aligned} \tag{2.40}$$

When the system has several scaling exponents, we can measure its spectrum by measuring generalized dimension determined from Rényi entropy

$$\mathcal{I}_q(t) = \frac{1}{1-q} \ln \int_{\mathbb{R}} \mathrm{d}x\, p(x,t)^q\,. \tag{2.41}$$

For monofractal distribution with scaling exponent δ is the Rényi entropy equal to

$$\begin{aligned}\mathcal{I}_q(t) &= \frac{1}{1-q} \ln \int_{\mathbb{R}} \mathrm{d}x\, \frac{1}{t^{q\delta}} \left[F\left(\frac{x}{t^\delta}\right)\right]^q = \\ &= \frac{1}{1-q} \ln \int_{\mathbb{R}} \mathrm{d}y\, \frac{1}{t^{(q-1)\delta}} \left[F\left(y\right)\right]^q = B_q + \delta \ln t\end{aligned} \tag{2.42}$$

For distributions with more scaling exponents, we generally obtain the scaling exponents depending on q, so

$$\mathcal{I}_q(t) = B_q + \delta(q) \ln t\,. \tag{2.43}$$

The procedure of estimation of scaling exponent $\delta(q)$ is straightforward: we estimate the empirical probability distribution $\hat{p}$ from the time series, calculate the Rényi entropy and extract the scaling exponents from linear regression $\mathcal{I}_q(t) \sim B_q + \widehat{\delta}(q) \ln t$. The main challenge is the estimation of probability distributions such that the empirical Rényi entropy is approximated optimally. This is slightly different situation from the ordinary procedures known from theory of histograms. For our purposes, depending on parameter q, we do not have to estimate only $\hat{p}$, but also its powers, i.e. $\hat{p}^q$. Generally, this is the important point for all methods based on entropy estimation, not only multifractal methods. The next section is devoted to the proper estimation of probability histograms for estimation of Rényi entropy and subsequently δ-spectrum.

2.4 Estimation of Rényi Entropy and δ-spectrum

Estimation of entropies in general brings about several aspects that have to be properly discussed. The discussion covers the topics of probability distribution estimations, limitations in estimation procedure according to the particular value of q, and calculation of optimal bin-width for estimation of probability histograms. The discussion was done in Ref. [21] in connection with Diffusion entropy analysis, but can be also helpful in connection with other methods based on estimation of Rényi entropy. Similar discussions about applicability of particular methods and all technical details are done for DFA in Ref. [39] and for GHE in Ref. [31] .

2.4.1 Fluctuation Collection Algorithm

Most of the methods used for probability distribution estimation are established on the principle of repeating experiment and law of large numbers. It sets down that the empirical probability converges to the underlying theoretical probability distribution. This can be a problem in the case of time series, because the evolution of time series does not exhibit such behavior. Nevertheless, when we confine ourselves to the case of *stationary* time series, the estimation becomes tractable, because the properties of the stationary process do not depend on the particulary position in the series. This can be usually achieved by taking the *successive differences* (or returns in financial terminology) $\xi_j = x_j - x_0$. For estimation of probability distribution, we use, as in the case of GHE, the method of *moving time-window*. The fluctuation functions are for $j = \{1, \ldots, N - s\}$ defined as

$$\sigma_j(s) = \sum_{i=j}^{j+s} \xi_{i+s} = x_{j+s} - x_j \,. \tag{2.44}$$

The first expression is used when we work directly with noise-like series, in the case of walk-like (non-stationary) series, it is equivalent to use both approaches. All fluctuations

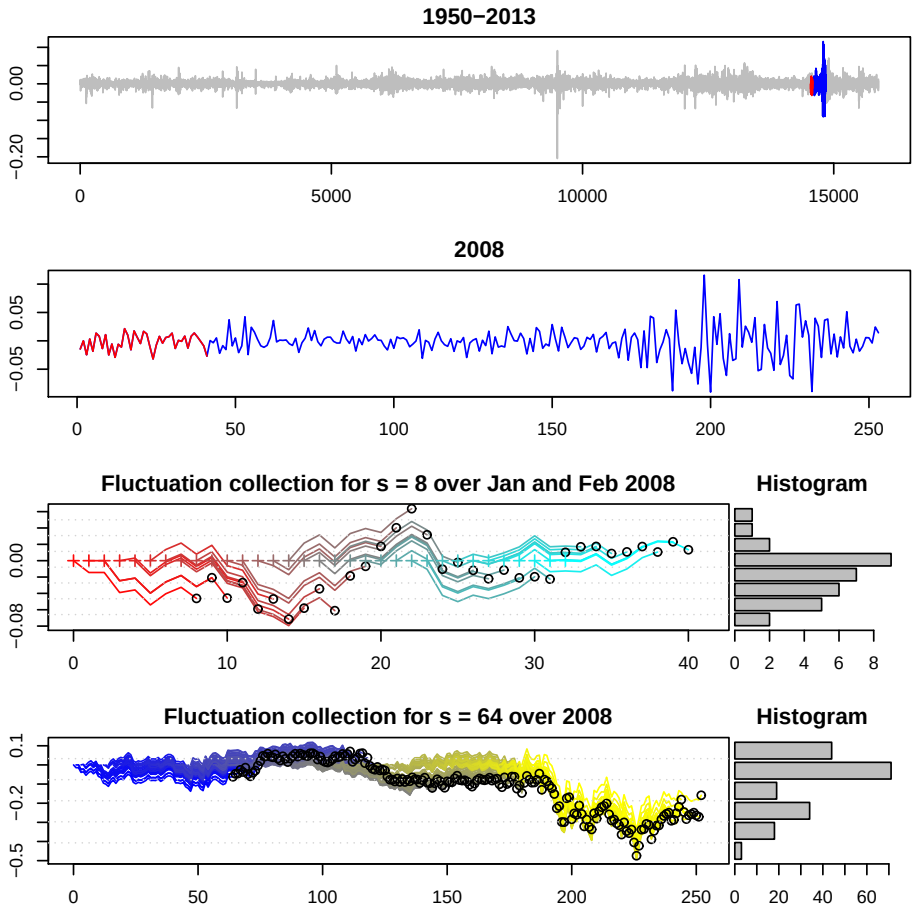


Figure 2.1: Fluctuation collection algorithm for the time series S&P500 in 2008. From above: a) Time series of S&P500 from January 1950 to March 2013, containing approximately 16000 entries.b) S&P500 for the year 2008. c) Fluctuation collection algorithm for the first two months of 2008 and $s = 8$ days. The series is partially integrated, i.e., fluctuation sums $\sigma_j(8)$ are collected into the histogram (right). d) Fluctuation collection algorithm for the whole year 2008 for $s = 64$ days. This histogram was estimated independently of the first histogram.

are divided into equidistant regions K_i of bin-width $h(s)$ and the probability is estimated as a normalized equidistant histogram

$$\hat{p}_i(s) \equiv \frac{\text{card}\{j \,|\sigma_j(s) \in K_i\}}{N - s + 1} \,. \tag{2.45}$$

For multidimensional data is the procedure similar, but K_i become hypercubes of volume h^D. The choice of bin-width influences substantially the estimated histogram and therefore it is necessary to find an optimal value of the bin-width. The algorithm is called *Fluctuation collection algorithm* because of its striking resemblance with diffusion of a

particle over the given time period. In the case of estimation of scaling exponents, we need to be able to estimate the scaling behavior simultaneously for several time scales. Fig. 2.1 illustrates the fluctuation collection algorithm on an example of financial time series S&P 500. The two histograms are estimated separately, i.e. on different scales leading to different bin-widths. We need to incorporate the parallel estimation on multiple scales to the calculation of optimal bin-width. This issue is broadly discussed in the next sections.

2.4.2 Histograms and Probability Distances

In this section we revise two classic topics of probability theory, namely histograms and distances on a probability space. Let us start with histograms. An equidistant histogram is a discrete approximation of an underlying probability distribution $p(x)$ defined as

$$\hat{p}(x) = \sum_{i=-\infty}^{\infty} \frac{p_i}{h} \chi_i(x)\,, \tag{2.46}$$

where χ_i is the characteristic function of K_i and $p_i = \int_{K_i} p(x)\mathrm{d}x$. In practical estimations, we work with finite data and the histogram is understood as an approximation of underlying PDF obtained from the data, so

$$\hat{p}(x) = \frac{1}{Nh} \sum_{i=1}^{n_B} \hat{\nu}_i \chi_i(x)\,, \tag{2.47}$$

where N is the length of the dataset $\{x_j\}_{j=1}^{N}$, n_B is number of bins, χ_i is the characteristic function of i-th bin $K_i = [x_{\min} + (i-1)h, x_{\min} + ih]$ and $\hat{\nu}_i$ is the number of elements that fall into K_i. The bin-width determines the number of bins, because it holds

$$n_B = \lceil \frac{x_{\max} - x_{\min}}{h} \rceil\,, \tag{2.48}$$

where $\lceil \cdot \rceil$ denotes the ceiling function, i.e. the smallest exceeding integer. Naturally, the q-th power of a histogram is equal to

$$\hat{p}^q(x) = \frac{1}{N^q h^q} \sum_{i=1}^{n_B} \hat{\nu}_i^q \chi_i(x)\,. \tag{2.49}$$

Our aim is to find such a histogram which is the optimal approximation of the underlying probability distribution with respect the Rényi entropy. The natural measure of discrepancy is the Rényi information divergence [40]:

$$D_q(p||\hat{p}) = \frac{1}{q-1} \ln \int_{\mathbb{R}} \mathrm{d}x\, \hat{p}^{1-q}(x) p(x)\,, \tag{2.50}$$

which represents the information lost (measured in Rényi entropy sense) when a distribution p is approximated by histogram $\hat{p}$. For $q \to 1$, we get the famous Kullback-Leibler information divergence [41]. Because we do not want to restrict ourselves to one histogram, which is only one representative outcome of N-times repeated joint experiment, we have to introduce the *expected Rényi information divergence,*

$$\langle D_q(p||\hat{p})\rangle_{\mathcal{H}} = \frac{1}{q-1}\left\langle \ln \int_{\mathbb{R}} \mathrm{d}x\, \hat{p}^{1-q}(x)p(x) \right\rangle_{\mathcal{H}}, \tag{2.51}$$

where $\langle\cdot\rangle_{\mathcal{H}}$ denotes the ensemble average over all admissible histograms. Unfortunately, working with this measure is intractable because of the $\log$ function in the expression. Therefore, we would have to really calculate the average over all possible histograms. The issue can be circumvented by approximation of Rényi divergence by other statistical distances with similar properties and yet computationally tractable. For this end, we firstly approximate the logarithm using Jensen inequality

$$1 - \frac{1}{z} \leq \ln z \leq z - 1 \tag{2.52}$$

and obtain that

$$|D_q(p||\hat{p})| \leq \frac{c_q}{q-1}\int_{\mathbb{R}} \mathrm{d}x\, |\hat{p}^q(x) - p^q(x)|\,. \tag{2.53}$$

This is a generalization of Csiszár—Kullback inequality [42] between Rényi divergence and L_1-distance between q-th powers. In Ref. [21] is shown that the constant c_q is equal to

$$c_q = \max\left\{1, \left(\int_{\mathbb{R}} \mathrm{d}x\, \hat{p}^{1-q}(x)p^q(x)\right)^{-1}\right\}. \tag{2.54}$$

Finally, from previous inequality together with Hölder inequality we obtain that

$$|D_q(p||\hat{p})|^2 \leq \frac{c_q^2}{(q-1)^2}\left(\int_{\mathbb{R}} \mathrm{d}x\, |\hat{p}^q(x) - p^q(x)|\right)^2 \leq \frac{c_q^2}{(q-1)^2}\int_{\mathbb{R}} \mathrm{d}x |\hat{p}^q(x) - p^q(x)|^2. \tag{2.55}$$

The main advantage of using L_2 (or L_1) norm consists in the fact that the ensemble average can be interchanged with the integral, so

$$\left\langle \|\hat{p}^q - p^q\|_{L_2}^2 \right\rangle_{\mathcal{H}} = \int_{\mathbb{R}} \mathrm{d}x \langle (\hat{p}^q(x) - p^q(x))^2\rangle_{\mathcal{H}} \tag{2.56}$$

and therefore ensemble averaging acts to the histogram only locally. Consequently, we do not have to average over all frequencies $\{\hat{\nu}_i\}_{i=1}^{n_B}$ such that $\nu_i \in \{1,\ldots,N\}$ and $\sum_{i=1}^{n_B}\nu_i = N$. We can average only over one frequency ν_i, which is a significant simplification in calculations.

Let us mention that it is also possible to work with the previously mentioned L_1-distance. The reason of working rather with L_2-distance is twofold: first, it is computationally slightly more tractable and second, most of the authors work with L_2-distance and therefore we can compare our results with classic results known from theory of histograms. Generally, it is plausible to assume that in this 1-dimensional optimization problem (the only parameter is the bin-width) are optimal results under previously mentioned distance measures to similar results.

2.4.3 Dependence of Bin-width on q and Optimal Bin-width

At this place arises a natural question: is the optimal bin-width depending on the Rényi parameter q or is it enough to estimate the optimal bin-width for one parameter, e.g., for $q = 1$, and use it for all entropies? In the following discussion we show that it is necessary to calculate the optimal bin-width separately for each q. We denote $\Delta(x) = p(x) - \hat{p}(x)$. The L_2 squared distance between probability distribution and histogram is equal to

$$\begin{aligned} \|p^q - \hat{p}^q\|_{L_2}^2 &= \int_{\mathbb{R}} \mathrm{d}x \, (p^q(x) - \hat{p}^q(x))^2 = \int_{\mathbb{R}} \mathrm{d}x \, \left(p^q(x) - \frac{1}{h^q} \sum_{i=1}^{n_B} \hat{p}_i^q \chi_i(x) \right)^2 = \\ &= \int_{-\infty}^{x_{min}} \mathrm{d}x \, p^{2q}(x) + \sum_{i=1}^{n_B} \int_{K_i} \mathrm{d}x \, \left(p^q(x) - \frac{\hat{p}_i^q}{h^q} \right)^2 + \int_{x_{max}}^{\infty} \mathrm{d}x \, p^{2q}(x) \end{aligned}$$

Assuming that $\Delta(x)$ is sufficiently small, we can approximate the distribution as

$$p(x)^q = \left[\frac{\hat{p}_i}{h} \right]^q + \binom{q}{1} \left[\frac{\hat{p}_i}{h} \right]^{q-1} \Delta(x) + \mathcal{O}(\Delta(x)^2) \tag{2.57}$$

Subsequently, the distance can be approximated as

$$\|p^q - \hat{p}^q\|_{L_2}^2 \approx \int_{-\infty}^{x_{min}} \mathrm{d}x \, p^{2q}(x) + q^2 \sum_{i=1}^{n_B} \left(\left[\frac{\hat{p}_i}{h} \right]^{2(q-1)} \Delta_i^2 \right) + \int_{x_{max}}^{\infty} \mathrm{d}x \, p^{2q}(x) \,,$$

where $\Delta_i^2 = \int_{K_i} \mathrm{d}x \Delta(x)^2$. We use the following notation

$$\|p^q - \hat{p}^q\|_{L_2}^2 \approx \Delta_0^{2q} + \mathfrak{S}_q^2 + \Delta_{n_B}^{2q} \,. \tag{2.58}$$

The middle sum $\mathfrak{S}_q^2$ depends only on the choice of histogram and therefore on h. We divide the discussion into three cases:

- $q \leq 0$**:** the sum accentuates extremely small probabilities $\hat{p}_i$. This can be compensated by larger bin-width. However, especially for distributions with extremely

small probabilities it is very hard to decide whether the probability is zero or not, and corresponding problem with definition 0^q for $q < 0$. Consequently, the estimation of Rényi entropy is extremely sensitive (for $q < 0$ is Rényi entropy not a properly defined information measure; see Sect. 4.3.1) and most authors do not calculate histograms for negative q's.

- $0 < q < 1$**:** the exponent in the sum is larger that -1, therefore the small probabilities are accentuated, but not in a drastic way.

- $1 \leq q$**:** the error is diminished, because the error in $p_i^{2(q-1)}$ is suppressed. Against this is the factor $h^{2(1-q)}$ which is accentuated for small h, thus it is convenient to choose larger bin width and not to over-fit the histogram.

The previous discussion indicates that it is necessary to choose different bin-widths for different values of q and one common bin-width for all Rényi parameters would not sufficiently approximate the underlying probability distribution.

In order to find the optimal bin-width, there have been used several approaches. In this connection it is necessary to mention the popular *Sturges rule* [43], which is based on estimation of histograms for binomial distributions. It estimates the optimal number of bins as $n_B = 1 + \log_2 N$. However, this rule is good rather for data visualization, but in case of probability distribution approximation, most authors prefer the approach based on *integrated mean square error minimization*. We utilize the previously discussed L_2 distance between q-th powers and formulate the problem as minimization of the term

$$\min_{h>0} \int_{\mathbb{R}} \mathrm{d}x \left\langle (p^q(x) - \hat{p}^q(x))^2 \right\rangle_{\mathcal{H}} = \min_{h>0} \sum_{i=1}^{n_B} \int_{K_i} \mathrm{d}x \left\langle \left(p_q(x) - \frac{\nu_i^q}{N^q h^q} \right)^2 \right\rangle_{\nu_i} . \quad (2.59)$$

First, the integrand, which is nothing else than the *local mean squared error*, can be rewritten as (we omit the subindex ν_i)

$$\left\langle \left(p_q(x) - \frac{\nu_i^q}{N^q h^q} \right)^2 \right\rangle = \left\langle \left(\frac{\nu_i^q}{N^q h^q} - \left\langle \frac{\nu_i^q}{N^q h^q} \right\rangle \right)^2 \right\rangle + \left(\left\langle \frac{\nu_i^q}{N^q h^q} \right\rangle - p^q(x) \right)^2 \quad (2.60)$$

where the first term represents variance of $\hat{p}^q(x)$ and the second term corresponds to squared bias of $\hat{p}^(x)$ with respect to $p(x)$. In both cases, we need to calculate at first the expectation value of ν_i^q. In the theory of histograms can be easily shown that the frequency fulfills the binomial distribution $\nu_i \sim Bi(N, p_i)$, where p_i is the probability of i-th bin. Hence, we have to calculate the fractional moment of binomial distribution, which is not analytically possible. In the case when we have enough statistics, we can approximate the distribution by Gaussian distribution (this is a consequence of

Central limit theorem), so $Bi(N,p) \sim \mathcal{N}(Np, Np(1-p))$. Consequently, we are able to calculate the fractional moment as

$$\langle \nu_i^q \rangle \approx \int_R \mathrm{d}z |z|^q \frac{1}{\sqrt{2Np_i(1-p_i)}} \exp\left(-\frac{(z-Np_i)^2}{2Np_i(1-p_i)}\right) \tag{2.61}$$

which can be expressed in the closed-form in terms of *confluent hypergeometric functions*. The procedure is also presented in [21]. We have used the absolute moment $\langle |z|^q \rangle$, because it ensures that the results remains real. Using the leading term approximation, the fractional moment can be expressed as

$$\langle \nu_i^q \rangle = N^q p_i^q \left(1 + \frac{q(q-1)}{2}\frac{1-p_i}{Np_i} + \mathcal{O}(N^{-2})\right). \tag{2.62}$$

With that is the local variance equal to

$$\left\langle \left(\frac{\nu_i^q}{N^q h^q} - \left\langle \frac{\nu_i^q}{N^q h^q} \right\rangle\right)^2 \right\rangle = \frac{q^2 p_i^{2q-1}(1-p_i)}{h^{2q}N} + \mathcal{O}(N^{-2}) \leq \frac{q^2 p_i^{2q-1}}{h^{2q}N} + \mathcal{O}(N^{-2}) \tag{2.63}$$

Similarly,

$$\left\langle \frac{\nu_i^q}{N^q h^q} \right\rangle - p^q(x) = \frac{p_i^q}{h^q} - p^q(x) + \mathcal{O}(N^{-1}). \tag{2.64}$$

When calculating the integrated error, we approximate the probability p_i, so for $\xi \in K_i$

$$p_i^q = h^q p^q(\xi) + q h^{q-1} p^{q-1}(\xi) h \left(\frac{h}{2} - \xi\right) \frac{\mathrm{d}p(\xi)}{\mathrm{d}\xi} + \mathcal{O}(h^{q+2}). \tag{2.65}$$

With this *leading order approximation* is possible to show that the mean squared integrated error is equal to

$$\int_{\mathbb{R}} \mathrm{d}x \, \langle p^q(x) - \hat{p}^q(x) \rangle_{\mathcal{H}} \overset{l.o.}{=} \frac{q^2}{Nh} \int_{\mathbb{R}} \mathrm{d}x \, p^{2q-1}(x) \mathrm{d}x + \frac{h^2}{12} \int_{\mathbb{R}} \mathrm{d}x \, \left(\frac{\mathrm{d}p^q(x)}{\mathrm{d}x}\right)^2. \tag{2.66}$$

The only dependence on the histogram parameters is now remaining on the bin-width h. The dependence on parameter h is depicted in Fig. 2.2. When we minimize the error with respect to h, we obtain

$$h_q^* = \left(\frac{6q^2}{N} \frac{\int_{\mathbb{R}} \mathrm{d}x \, p^{2q-1}(x)}{\int_{\mathbb{R}} \mathrm{d}x (\mathrm{d}p^q(x)/\mathrm{d}x)^2}\right)^{1/3}. \tag{2.67}$$

When we assume that the underlying model is driven by the normal distribution $\mathcal{N}(\mu, \sigma^2)$, the integral converges for $q > \frac{1}{2}$, and the formula can be rewritten as

$$h_q^* = \sigma N^{-1/3} \sqrt[3]{24\sqrt{\pi}} \frac{q^{1/2}}{\sqrt[6]{2q-1}} = h_1^* \rho_q. \tag{2.68}$$

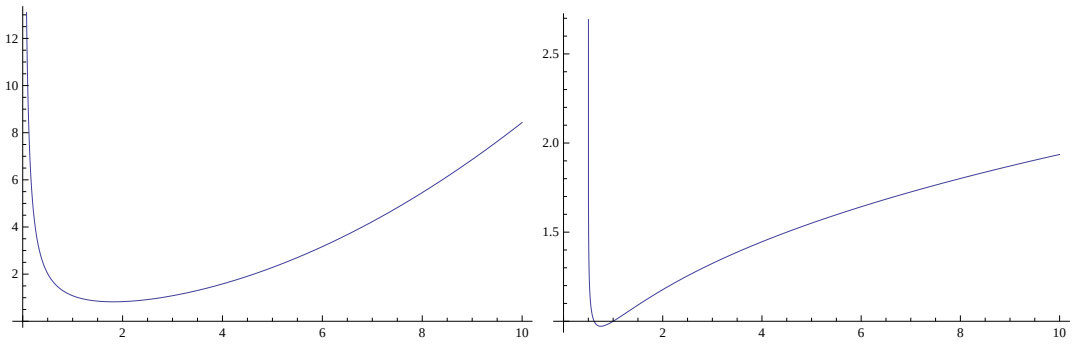


Figure 2.2: Left: Shape of asymptotic mean squared error for $q = 1$ as a function of h (the choice of N and σ determines $MSE(h) = \frac{1}{h} + \frac{h^2}{12}$). Right: Plot of ρ_q. For large q's it is similar to $q^{1/3}$, but it starts to diverge for values close to $q = \frac{1}{2}$.

For $q = 1$ we recover the classic result of Scott [44], which is for other values of q only multiplied by factor $\rho_q = \frac{q^{1/2}}{\sqrt[6]{2q-1}}$ (the function ρ_q is shown in Fig. 2.2). In practical estimation, the theoretical standard deviation is replaced its empirical counterpart, so we obtain a generalization of familiar *Scott rule* [44]

$$\hat{h}_q^{Sc} = 3.5\hat{\sigma}N^{-1/3}\rho_q\,. \tag{2.69}$$

Alternatively, in cases when the standard deviation is not a good statistics (because of distribution kurtosis, presence of heavy-tails or asymmetry), we can replace the standard deviation by a multiple of the *interquartile range* (IQR), i.e. the difference between first and third quartile of the distribution. The transformation coefficient is given by the IQR of normal distribution, which is

$$IQR(\mathcal{N}(\mu,\sigma^2)) = 2\sqrt{2}erfc^{(-1)}(1/2)\sigma \approx 1.349\sigma\,. \tag{2.70}$$

With replaced interquartile range, the bin-width rule is expressible as

$$\hat{h}_q^{FD} = 2.6\,\widehat{IQR}\,N^{-1/3}\rho_q\,. \tag{2.71}$$

The approach is inspired by the original method of Freedman and Diaconis [45].

When $q \leq \frac{1}{2}$, the integral in Eq. 2.67 does not converge for distributions with unbounded support. The situation can be in principle patched by the assumption of *truncated distribution*, i.e. distribution with finite support. Nonetheless, the choice of the particular distribution heavily influences the optimal bin-width and one would need to know exactly the theoretical form of the underlying distribution.

For comparison, when the Normal distribution is replaced by the Lévy-stable distribution with stability parameter $\alpha < 2$, one immediately derives a new limit for convergence of the integral in Eq. 2.67, which is

$$q_L > \frac{1}{2} + \frac{1}{2(\alpha+1)}\,. \tag{2.72}$$

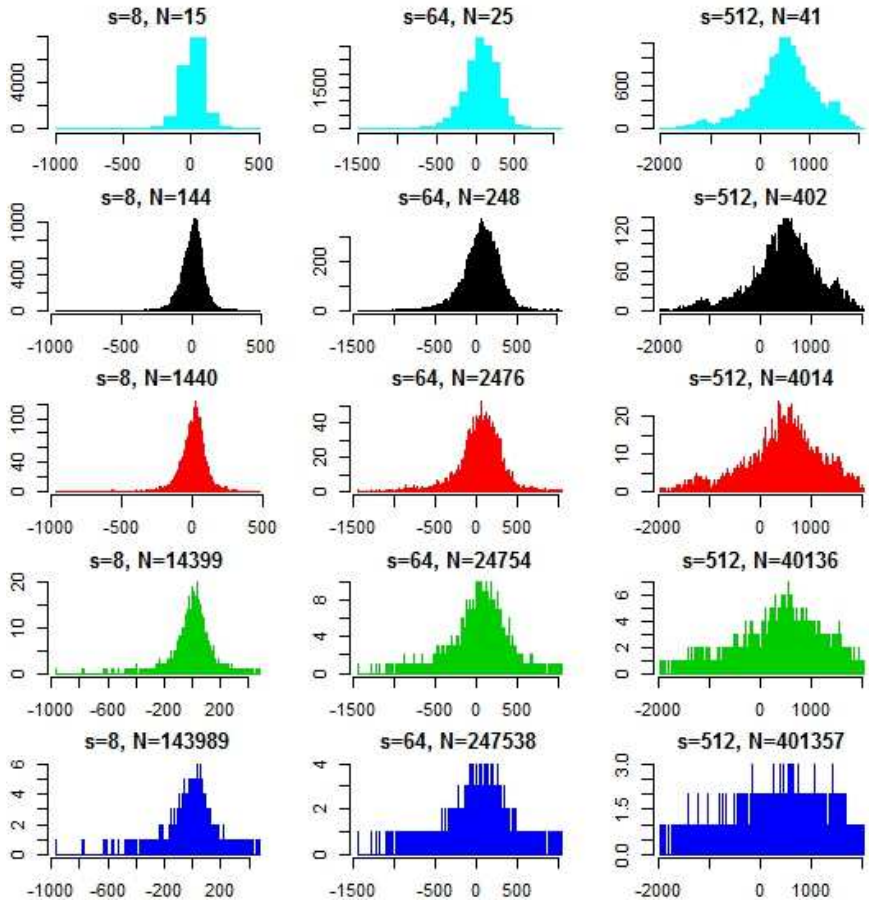


Figure 2.3: Un-normalized (frequency-based) histograms of the fluctuation sums obtained from time series $S\&P500$, with $s = 8$, 64 and 512 with bin-widths $h = 100$, 10, 1, $0,1$ and $0,01$; measured in units $u = 3 \times 10^4$ for better visualization. We can observe underfitted and overfitted histograms.

In the case of estimation of δ-spectrum, one has to estimate the Rényi entropy on several bin-widths to be able to estimate the scaling exponent from the linear regression. Let us have a set of characteristic scales $Sc = \{s_i\}_{i=1}^{m}$. The particular choice of characteristic scales depends on the problem, but one can find a general rule which is working in most cases that $Sc = \{K2^i\}_{i=1}^{i_{max}}$, where i_{max} is determined by the length of the dataset. This choice is desirable because of two reasons: in log-linear plot, the entropies are distributed uniformly, and the complexity of algorithm remains $\mathcal{O}(N \log N)$. The optimal bin-width is determined by the *total asymptotic mean integrated squared error*, so we

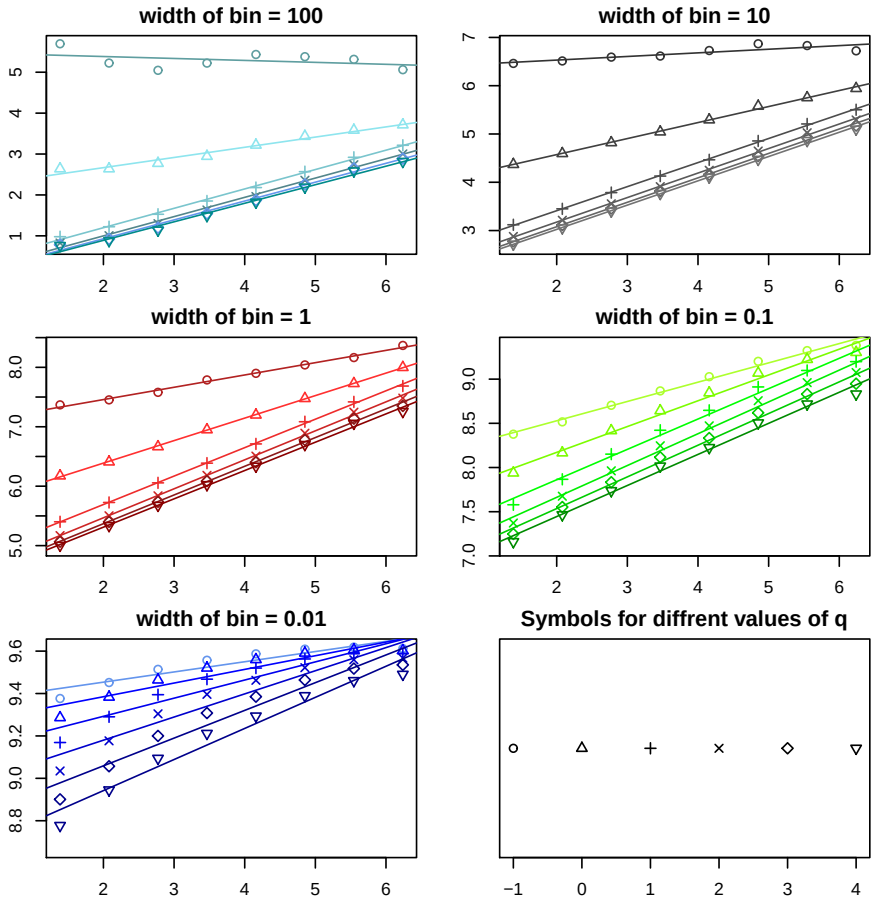


Figure 2.4: Linear fits of estimated RE vs. $\ln s$. The error from is from histograms distributed to fitting procedure of $\delta(q)$ spectrum.

have to optimize

$$\min_{h>0} \sum_{i=1}^{i_{max}} \left(\frac{q^2 (2\pi)^{1-q} \sigma_{s_i}^{2(1-q)}}{N_{s_i} h \sqrt{2q-1}} + \frac{h^2}{12} q^{1/2} \pi^{-(1/2+q)} \sigma_{s_i}^{-(1+2q)} \right) . \tag{2.73}$$

Where $N_{s_i} = N - s_i + 1$ and σ_{s_i} is the standard deviation on the scale s_i. From previous relations one immediately obtains the optimal bin-width as

$$h_q^*(\mathcal{S}) = (24\sqrt{\pi})^{1/3} \rho_q \sqrt[3]{\frac{\sum_{i=1}^{i_{max}} \sigma_{s_i}^{2(1-q)} / N_{s_i}}{\sum_{i=1}^{i_{max}} \sigma_{s_i}^{-(1+2q)}}} \tag{2.74}$$

Unfortunately, we do not obtain the bin-width in the factorized form, i.e., as the product of ρ_q and a q-independent part. The empirical bin-width is obtained similarly to

the previous cases. To illustrate necessity of proper estimation of histogram bin-width, Figs. 2.3 show the histograms of one particular financial time series (S&P 500) estimated for several bin-widths. One clearly distinguishes that the histograms for too large bin-width are *underfitted* (i.e., loose too much information), while histograms for too small bin-width are *overfitted* (i.e., we do not obtain enough statistics for most bins). Moreover, Fig. 2.4 shows subsequent fits of $\delta(q)$ estimated from the presented histograms. We see that the errors are transferred to spectrum estimations, too.

2.5 Applications of Multifractals in Physics

In this section, we discuss possible applications of multifractal analysis into physics and other related fields. Interestingly, the presented concepts find their applications also in financial models. This conjunction was a cornerstone for the formation of econophysics, and multifractal models still remain one of the most important parts in the branch. We focus on applications in hydrodynamics and meteorology conveyed by the concept of multiplicative cascades and the connection of multifractals with thermodynamical systems.

2.5.1 Multifractal Cascades and Deformations

The theory of multiplicative cascades was formulated by A.N. Kolmogorov in 1940 [3]. The theory was originally used in the connection with description of fully developed turbulence, however, it found many other applications as e.g. description of chaotic systems [46], or rainfalls in climatic models [47]. The theory is based on assumption that large vortices are compound of eddies on smaller scale in some characteristic way. We define a sequence of typical scales $r_0 > r_1 > \dots r_n$. One can define a typical ratio between two typical scales, i.e., $l_i = \frac{r_n}{r_{n-1}} < 1$, so $r_m = r_0 \prod_{j=1}^{m} l_j$. These scales define a set of distinct regions on each scale r_j, which is denoted as $\{K_i^j\}_{i=1}^{i_{max}}$, where i_{max} is determined by the nature of the system. We denote a characteristic quantity (often energy of the system) as E. This quantity is defined by its density function $\epsilon(x)$, so

$$E(\Omega) = \int_{x\in\Omega} \epsilon(x)\mathrm{d}x\,. \tag{2.75}$$

In the framework of multiplicative cascades, the quantity is defined on the typical scales as a product of *multipliers*, so

$$E_{r_n}(K_i^j) = E_{r_0} \prod_{j=1}^{n} \mathcal{M}_{j,i_j}\,. \tag{2.76}$$

The limit $n \to \infty$ should converge to the density function. Thus, the cascade is defined set of scales l_j and multipliers $\mathcal{M}_j$. There are two classes of multiplicative cascades. In

the first case, we assume that all multipliers $\mathcal{M}_{j,i}$ are for all regions K_i^j deterministic functions. The constant E_{r_0} therefore represents a normalization of the quantity in the region K^0. The straightforward generalization enables to define multipliers as random variables. The normalization is then determined by the mean values $\langle E_{r_0} \rangle$.

Let us mention a few popular models of multiplicative cascades. In the original work of Kolmogorov [3] was considered an isotropic distribution of multipliers, so the only parameter of the model is the normalization $\langle E_{r_0} \rangle$. The other popular examples provide cascades with multipliers obeying log-normal distribution, β-model, where a fraction (usually denoted as β) of multipliers $\mathcal{M}_j$ is nonzero and the rest is equal to zero.

Let us turn the attention to another class of multiplicative cascades which incorporates several characteristic scaling exponents and therefore with a good potential to describe multifractal systems. The definition of cascades based on scale multipliers is naturally predestined for modeling multifractal systems. The simplest version of *multifractal cascade* is *binomial cascade*, which serves as a springboard for more sophisticated models. It is a deterministic cascade with binomial division rule (i.e., $l_j = \frac{1}{2}$), when the multipliers are $\mathcal{M}_{j,1} = p$ and $\mathcal{M}_{j,2} = (1 - p)$. Analogously, one could define a *multinomial cascade* for $l_j = \frac{1}{n}$. The important property is the *conservation* of the cascade, so

$$\sum_{i=1}^{n_j} \mathcal{M}_{j,i} = 1 \,. \tag{2.77}$$

A straight generalization of binomial cascade is the *microcanonical cascade*, where we assume that multipliers p and $1 - p$ are randomly assigned to $\mathcal{M}_{j,1}$ and $\mathcal{M}_{j,2}$. Also this model represents a cascade with conservation. The disadvantage of the system is the fact that the multipliers are not statistically independent random variables. The statistical independence of variables can be reanimated when we assume only *statistical conservation*, i.e., we assume only

$$\langle \sum_{i=1}^{n_j} \mathcal{M}_{j,i} \rangle = 1 \,. \tag{2.78}$$

If the multipliers are identically distributed, we obtain $\langle \mathcal{M}_{j,\cdot} \rangle = l_j$. This model is called *canonical cascade*, because the analogy with (micro)canonical ensembles in thermodynamics, where the conservation rules are also expressed either in the strict form or in the statistical form.

Multifractal properties can be naturally investigated with help of codimension function defined in Sect. 2.3.3. When we assume that $l_j = \frac{1}{\lambda}$, i.e., $r_n = \frac{r_0}{\lambda^n}$. We suppose that moments of multipliers fulfill the following scaling rule

$$\langle \mathcal{M}^q \rangle \propto \lambda^{K(q)} \tag{2.79}$$

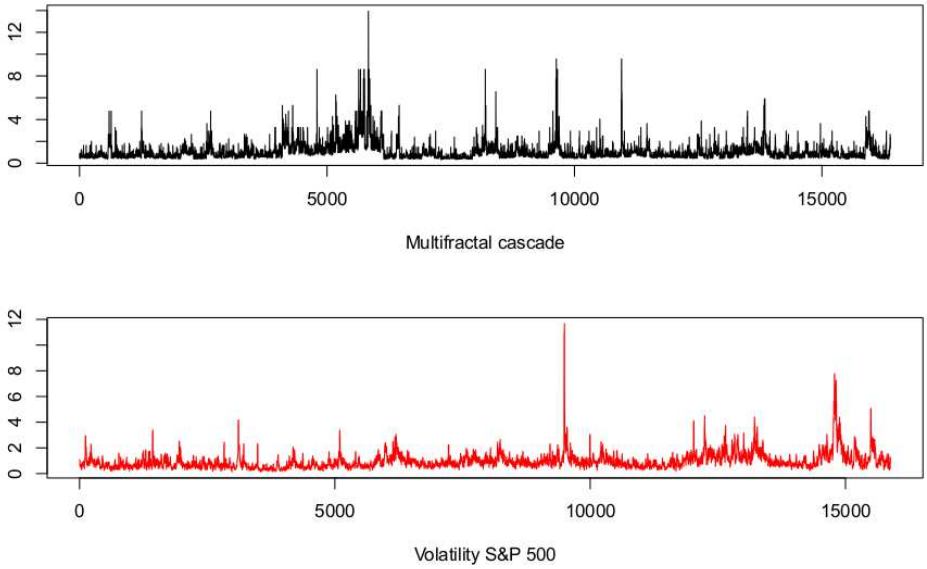


Figure 2.5: Simulation of time-dependent volatility modeled as a multifractal cascade and comparison with 20-day volatility of S&P 500 index.

When we generalize the scaling rule to any positive scale λ, the scaling exponent $K(q)$ corresponds to cumulant generating function (defined in Sect. 2.3.3). The relation to multifractal spectrum is discussed in [36].

In one-dimensional case, the cascade defines a *multifractal measure*, which can be successfully used in modeling of multifractal systems. Let us have a multifractal canonical cascade $m_n(x) = m_0 \prod_{j=1}^{n} \mathcal{M}_i^j(x)$. The cascade forms a sequence of measures, so we have

$$\mu_n[a,b] = \int_a^b m_n(x)\mathrm{d}x\,. \tag{2.80}$$

The limit $\mu = \lim_{n\to\infty} \mu_n$ is defined in the natural sense of the measure theory. We define the *time deformation* as

$$\theta(t) = \mu[0,t]\,. \tag{2.81}$$

The time deformation can bring the multifractal nature into a process $X(t)$ with a simple scaling. The time deformation has a nice interpretation, which says that it is a transformation between two times: one, physical, objective time of external observer and the second, inner time of the system. In the inner time is the process simply $X(t)$, but of an external observer, one has to transform the internal time into the clock time $\tau = \theta(t)$, so the process becomes $X(\tau) = X(\theta(t))$. In many systems, the time difference is proportional to standard deviation of the system, so alternatively, the time deformation can be interpreted as time-dependent standard deviation (in financial theory known as

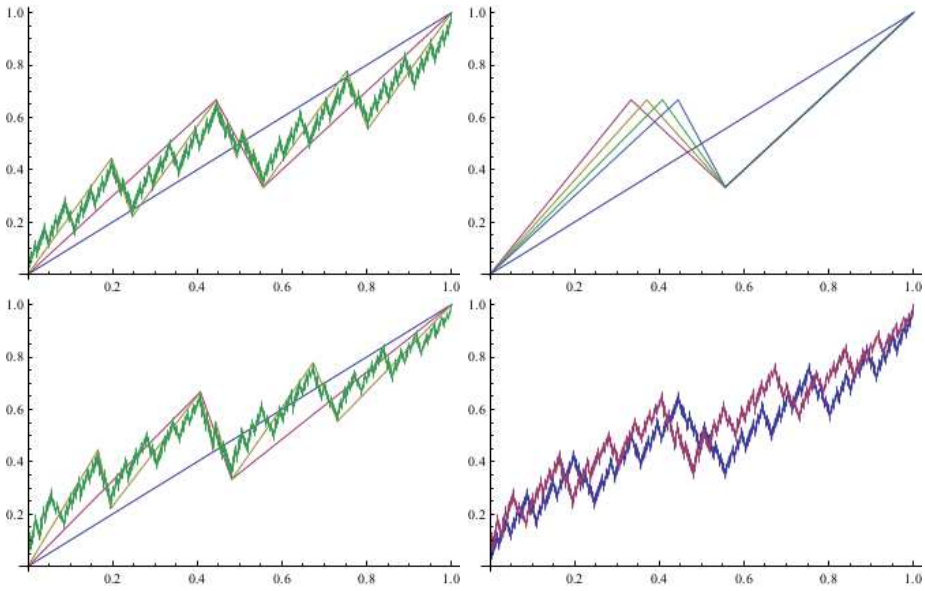


Figure 2.6: Example of recursive generation of multifractal patterns. The top-left figure represents the Wiener pattern with constant scaling $\Delta x = (\Delta t)^{1/2}$. The top-right figure shows possible changes of Wiener pattern to obtain multifractal patterns. These patterns are chosen randomly in each step. The bottom-left pattern represents the resulting multifractal pattern. The difference between the Wiener pattern and a representative multifractal pattern (displayed in bottom-right figure) generates the time deformation.

volatility) of the underlying probability distribution. As an example, Fig. 2.5 shows comparison of multifractal cascade and volatility of a financial series.

Alternatively, the time deformation can be created by generation of so-called *multifractal patterns*. This approach was invented by B. B. Mandelbrot [48] and is based on generation of patterns with typical scaling exponents. An example of such multifractal pattern is illustrated in Fig. 2.6.

2.5.2 Multifractal Thermodynamics

The connection of multifractal formalism with concept of thermodynamics represents another important interpretation of multifractal analysis and shows us possible applications in thermodynamical systems. Identification of multifractal scaling exponents with thermodynamical quantities was a starting point for many applications in many fields, chaotic systems are just one example [49]. It is also a good argument for using associated Rényi entropy in thermodynamical systems [50]. The connection to thermodynamics can be established via the partition function

$$Z(q,s) = \sum_i p_i^q(s) = \sum_i \exp(-\beta E_i) \tag{2.82}$$

where E_i are energies of the system. When probabilities are scaling as $p_i(s) \propto s^{\alpha_i}$, we can immediately identify multifractal exponents with thermodynamical quantities. Therefore we obtain

$$E_i(s) = -\ln(p_i(s)) = -\alpha_i \ln s\,. \tag{2.83}$$

Additionally, we can interpret the Rényi parameter q as

$$q = \beta\,. \tag{2.84}$$

where β is the inverse temperature. The connection to Rényi entropy is given as

$$\mathcal{I}(q,s) = \frac{1}{1-q}\Psi(q,s) = \frac{1}{q-1}\ln Z(q,s)\,. \tag{2.85}$$

The function $\Psi(q,s)$, the negative logarithm of partition function, is nothing else than a multiple of thermodynamical *free energy*

$$F_q = \frac{1}{\beta}\ln Z(q,s) = -\frac{1}{\beta}\Psi(q,s)\,. \tag{2.86}$$

It is also connected to the escort distribution

$$\rho_i^q(s) = \frac{p_i^q(s)}{\sum_j p_j^q(s)} \sim \exp(\Psi(q,s) - \beta E_i(s))\,. \tag{2.87}$$

An interesting is the relation to the multifractal spectrum. When we use abbreviation

$$V = -\ln s\,, \tag{2.88}$$

the function $\Psi(q,s)$ can be rewritten (similarly to Sect. 2.2)

$$\Psi(q,s) = -\ln \int \mathrm{d}\alpha\, \exp[(f(\alpha) - q\alpha)V]\,. \tag{2.89}$$

According to stationary phase approximation we obtain that

$$\Psi(q) \sim [q\alpha(q) - f(\alpha(q))]V\,. \tag{2.90}$$

From the correspondence of $\Psi(q)$ to free energy and the fact that the Legendre structure of the thermodynamics is preserved even for the general case [51], we end with

$$\Psi(q) = qU_q - S_q = q\alpha(q)V - f(\alpha(q))V\,. \tag{2.91}$$

Naturally, the term $\alpha(q)V = \alpha(q)\ln s$ represents the average energy given by q-averaging

$$a(q) = \langle a \rangle_q = \sum_i \rho_i^q a_i \tag{2.92}$$

and correspondingly we obtain that

$$U_q = \langle E \rangle_q = \sum_i \rho_i^q E_i = a(q) \ln s \,. \tag{2.93}$$

The second term in Eq. (2.90) can be interpreted as the thermodynamical entropy of the system, so we obtain that

$$\lim_{V\to\infty} (S_q/V) = f(\alpha(q)) \tag{2.94}$$

$$\lim_{V\to\infty} (\Psi(q)/V) = \tau(q) \,. \tag{2.95}$$

The limit $V \to \infty$ corresponds to multifractal limit $s \to 0$. As a result, we obtain analogical relations to thermodynamic Maxwell equations

$$\frac{\partial \Psi(q)}{\partial q} = U_q \tag{2.96}$$

$$\frac{\partial S_q}{\partial U_q} = q \,. \tag{2.97}$$

These thermodynamical relations ordain the relation between informational entropy and thermodynamical entropy, because we have

$$\mathcal{I}_q = \frac{S_q - q\, U_q}{q-1} \,. \tag{2.98}$$

For $q \to 1$ the relation boils down to the classic relation between thermodynamical and information entropy. Apart from multifractal thermodynamics, there exist other concepts of thermodynamics going beyond classic scope of Shannon entropy, for example non-extensive thermodynamics based on Tsallis entropy which is briefly discussed in Section 4.3.2.

This section has presented some possible applications of multifractals in physical systems. Indeed, there exist many other interesting multifractal models in biology, cosmology, theory of complex systems, etc. A nice overview of applications of multifractals provide Refs. [25].

Chapter 3

Models of Anomalous Diffusion

Diffusion can be observed in many processes in the nature. Nevertheless, sometimes is the description quite complicated because of emergent phenomena, long-range correlations, etc. In this chapter we go through several models of anomalous diffusion and discuss their properties. For this end, we also introduce so-called fractional calculus, a mathematical tool which is a generalization of ordinary calculus for non-natural orders. Thereafter, we compare several models of generalized diffusion. Particularly interesting is the double-fractional model, which incorporates both spatial and temporal anomalous scaling exponents and can be expressed in several representations, including kernel representation and integral representation.

3.1 Brownian Motion and Diffusion Equation

Brownian motion is the most popular and easy-to-understand model of random movement. It was firstly experimentally discovered by a biologist R. Brown during observation of pollen grains in the water. Since that time, it has found many theoretical descriptions as well as practical applications in many fields not only including physics, but basically in every scientific branch, where some uncertainty is present in the system. Theoretical description of the Brownian motion was done by A. Einstein and M. Smoluchowski at the beginning of twentieth century. They have found that the mean squared displacement is proportional to time $\langle x(t)^2 \rangle \propto t$. Theoretical description was done by P. Langevin, N. Wiener and many other scientists. The most common mathematical description of diffusion processes is given by the diffusion equation

$$\frac{\partial p(x,t)}{\partial t} = D \frac{\partial^2 p(x,t)}{\partial x^2} \,. \tag{3.1}$$

To determine the solution completely, it is necessary to impose some boundary conditions. The most common is to set two initial conditions

$$p(x,t)|_{t=0} = f_0(x)\,, \qquad \frac{\partial p(x,t)}{\partial t}|_{t=0} = f_1(x)\,. \tag{3.2}$$

When $f_0(x) = \delta(x)$ and $f_1(x) = 0$ we obtain the well-known Gaussian distribution

$$p(x,t)\mathrm{d}x = \frac{1}{\sqrt{2\pi D}} \exp\left(-\frac{x^2}{2Dt}\right) \mathrm{d}x\,. \tag{3.3}$$

From the mathematical point of view, Brownian motion can be described as a stochastic process. This process is called *Wiener process* [1] and is usually denoted as $W(t)$. The process is defined as a process with stationary increments with Gaussian distribution proportional to time. As discussed in the chapter about multifractals, the Wiener process has the fractal dimension equal to 2 in two or more dimensions, and therefore the representative trajectories are not differentiable. This is easy to see from the scaling relation

$$\frac{\langle |x(t+h) - x(t)| \rangle}{h} \propto \frac{\sqrt{h}}{h} \to +\infty \quad \text{for } h \to 0\,. \tag{3.4}$$

To the important properties of Wiener process belongs the *Markov property* which points to the absence of long-term memory in the diffusion. Hence, the full information about the process is encoded in the last observed value. As discussed in Sect. 2.1, the diffusion process has the scaling discovered by Einstein, i.e. $|\Delta x| \propto \Delta t^{1/2}$. The scaling properties are the most important in the possible generalizations of diffusion processes. The resulting scaling is determined by the *Central limit theorem*. From this perspective, the Diffusion process can be seen as a limit of a discrete process of random variables with independent increments and finite variance. The diffusion process is also important from the perspective of entropies, because it is the MaxEnt distribution under the constraint of zero mean and standard deviation proportional to time, i.e. $\langle x^2(t) \rangle = Dt$.

T Brownian motion the most popular diffusion process, nevertheless, models based on Brownian motion are not able to describe certain types of systems. It is usually the situation when some kind of complex behavior is observed. As an example, let us mention processes with presence of memory effects. This is usually the motivation for using some generalizations of Brownian motion which serves as a springboard for more sophisticated methods. It is possible to follow two directions: the most common is to allow correlations/memory effects. This can be done in plenty of ways; nonetheless, we introduce the approach based on scaling properties. The second possibility is to admit distributions with infinite variance. Models with these distributions can be for long times, i.e., many independent increments, described via the class of *Lévy processes*. The is a consequence of *Generalized central limit theorem*. We introduce both previous concepts and briefly show some differences. Finally, we combine both concepts in the

model of anomalous *double-fractional* diffusion. Before we turn our attention to the particular models, we have introduce a mathematical apparatus which will be used in generalizations of diffusion equations. It is based on definition of derivative operators for non-natural values. Because there are several existing definitions, we choose a few of them and compare their properties.

3.2 Fractional Calculus

In order to describe the models that generalize the diffusion process to anomalous regime, it is necessary to generalize the classical calculus to operators (integrals and derivatives) of non-natural orders. These operators were studies for quite a long time, for example the "half-derivative" was studied by Leibnitz. The first systematic attempts were done by Liouville and Riemann in the first half of nineteenth century. Presently, the exhaustive overview of fractional calculus is given e.g., by Ref. [52].

We begin with the definition of fractional integral. Let us remind the well-known Cauchy formula for repeated integration:

$$\int_{x_0}^{x}\int_{x_0}^{x_1}\dots\int_{x_0}^{x_{n-1}} f(x_n)\mathrm{d}x_n\dots\mathrm{d}x_1 = \frac{1}{(n-1)!}\int_{x_0}^{x}(x-y)^{n-1}f(y)\mathrm{d}y\,. \tag{3.5}$$

Indeed, it is possible to use the similar expression for the integrals with upper bound. The Cauchy formula can be naturally generalized for fractional orders

$${}_{x_0}\mathfrak{I}_x^{\nu}f(x) := \frac{1}{\Gamma(\nu)}\int_{x_0}^{x}(x-y)^{\nu-1}f(y)\mathrm{d}y\,. \tag{3.6}$$

It is apparent that the fractional integral is a linear operator. The fractional integrals form a semigroup, because

$${}_{x_0}\mathfrak{I}_x^{\nu_1}\circ{}_{x_0}\mathfrak{I}_x^{\nu_2} = {}_{x_0}\mathfrak{I}_x^{\nu_1+\nu_2}\,. \tag{3.7}$$

The baseline for definition fractional derivative is the relation between ordinary derivative and fractional integral

$$\frac{\mathrm{d}}{\mathrm{d}x}\left({}_{x_0}\mathfrak{I}_x^{\nu+1}\right) = {}_{x_0}\mathcal{I}_x^{\nu}\,. \tag{3.8}$$

We want to generalize the relation also for negative values of ν. Nevertheless, the generalization is not unique and there are several possible ways which are not equal. We introduce a few types of fractional derivatives in the following sections, discuss their main properties.

3.2.1 Riemann-Liouville Derivative

The relation between ordinary derivative and fractional integral is the motivation for introduction of *Riemann-Liouville* fractional derivative as a derivative of fractional integral with exponent $\nu_0 \in (0,1)$. Similarly for $\nu_n \in (n, n+1)$, we use the derivative of order $n+1 = \lceil \nu_n \rceil$, where $\lceil n \rceil$ is the ceiling function, i.e. the smallest integer exceeding ν_n. For arbitrary ν, the definition of derivative is given as follows:

$$ {}_{x_0}\mathcal{D}_x^{\nu} f(x) := \frac{\mathrm{d}^{\lceil \nu \rceil}}{\mathrm{d}^{\lceil \nu \rceil} x} \left({}_{x_0}\mathfrak{I}_x^{\lceil \nu \rceil - \nu}[f] \right)(x) \,, \tag{3.9} $$

Unfortunately, the Riemann-Liouville fractional derivative does not follow all properties of ordinary derivatives. For example, the derivative is not commutative

$$ {}_{x_0}\mathcal{D}_x^{\nu_1} \circ {}_{x_0}\mathcal{D}_x^{\nu_2} \neq {}_{x_0}\mathcal{D}_x^{\nu_2} \circ {}_{x_0}\mathcal{D}_x^{\nu_1} . \tag{3.10} $$

which means that the derivative operator does not form the semigroup. On the other hand, for particular values of x_0 is possible to recover some of the properties of ordinary derivatives. For $x_0 = 0$, we recover the derivative of polynomial function, because

$$ {}_0\mathcal{D}_x^{\nu} x^{\mu} = \frac{\Gamma(\mu+1)}{\Gamma(\mu-\nu+1)} x^{\mu-\nu}. \tag{3.11} $$

This is not only true for $\mu > 0$ but also for any real value. Paradoxically, the derivative of a constant is not zero:

$$ {}_0\mathcal{D}_x^{\nu} 1 = \frac{x^{-\nu}}{\Gamma(1-\nu)} \,. \tag{3.12} $$

The expression becomes zero only for natural values of ν, because the Gamma function has poles for $\nu \in \mathbb{N}$. we omit the subindex in the rest of section and assume only the case when $x_0 = 0$.

Another paradox is connected with fractional diffusion equations. In the Laplace image, it is apparent that using Riemann-Liouville derivative demands to impose so-called *fractional initial conditions*, i.e. values of fractional derivative in the initial point, because:

$$ \mathcal{L}\left[\mathcal{D}_x^{\nu} f(x); s\right] \equiv \widehat{[\mathcal{D}_x^{\nu} f]}(s) = s^{\alpha} F(s) - \sum_{k=0}^{\lfloor \nu \rfloor} s^k \left[\mathcal{D}_x^{\nu-k-1} f(x)\right]_{x=0} \,. \tag{3.13} $$

One has to note that these initial conditions do not have any clear physical meaning, as position and velocity in the case of ordinary derivatives [53]. The previous issues motivate the introduction of some other definitions that would overcome these problems.

3.2.2 Caputo Derivative

Due to the objectionable properties of Riemann-Liouville derivative which limits the applicability especially for physical systems, it is necessary to form another definition of fractional derivative which would recover some properties of ordinary derivatives. The main idea is to interchange ordinary derivative operator and fractional integral in the definition of Riemann-Liouville derivative. The resulting derivative is called *Caputo derivative* and is defined as

$$ {}_{x_0}{}^{*}\mathcal{D}^{\nu}_{x} f(x) := {}_{x_0}\mathfrak{I}^{\lceil\nu\rceil-\nu}_{x} \left(\frac{\mathrm{d}^{\lceil\nu\rceil} f(x)}{\mathrm{d}^{\lceil\nu\rceil} x} \right) = \frac{1}{\Gamma(\lceil\nu\rceil-\nu)} \int_{x_0}^{x} \frac{f^{\lceil\nu\rceil}(y)}{(x-y)^{\nu+1-\lceil\nu\rceil}} \,\mathrm{d}y \,. \quad (3.14) $$

Again, unless specified differently, we assume $x_0 = 0$. The Caputo derivative is more restrictive on its domain, because the function f has to have alt least $\lceil\nu\rceil$ derivatives. On the other hand, because the derivative is inside the integral, the derivative of constant function is now zero ${}^{*}\mathcal{D}^{\nu}_{x} 1 = 0$. Laplace transform of Caputo derivative is

$$ \mathcal{L}\left[{}^{*}\mathcal{D}^{\nu}_{x} f(x); s\right] \equiv \widehat{\left[{}^{*}\mathcal{D}^{\nu}_{x} f\right]}(s) = s^{\nu} F(s) - \sum_{k=0}^{\lfloor\nu\rfloor} s^{\nu-k-1} f^{(k)}(0) \quad (3.15) $$

so the natural initial conditions are recovered. Caputo differential operators and fractional differential equations of Caputo type have been studied e.g. in Ref. [54]. The eigenfunctions of Caputo derivative operators

$$ {}^{*}\mathcal{D}^{\nu}_{x} f(x) = \lambda f(x) \quad (3.16) $$

are expressible in terms of *Mittag-Leffler functions* (defined in Appendix C)

$$ f_{\lambda}(x) = E_{\nu}(\lambda x^{\nu}) \,. \quad (3.17) $$

Finally, Riemann-Lioville derivative and Caputo derivative can be connected through the relation (the proof can be found in Ref. [55])

$$ {}^{*}_{x_0}\mathcal{D}^{\nu}_{x} f(x) = {}_{x_0}\mathcal{D}^{\nu}_{x} f(x) - \sum_{k=0}^{\lfloor\nu\rfloor} \frac{x^{k}}{k!} f^{(k)}(x_0) \,. \quad (3.18) $$

In the next section we show yet another definition of the fractional derivative operator, which is the most common in physical applications.

3.2.3 Riesz-Feller Derivative

Previous definitions of derivatives depend on the particular value of the lower bound of the integral, which influence the necessary initial conditions. In many cases, as e.g.

in probability theory, we want to set the normalization conditions rather than particular function values. This can be reached, when we send the point x_0 to minus infinity, so

$$\mathfrak{D}^{\nu}_{x} f(x) := \lim_{x_0 \to -\infty} {}_{x_0}\mathcal{D}^{\nu}_{x} f(x)\,. \tag{3.19}$$

The derivative operator is called *Riesz-Feller derivative*. Because of the Eq. (3.18), the Riesz-Feller derivative operator can be alternatively defined by the Caputo derivative. It is clear, that because of the convergence of the integral, the first $\lceil \nu \rceil$ derivatives has to vanish in minus infinity. Thus, the domain of such functions is much smaller than in case of Riemann-Liouville or Caputo derivative.

Riesz-Feller derivative possesses several important properties. First, eigenfunctions of Riesz-Feller derivative are exponentials function, similarly to ordinary derivatives

$$\mathfrak{D}^{\nu}_{x} \exp(\lambda x) = \lambda^{\nu} \exp(\lambda x)\,. \tag{3.20}$$

Second, the Riesz-Feller derivative naturally generalizes derivative operator in the Fourier space, because its Fourier transform is equal to

$$\mathcal{F}\left[\mathfrak{D}^{\nu} f(x); p\right] \equiv \overline{[\mathfrak{D}^{\nu} f]}(p) = \int_{\mathbb{R}} \mathrm{d}x\, e^{ipx} \int_{-\infty}^{x} \mathrm{d}y\, (x-y)^{-\nu-1} f(y) = (-ip)^{\nu} \overline{f}(p). \tag{3.21}$$

This is shown in Ref. [53]. Particularly important are Riesz-Feller derivatives in connection with Lévy processes, because they belong to the wider class of pseudo-differential operators defined through the Fourier transform. Definition of these processes, also with help of fractional calculus, is the subject of the next section.

3.3 Anomalous Diffusion

In the first section of this chapter was presented the description of regular diffusion process. Nonetheless, as objected before, in the case of complex processes as processes with long-term correlations, memory effects or sudden jumps, it is necessary to use more appropriate diffusion models that are capable to describe the aforementioned phenomena. In the rest of the chapter are introduced some examples of these processes. In is not the aim of this chapter to describe every single existing generalization of diffusion processes (which is anyway not possible due to the enormous number of existing processes), but to show some possible directions and concepts used in the theory of generalized diffusion. We start with the *fractional Brownian motion*, model with long-range correlations and Hurst exponent not equal to $\frac{1}{2}$. Then, we move to the class of *Lévy processes*, based on stable distributions with power-laws. Finally, we generalize the Lévy processes to anomalous *double-fractional diffusion*, is a straight generalization of Lévy process and combines the elements of both previous models in a way. These models are not only important in physics, but also play a crucial role in biology, sociology and economics.

3.3.1 Fractional Brownian Motion

One possible direction in diffusion process generalization is to introduce correlations into the system. We have shown that for Brownian motion are the increments statistically independent. When we put correlations into the system, it also affects the scaling properties. Positivite/negative correlations cause the persistent/antipersistent behavior which is reflected in the different scaling exponents. As a result, the fractal dimension of the process changes as discussed in Sect. 2.1. From the mathematical point of view, the process is defined as a fractional integral of the stochastic Wiener measure, originally introduced by Mandelbrot [8]

$$W_H(t) := \mathfrak{I}^{H-\frac{1}{2}}(\mathrm{d}W(t)) = \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-s)^{H-\frac{1}{2}}\mathrm{d}W(s)\,. \tag{3.22}$$

The Hurst exponent is defined in the interval $H \in [0,1]$. More details about definition of stochastic measures and the stochastic calculus can be found in Refs [56].

The definition leads to non-trivial correlations in increments. When we calculate the autocorrelation function of the process, we obtain

$$\langle W_H(t)W_H(s)\rangle = \frac{1}{2H\Gamma(H+\frac{1}{2})^2}\left(s^{2H}+t^{2H}-|s-t|^{2H}\right). \tag{3.23}$$

According to the parameter H, which corresponds to the Hurst exponent, we can divide the fractional Brownian motion into three classes:

- for $H \in [0,\frac{1}{2})$ has the process negative correlations and anti-persistent, sub-diffusive behavior, which causes larger fractal dimensions
- for $H = \frac{1}{2}$ we recover the Brownian motion with uncorrelated increments
- for $H \in (\frac{1}{2},1]$ is the process positively correlated, super-diffusive with presence of more trends than in case of uncorrelated Brownian motion.

These processes based on fractional Brownian motion are observed in finance, biology, dynamical systems and in many other fields. The fractional Brownian motion is only one simple example of processes with long-term memory. There exist a broad literature about stochastic processes and applications to physics, e.g. [57].

Alternatively, one can assume processes with uncorrelated increments, but with limiting distributions with infinite variance. This opens another whole class of processes with heavy tails driven by so-called Lévy distributions. These processes are described in the next section.

3.3.2 Lévy Flights

Lévy flights are processes with uncorrelated increments based on stable distributions. Lévy distributions constitute a class of distributions obtained as the limiting distributions of sums of i.i.d. random variables. This statement is a consequence of the *Generalized central limit theorem*. Gnedenko and Kolmogorov [12] have shown that these distributions corresponds to the class of distributions which are functionally invariant under convolution. This is not surprising, because the probability distribution of sum of two independent random variables is given by their convolution. Unfortunately, the probability density function is not expressible in most cases. It is necessary to use the Fourier representation, i.e., the characteristic function. According to the analogy with physics, the logarithm of the characteristic function is called *stable Hamiltonian*. The properties of stable distributions are summarized in Appendix A. Here we only mention the most important aspects necessary to definition of Lévy flights. At first, the stable Hamiltonian is expressible as

$$\mathcal{H}_{\alpha,\beta;\bar{x},\bar{\sigma}}(p) \equiv \ln\langle e^{ipx}\rangle = i\bar{x}p - \bar{\sigma}^{\alpha}|p|^{\alpha}\left(1 - i\beta\mathrm{sign}(p)\omega(p,\alpha)\right) , \tag{3.24}$$

where the exact form of function ω is shown in Appendix A. The four parameters of the distribution have the following meaning: $\alpha \in (0, 2]$ is the stability parameter, which influences the shape of the distributions, the decay of tails parts and existence of fractional moments $\langle x^{\mu}\rangle$. Parameter $\beta \in [-1, 1]$ is the asymmetry parameter, for $\beta = 0$ we obtain a symmetric distribution around its mean value (or location parameter), for $\beta = \pm 1$ we have totally asymmetric distribution. This means that for $\alpha \in (1, 2)$ one tail decay exponential and the other tail decays polynomially (this shows Eq. A.12 in Appendix). The parameters $\bar{x} \in \mathbb{R}$ and $\bar{\sigma} \in \mathbb{R}_+$ are location and scale parameters and are equal to mean and variance, whenever these moments exist and are finite. In Appendix A is also presented an alternative representation of stable Hamiltonian, which is sometimes more advantageous to use.

There is a tight relation between Lévy distributions and Riesz-Feller fractional derivatives [58]. In Sect. 3.2.3 is shown that the representation of Riesz-Feller derivative in Fourier image is equal to multiplication by term $(\pm ip)^{\nu}$, which is exactly the stable Hamiltonian of totally asymmetric Lévy distribution. This allows to define the class of pseudo-differential operators [53] defined in the Fourier image as

$$\overline{[{}^{\beta}D_x^{\nu} f]}(p) = \mathcal{H}_{\nu,\beta}(p)\overline{f}(p) , \tag{3.25}$$

which is for $\beta = \pm 1$ equal to the Riesz-Feller fractional derivative (for $\beta = 1$ we have the fractional derivative with integration from x to $+\infty$.) Consequently, the solution of generalized diffusion equation

$$\frac{\partial}{\partial t} f(x,t) = {}^{\beta}D_x^{\nu} f(x,t) \tag{3.26}$$

is the Lévy distribution with stable Hamiltonian $\mathcal{H}_{\nu,\beta}(p)$.

Regarding the scaling properties of Lévy processes, it has been already discussed that sample paths of Lévy processes are equal to $\max\{1,\alpha\}$ and for process in x-t space is $\max\{1, 2-\frac{1}{\alpha}\}$, which is smaller than $\frac{3}{2}$. This is not surprising, because presence of polynomial tails in the distribution cause large jumps in the process and these trends cause the decrease of fractal dimension. This is in contrast to fractional Brownian motion, because fractional Brownian motion can also acquire fractal dimension larger than $\frac{3}{2}$. In the case of anti-persistent behavior is necessary to introduce correlations to the system.

Usually, the real systems are not described exactly by Lévy processes, but they can be used as limiting process, especially for large timescales. Thus, it is convenient to introduce other more realistic models valid also for short timescales. We introduce a concept of double-fractional diffusion, where we use not only the spatial fractional derivative, but also a temporal derivative operator. This gives us wider class of diffusion processes which possess more realistic behavior.

3.3.3 Double-Fractional Diffusion

Double-fractional diffusion is a model based on diffusion equation with fractional derivatives in both spatial and temporal coordinates. The Green function (also called fundamental solution) is therefore the solution of equation

$$\left({}^{K}\partial_{\tau}^{\gamma} + \mu[{}^{\beta}D_{x}^{\alpha}]\right) g(x,t) = 0\,, \tag{3.27}$$

where γ is the degree of temporal derivative called *speed diffusion parameter*, α is the degree of Riesz-Feller spatial fractional derivative called *stable parameter*, μ is the diffusion parameter (for $\alpha = 2$ is proportional to parameter D) and K denotes the type of temporal fractional derivative. We consider two types of temporal derivatives, namely Riesz-Feller and Caputo derivative. Both diffusion equations belong to the wide class of pseudo-differential operators which can be expressed in the Laplace-Fourier image (i.e. Laplace transform $t \to s$ in temporal coordinate and Fourier transform $x \to p$ in spatial coordinate) as

$$a(s)\hat{\bar{g}}(p,s) - a_0(s)\bar{g}_0(p) \;=\; b(p)\hat{\bar{g}}(p,s) \tag{3.28}$$

where $a(s)$ is the Laplace representation of the temporal fractional derivative $a_\gamma(s) = s^\gamma$; $b(p)$ is the Fourier representation of spatial fractional derivative. It is for Riesz-Feller derivative equal to $b_\alpha(p) = \mathcal{H}_{\alpha,\beta}(p)$. $\bar{g}_0(p)$ is the Fourier transform of the first initial condition, which is usually equal to $\bar{g}_0(p) = \mathcal{F}[\delta(x)] \equiv 1$. Finally, $a_0(s)$ is the term depending on the type of derivative. It is expressible as $a_0(s) = s^{\gamma-\kappa}$, where $\kappa = 1$ for Caputo derivative and $\kappa = \gamma$ for Riesz-Feller derivative. We have to mention that for $1 < \gamma \leq 2$ is necessary to impose another initial condition, which adds another term to

Eq. (3.28). In order to preserve the above presented form of double fractional equation also in this case, we have to assume that the second initial condition has the following form:

$$\frac{\partial g}{\partial t}(x,t)|_{t=0} \equiv 0\,. \tag{3.29}$$

Nevertheless, this type of initial condition is quite natural, so it is reasonable to consider only this type of diffusion . The important question is, whether the solution of this class of double-fractional diffusion equations is positive so that it is interpretable as a Green function. In Ref. [59] is possible to find that this is possible if the two parameters fulfill the condition

$$0 < \gamma < \alpha \leq 2\,. \tag{3.30}$$

We turn our attention to a kernel representation of the fundamental solution, which is useful for $\gamma < 1$, because in this case is the distribution a continuous superposition of Lévy distributions. Sometimes it can also be called "superstatistical" representation because of similarity to superstatistics. We have to note, that this representation can be only formal and the real superstatistics can be observed only in case when we can recognize two distinct characteristic time scales [60].

The kernel representation can be obtained from the Laplace-Fourier image, because according to Eq. (3.28), the Green function can be represented with help of Schwinger formula ($\frac{1}{A} = \int_0^\infty e^{-A}$ for $\Re(A) > 0$) as

$$\begin{aligned}\hat{\bar{g}}(p,s) = \frac{a_0(s)\bar{g}_0(p)}{a(s)-b(p)} &= \int_0^\infty \mathrm{d}l\, \left[a_0(s)e^{-la_\gamma(s)}\right]\left[\bar{g}_0(p)e^{lb_\alpha(p)}\right] \\ &= \int_0^\infty \mathrm{d}l\, \hat{g}_1(s,l)\bar{g}_2(l,p)\,. \end{aligned} \tag{3.31}$$

The solution is given by superpositions of Lévy stable distribution with stable parameter α at different times, weighted by *smearing kernel* $g_1(s,l)$. The double-fractional diffusion is decomposed into set of two fractional equations for two kernels

$$\frac{\mathrm{d}}{\mathrm{d}l}\,\hat{g}_1(s,l) = -a(s)\hat{g}_1(s,l), \qquad \hat{g}_1(s,0) = a_0(s) \tag{3.32}$$

$$\frac{\mathrm{d}}{\mathrm{d}l}\,\bar{g}_2(l,p) = b(p)\bar{g}_2(l,p), \qquad \bar{g}_2(0,p) = \bar{g}_0(p) \tag{3.33}$$

and the connection to the resulting Green function is given by the Eq. (3.31). Now, we discuss two kinds of considered fractional derivatives. When we take into account the Riesz-Feller derivative, i.e. $\kappa = \gamma$, we obtain the fractional equation exactly same as in case of Lévy stable process, but only with stable parameter γ and asymmetry parameter $+1$. For $\gamma < 1$ is the support of such distribution bounded to the positive half-line. The normalization of the Green function requires normalization of the smearing kernel [61], so we have

$$\int_0^\infty \mathrm{d}l\, g_1^{RF}(t,l) = \int_0^\infty \mathrm{d}l \int_{\mathbb{R}} \mathrm{d}p\, \frac{e^{-ipt}}{2\pi} e^{-l\mu(ip)^\gamma} = \int_{\mathbb{R}} \mathrm{d}p \frac{e^{-ipt}}{2\pi} \frac{1}{\mu(ip)^\gamma} = \frac{t^{\gamma-1}}{\mu\Gamma(\gamma)}. \tag{3.34}$$

Thus, the smearing kernel is for the Riesz-Feller derivative equal to

$$g_1^{RF}(t,l) = \left(\frac{\Gamma(\gamma)}{t^{\gamma-1}}\right) \frac{1}{l^{1/\gamma}} L_{\gamma,1}\left(\frac{t}{l^{1/\gamma}}\right) . \tag{3.35}$$

In the case of Caputo derivative is the solution slightly different. According to Ref. [59], the solution is expressible via Wright M-function

$$g_1^{C}(t,l) = \frac{1}{t^\gamma} M_\gamma\left(\frac{l}{t^\gamma}\right) , \tag{3.36}$$

where Wright M-function can be defined as an infinite series:

$$M_\nu(z) = \sum_{n=0}^{\infty} \frac{(-z)^n}{n!\Gamma(-\nu n + (1-\nu))} . \tag{3.37}$$

The M-function has a tight relation to Lévy distribution, because

$$\frac{1}{c^{1/\nu}} L_{\nu,1}\left(\frac{x}{c^{1/\nu}}\right) = \frac{c\nu}{x^{\nu+1}} M_\nu\left(\frac{c}{x^\nu}\right) \tag{3.38}$$

for $\nu \in (0,1)$, $c > 0$ and $x > 0$. Altogether, the smearing kernel for Caputo derivative can be represented also through Lévy stable distribution with slightly different coefficients

$$g_1^{C}(t,l) = \left(\frac{t}{l\gamma}\right) \frac{1}{l^{1/\gamma}} L_{\gamma,1}\left(\frac{t}{l^{1/\gamma}}\right) . \tag{3.39}$$

In Appendix E are compared the properties of Riesz-Feller and Caputo smearing kernels. These two kernels are depicted in Fig. 3.1. It is necessary to note that to the main differences belongs different behavior for $l = 0$. Riesz-Feller kernel goes to zero while the Caputo kernel does not vanish. This difference also influences the possible applications to the real systems. When the dependence on the initial configuration of the system remains strong also in later times, we use the Caputo derivative, on the other hand, if the most contributing parts are the pseudotimes $l \approx t$ we use Riesz-Feller derivative.

For practical calculations as well as for theoretical description is convenient to use another representation of double-fractional Green function based on *Mellin-Barnes* integral representation. Eq. (3.28) has for Double-fractional diffusion the following form

$$\hat{\tilde{g}}(p,s) = \frac{s^{\gamma-\kappa}}{s^\gamma - \mathcal{H}_{\alpha,\beta}(p)}. \tag{3.40}$$

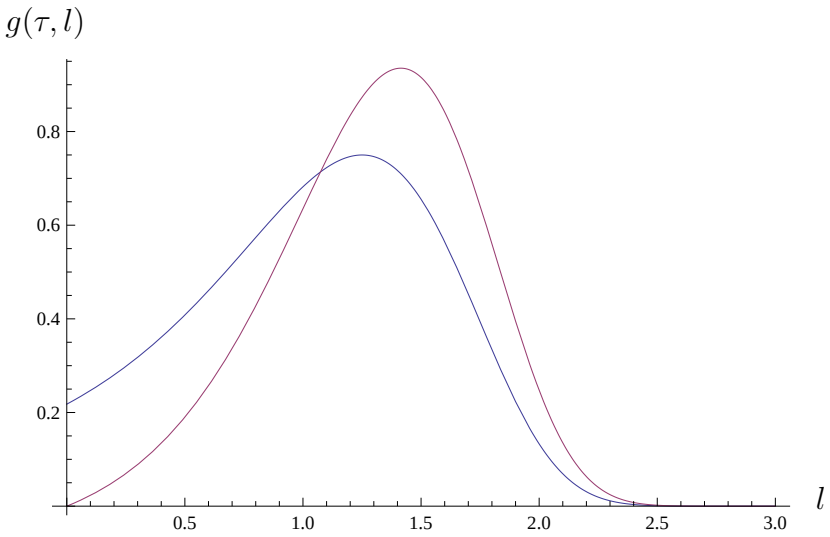


Figure 3.1: Comparison Riesz-Feller kernel $g^{RF}(\tau, l)$ (purple line) and Caputo kernel $g^{C}(\tau, l)$ (blue line) for $\tau = 1$.

From symmetry reasons, we can take into account only solutions for positive values of x, because the negative part can be obtained from relation $g_{\alpha,\beta}(x,t) = g_{\alpha,-\beta}(-x,t)$, and therefore we leave the asymmetry parameter β formally undetermined. In Appendix C is shown that the inverse Laplace transform of Eq. (3.40) is expressible as the Mittag-Leffler function, so

$$\hat{g}(p,t) = t^{\kappa-1} E_{\gamma,\kappa}\left(\mathcal{H}_{\alpha,\beta}(p) t^{\gamma}\right) . \tag{3.41}$$

It is advantageous to represent the Mittag-Leffler function through an integral form called Mellin-Barnes representation. It is based on the Mellin transform introduced in Appendix B. According to Eq. (C.7), it is possible to rewrite Eq. (3.41) as

$$\hat{g}(p,t) = \frac{t^{\kappa-1}}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\Gamma(s')\Gamma(1-s')}{\Gamma(\kappa-\gamma s')} \left[-\mu|p|^{\alpha} exp\left(-\frac{i\pi\theta\,\mathrm{sign}(p)}{2}\right)\tau^{t}\right]^{-s'} \mathrm{d}s' \tag{3.42}$$

where $0 < \Re(c) < 1$ and $\theta = 2-\alpha$ for $\beta = -1$ and $\alpha > 1$; resp. $\theta = \alpha - 2$ for $\beta = +1$ and $\alpha > 1$. Note that the parameter θ is known from an alternative representation of stable Hamiltonian introduced in Appendix A. Inverse transform is straightforward, because

$$g(x,t) = \mathcal{N}\frac{t^{\kappa-1}}{2\pi i x} \int_{c-i\infty}^{c+i\infty} \frac{\Gamma(s')\Gamma(1-s')}{\Gamma(\kappa-\gamma s')\Gamma(s'\alpha)} \left[-\mu\frac{t^{\gamma}}{x^{\alpha}}\right]^{-s'} \mathrm{d}s' \tag{3.43}$$

where $\mathcal{N} = \frac{\tau^{\kappa-1}}{\Gamma(\kappa)}$ is a normalization constant. Finally, we apply a change of variables

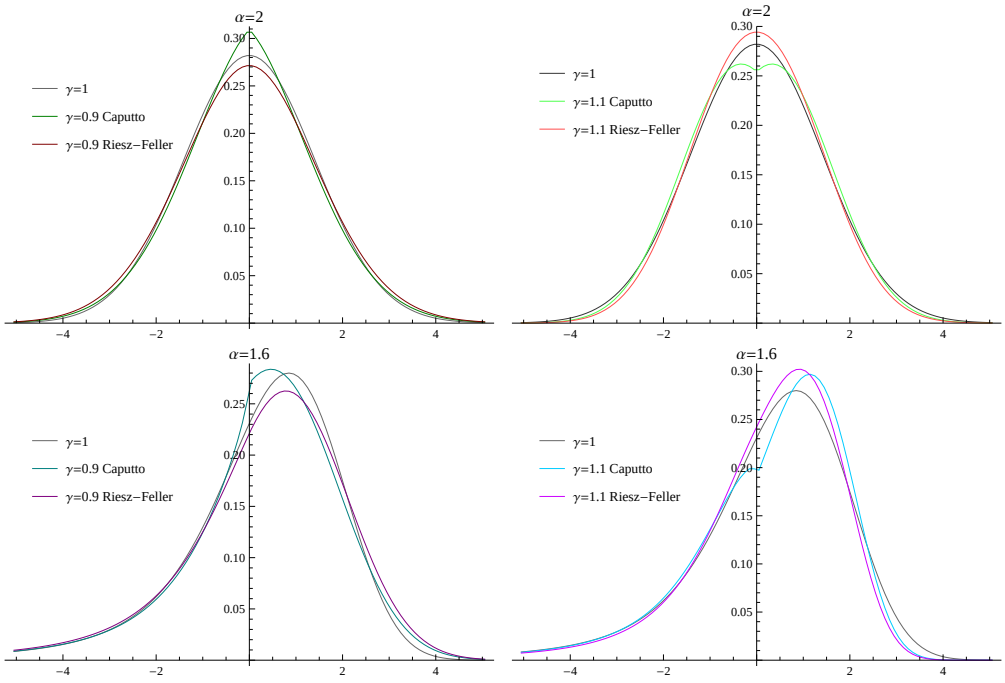


Figure 3.2: Comparison of Green functions for ordinary derivative ($\gamma = 1$), Riesz and Caputo derivative for $\gamma = 0.9$ (slow diffusion) and $\gamma = 1.1$ (fast diffusion) for $\alpha = 2$ and $\alpha = 1.6$. The Caputo Green function highlights the peak of the distribution, while Riesz-Feller Green function has slower decay in tails of the distribution. Note that for $\gamma > 1$, the green function exhibits fast-diffusion behavior with two peaks receding in time.

$\alpha s' = s$ and we end with

$$g^{DF}(x,t) = \frac{\Gamma(\kappa)}{2\alpha\pi i|x|}\int_{c-i\infty}^{c+i\infty} \frac{\Gamma\left(\frac{s}{\alpha}\right)\Gamma\left(1-\frac{s}{\alpha}\right)\Gamma(1-s)}{\Gamma\left(\kappa-\frac{\gamma}{\alpha}s\right)\Gamma\left(\frac{(\alpha-\theta)s}{2\alpha}\right)\Gamma\left(1-\frac{(\alpha-\theta)s}{2\alpha}\right)}\left[\frac{x}{(-\mu t^{\gamma})^{1/\alpha}}\right]^{s}\mathrm{d}s. \tag{3.44}$$

We see that the Green function follows the scaling rule $g^{DF}(x,t) \propto t^{\Omega}$, where $\Omega = \frac{\gamma}{\alpha}$ is called *diffusion scaling exponent* and plays the similar role as Hurst exponent. The main advantage of the Mellin-Barnes representation is the fast convergence of the complex integral, which allows to calculate the values of Green function much faster than in other representations. In Fig. 3.2 are displayed Green functions for several parameters. We can also compare differences between Riesz-Feller and Caputo derivatives.

Chapter 4

Generalized Entropies and Applications in Thermodynamics

Role of entropy in mathematics and physics is extremely important, since it is a cornerstone for the whole statistical physics and many other disciplines. This chapter describes origins of the concept of entropy and presents several possible generalizations that are able to describe nonextensive systems, open systems or systems with long-range correlations. We discuss the properties of the entropy for all presented generalizations and derive corresponding MaxEnt distributions.

4.1 Role of Entropy in Physics and Mathematics

The motivation for using entropy is coming from several scientific branches. Especially its role in in physics, statistics and other fields is extremely important. This section summarizes the main arguments for introduction of entropy and discusses its main principles. We start with the classical theory of thermodynamics and information theory. The original works of Clausius, Boltzmann and Gibbs defined the classic role of entropy in the theory of thermodynamics, including several formulations of second law of thermodynamics. Probably the most popular definition, carved on the gravestone of Ludwig Boltzman, introduces the entropy of a microcanonical ensemble as

$$S = k_B \ln W \,, \tag{4.1}$$

where k_B is the Boltzman constant and W is the number of states. The other important definition came from the information theory where the term entropy is defined as a measure of ignorance. In other words, it is the amount of information which is not not known about the system. As introduced by Shannon in his paper [62] and followed by Feinstein [63], the entropy can be interpreted as the minimal amount of information needed

in order to fully determine the system. Statistical physics establishes the relation between Boltzmann thermodynamical entropy and Shannon informational entropy - they are (up to the multiplicative factor) the same. The entropy is also crucial in a general procedure proposed by Jaynes [64] which is used for calculation of the most descriptive distribution of a system. The procedure is called *Maximum entropy principle* (MaxEnt) and determines the most probable probability distribution as a distribution which maximizes the entropy under given constraints. In other words, the MaxEnt probability distribution contains only information included in the set of constraints. The important point in the theoretical description of entropy was given by Khinchin [65] who introduced an axiomatic definition of entropy. We dedicate the next section to the axiomatic definition of the entropy, because it serves as a springboard for various generalizations.

The concept of entropy one of the most important tools not only for physicists but also for many other scientists. In information theory represents the concept of divergences (and derived information measures and entropies) an important way how to measure distances between probability distributions a the amount of information encoded in the probability distribution. Moreover, disciplines as statistics, numerical mathematics or theory of partial differential equations have adopted entropy as one of the successful methods for solution of various problems. Last but not least, applied sciences which are using some mathematical or physical methods for modeling and analysis use entropy in modeling as well. Among others, let us mention biology, sociology, theory of networks, econometrics or applications in finance. For all previously mentioned fields is important to find some appropriate techniques and models that would be able to describe the complex behavior appearing in the systems. Thus, we introduce several generalizations of classic Shannon(-Boltzmann-Gibbs) entropy in the following sections to be able to deal with systems which are not isolated or are not in equilibrium.

4.1.1 Axiomatic Definition of Shannon Entropy

It is possible to define entropy in several ways. We follow the approach of A. Khinchin [65], who expressed Shannon entropy $\mathcal{H}(\mathcal{P})$ uniquely by four axioms. Let us denote the discrete arbitrary probability distribution as $\mathcal{P} = (p_1, \ldots, p_n)$. The four axioms are formulated in the following way:

1. *Continuity axiom*: for given n and probability distribution $\mathcal{P}$ is $\mathcal{H}(\mathcal{P})$ a continuous function with respect to all its arguments.

2. *Maximality axiom*: for given n takes $\mathcal{H}(\mathcal{P})$ the largest value for uniform distribution, i.e. $\mathcal{P}_n = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$.

3. *Expansibility axiom*: $\mathcal{H}(p_1, \ldots, p_n, 0) = \mathcal{H}(p_1, \ldots, p_n)$.

4. *Additivity axiom*: $\mathcal{H}(A \cup B) = \mathcal{H}(A) + \mathcal{H}(B|A)$,
   where $\mathcal{H}(B|A) = \sum_i p_{i,A} \mathcal{H}(B|A = a_i)$ is the conditional entropy
   and $\mathcal{P}_A = (p_{1,A}, \ldots, p_{n,A})$ is the distribution corresponding to experiment A.

In the last axiom, we adopt the abbreviation that $S(A)$ denotes the entropy belonging to probability distribution $\mathcal{P}_A$ of the random variable A. Similarly, $S(A \cup B)$ is the entropy belonging to joint distribution $\mathcal{P}_{A \cup B}$. If A is independent of B, the conditional entropy reduces to $S(B)$.

Alternatively, Shannon [66] and other authors use slightly different set of axioms, which are equivalent to Khinchin's. The four axioms determine uniquely the functional form of entropy (up to normalization constant) which can be expressed as

$$\mathcal{H}(\mathcal{P}) = -\sum_{i=1}^{n} p_i \ln p_i \,. \tag{4.2}$$

It should be mentioned that the Shannon entropy has also the operational definition [63]. The entropy (defined in terms of binary logarithm, i.e. $\sum_i p_i \log_2 p_i$) represents the amount of information (measured in bits) which is necessary to fully determine the system. In other words, it is the minimal number of binary YES/NO question that has to be answered to reduce all uncertainty. Once can also say that it represents the minimal length of binary code that uniquely describes the system. As a consequence, the Shannon entropy is a measurable quantity. In the next section are discussed some of the properties of information measures particularly interesting for applications in thermodynamics.

4.2 Important Properties of Entropies

Shannon(-Boltzmann-Gibbs) entropy is the most important information measure with enormous number of applications. It is the central concept in the theory of classical thermodynamics and statistical physics. Nevertheless, complex systems, systems with long-range interactions or systems far from equilibrium cannot be fully described within the framework of classical thermodynamics. As a consequence, these systems require more sophisticated description based on generalized entropies that go beyond standard thermodynamics. In this section, we discuss the main properties of entropy classes, which are important in description of non-equilibrium systems. Among the other properties, we discuss additivity, extensivity and Legendre structure of thermodynamics resulting from the MaxEnt procedure. Finally, we present some properties sufficient for validity of maximality axiom. Most of the properties are discussed in general case. Only if necessary, we restrict the discussion to some more specific classes of entropies. As an example, in some cases is advantageous to work with the class of *trace entropies* (used

e.g., in Ref. [67]) which can be defined in a simple form

$$S_g(\mathcal{P}) = \sum_i g(p_i)\,. \tag{4.3}$$

This class covers many important classes of entropies, including Shannon and Tsallis entropy. It has some nice properties. For example, the concavity of the entropy functional is equal to concavity of function g, because the Hessian matrix (matrix of second derivatives) has the diagonal form

$$\mathbb{H}(S_g) = \mathrm{diag}\left(\frac{\mathrm{d}^2 g(p_1)}{\mathrm{d}p_1^2}, \ldots, \frac{\mathrm{d}^2 g(p_n)}{\mathrm{d}p_n^2}\right)\,. \tag{4.4}$$

On the other hand, not all entropies belong to the class of trace entropies. Still, some of them are expressible as *generalized trace entropies*, i.e. in the form

$$S_{G,g}(\mathcal{P}) = G\left(\sum_i g(p_i)\right)\,. \tag{4.5}$$

For instance, Rényi entropy belongs to the class of generalized trace entropies.

4.2.1 Additivity versus Extensivity

Additivity and extensivity are widely discussed properties of all entropies, but there exist some misconceptions about these two terms. One should clearly distinguish between them and discuss their relation [4]. First, we start with the term *additivity*, which is connected more with the informational origin of entropy. In Khinchin axiomatic, the additivity of the entropy means that

$$S(A \cup B) = S(A) + S(B|A) = S(B) + S(A|B) \tag{4.6}$$

where $S(B|A)$ is the conditional entropy. For independent events, the entropy is simply the sum of entropies of particular subsystems. Additivity is the major property of Shannon entropy and it is also valid for the Rényi entropy. Generally, the consequence of additivity is that the conditional entropy is defined in the usual way

$$S(B|A) = S(A \cup B) - S(A)\,. \tag{4.7}$$

For other entropies is the formula not valid. We define for many cases a generalized form of additivity. Tempesta [68] and other authors introduce for this end a term *composability*, which means that the entropy of a composed system is expressible in terms of entropies of its subsystems, so it is possible to write

$$S(A \cup B) = \Phi(S(A), S(B))\,. \tag{4.8}$$

As an example, in Refs. [68, 69] are discussed properties of such general classes of entropies. One particular case of the generalized additivity law represents Tsallis entropy. The generalized additivity is for Tsallis entropy defined as (see Sect. 4.3.2)

$$\mathcal{S}_q(A \cup B) = \mathcal{S}_q(A) \oplus_q \mathcal{S}_q(B) = \mathcal{S}_q(A) + \mathcal{S}_q(B) + (1-q)\mathcal{S}_q(A)\mathcal{S}_q(B)\,. \tag{4.9}$$

The addition law can be described with help of so-called q-sum, q-deformation of addition, which is defined as

$$\Phi_q(x, y) = x \oplus_q y = x + y + (1-q)xy\,. \tag{4.10}$$

Function Φ is nothing else than a group operation. As a consequence, the entropies can be classified with respect to its generalized additivity law [67].

Extensivity is, on the other hand, a property which is connected with the thermodynamical properties of the system. Let us have a compound system $A = \bigcup_{i=1}^{N} A_i$ of not necessarily independent variables. Let us denote a state space of N variables as $W(N)$. If the maximality axiom holds, then the entropy becomes maximal for the uniform distribution $\{1/W(N), \ldots, 1/W(N)\}$. We say that the entropy is extensive if

$$\lim_{N\to\infty} \frac{S(W(N))}{N} = \omega \tag{4.11}$$

where $\omega \in (0, \infty)$. That means that the entropy scales for large systems (i.e. systems with $N \gg 1$) as

$$S(W(N)) \propto N\,. \tag{4.12}$$

This condition ensures that the thermodynamical entropy (in the limit for large N) is an extensive function of its variables, i.e. $S(\alpha N, \alpha E, \alpha V) = \alpha S(N, E, V)$. Indeed, contrary to additivity, extensivity is property depending on the actual system, i.e. depending on the state function $W(N)$. When the system is compound of independent variables with no restrictions, then the state space grows exponentially, because it holds that $W(N) = W(N_1)W(N_2)$ for $N = N_1 + N_2$, which determines the state space volume as $W(N) \propto \mu^N$. Hence, Shannon entropy is extensive for such systems, because

$$\mathcal{H}(W(N)) = -\sum_{i=1}^{W(N)} \frac{1}{W(N)} \log \frac{1}{W(N)} = N \log \mu\,. \tag{4.13}$$

If the state space grows not exponentially, but rather polynomially, i.e., as $W(N) \propto N^\rho$, then we should use Tsallis entropy (see, e.g., Ref. [70]), because the Tsallis entropy is for these systems extensive:

$$\mathcal{S}_{1-1/\rho}(W(N)) = \frac{1}{1/\rho} W(N) \cdot W(N)^{1/\rho - 1} - 1 \propto \rho N\,. \tag{4.14}$$

Similarly, if the state space grows subexponentially, i.e. $W(N) \propto \nu^{N^\gamma}$, then so-called δ-entropy $\sum_{i=1}^{W} p_i(\ln p_i)^\delta$ represents an extensive entropy for this system.

Although some authors interchange terms additivity and extensivity, it is important to distinguish them. On the other hand, they are often tightly connected. Of course, there exist nonextensive systems with additive entropy (Shannon entropy for systems with long-range correlations) and non-additive but extensive systems (Tsallis entropy for systems with long-range correlations). Indeed, the most common case is the case of an additive and extensive system described by Shannon(-Boltzmann-Gibbs) entropy which leads to classical thermodynamics. Anyway, for complex systems with long-range correlations, which are nonextensive under Shannon entropy, is advantageous to use non-additive entropies, because many thermodynamical properties remain preserved.

4.2.2 MaxEnt Principle and Legendre Structure

The importance of entropy in statistical physics lies in the fact that the realized distribution maximizes the entropy under given constraints. This principle is called *Maximum entropy principle* (MaxEnt) and was firstly formulated by Jaynes [64]. The essence of the principle consists in the fact that the resulting distribution obtained from the MaxEnt procedure contains only information included in the constraints and does not contain any other additional information. Consequently, the particular entropy determines the form of MaxEnt distribution and the constraints only change the parametric description. This classic procedure is one of the basic techniques in statistical physics. Let us consider a particular form of entropy, for example Shannon entropy. We maximize the entropy with respect to the given constraints. This can be done through the techniques of Lagrange multipliers. Let us restrict ourselves into the most common class of constraints, i.e. $f_i(\mathcal{P}) = 0$, for $i \in \{1, \ldots, m\}$. We define the Lagrange functional as

$$\mathcal{L}(\mathcal{P}, \lambda) = G(\mathcal{P}) - \lambda \cdot f(\mathcal{P}) = -\sum_{j=1}^{n} p_j \ln p_j - \sum_{i=1}^{m} \lambda_i f_i(\mathcal{P}) \tag{4.15}$$

where $G(\mathcal{P})$ is the given entropy functional and λ_i are Lagrange multipliers. The maximization of Lagrange function leads to set of equations:

$$\frac{\partial \mathcal{L}(\mathcal{P}, \lambda)}{\partial p_j} = 0 \quad \text{for } j \in \{1, \ldots, n\}, \tag{4.16}$$

$$\frac{\partial \mathcal{L}(\mathcal{P}, \lambda)}{\partial \lambda_i} = 0 \quad \text{for } i \in \{1, \ldots, m\}. \tag{4.17}$$

The type of constraints determines the resulting MaxEnt distribution. In all cases is necessary to normalize the probability distribution, so we demand

$$\sum_{j=1}^{n} p_j = 1\,. \tag{4.18}$$

When we demand only the normalization condition, we end with the uniform distribution $p_i = \frac{1}{n}$. In thermodynamics, we usually impose the constraint on the average energy of the system, so

$$\sum_{j=1}^{n} p_j E_j = \langle E \rangle\,. \tag{4.19}$$

In the case of Shannon entropy leads the condition to the well-known Boltzmann-Gibbs distribution

$$p_i = \frac{1}{Z}\, e^{-\beta E_i} = \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}} \tag{4.20}$$

where Z is called partition function and β is the Lagrange multiplier belonging to the energy constraint and is connected to the temperature $\beta = \frac{1}{k_B T}$. As a consequence, we obtain typical thermodynamical relations of macroscopic quantities which can be expressed in terms of partition function and its derivatives:

$$U = \langle E \rangle = -\frac{\partial \ln Z}{\partial \beta} \qquad (\text{internal energy}) \tag{4.21}$$

$$F(U,T) = -\frac{1}{\beta} \ln Z \qquad (\text{free energy}) \tag{4.22}$$

$$S(U,T) = k_B(\ln Z + \beta U) = \frac{U - F}{T}\,. \qquad (\text{thermodynamic entropy}) \tag{4.23}$$

The last relation is known as the Legendre transform between thermodynamical potentials U and F, because $F = U - TS$. We also obtain that the temperature can be defined as the derivative of entropy with respect to internal energy

$$\frac{\partial S(U,T)}{\partial U} = \frac{1}{T}\,. \tag{4.24}$$

The previous set of relations and the so-called Legendre structure of thermodynamics is valid not only for Shannon entropy, but it is preserved for a wider class of entropies [51]. We have already observed this structure in the case of multifractal thermodynamics in Sect. 2.5.2 and we will discuss it also for nonextensive thermodynamics based on Tsallis entropy in Sect. 4.3.2. Once we are able to calculate the partition function, we are able to calculate all other thermodynamical quantities (see e.g. Ref. [71])

4.2.3 Concavity and Schur-concavity

Concavity is an important concept that is widely discussed in connection with entropy and is crucial in equilibrium thermodynamics as well as in information theory. At this place, we need to distinguish between two types of concavity. The first type is the concavity of thermodynamical entropy $S(E)$. In equilibrium thermodynamics, it is demanded that the thermodynamical entropy is strictly concave function of its extensive variables (energy, volume, number of particles, etc.). In the case of homogenous entropy we have

$$S(2E) = 2S(E) \geq S(E - \Delta E) + S(E + \Delta E)\,. \tag{4.25}$$

Thus, the system remains in equilibrium state and existence of subsystems in inhomogeneous states is suppressed. More discussion is e.g., in Ref. [72].

In information theory ensure the concavity of entropy $G(\mathcal{P})$ (together with symmetry in all arguments) validity of the maximality axiom, i.e.,

$$\mathrm{argmax}_{\mathcal{P}}\, G(\mathcal{P}) = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)\,. \tag{4.26}$$

Nevertheless, concavity condition is only sufficient but it is not necessary. An alternative approach, weaker than concavity, is called Schur-concavity [73] and it is based on the concept of *majorization* [74]. We define majorization in the following way: a distribution $\mathcal{P} = (p_1, \ldots, p_n)$ is majorized by a distribution $\mathcal{Q} = (q_1, \ldots, q_n)$, i.e., $\mathcal{P} \prec \mathcal{Q}$, if for ordered probability vectors $p_{(1)} \geq p_{(2)} \geq \cdots \geq p_{(n)}$, resp. $p_{(1)} \geq p_{(2)} \geq \cdots \geq p_{(n)}$ holds

$$\sum_{k=1}^{j} p_{(k)} \leq \sum_{k=1}^{j} q_{(k)} \quad \text{for } j \in \{1, \ldots, n\}\,. \tag{4.27}$$

For $j = n$ is the inequality automatically fulfilled because of the normalization condition. A symmetric function $G(p_1, \ldots, p_n)$ is called *Schur-concave* if for all $\mathcal{P} \prec \mathcal{Q}$ is $G(\mathcal{P}) \geq G(\mathcal{Q})$ (Analogously, the function is Schur-convex if for all $\mathcal{P} \prec \mathcal{Q}$ is $G(\mathcal{P}) \leq G(\mathcal{Q})$). There exists also a handy criterion for Schur-concavity. A symmetric function is Schur-concave if for all probabilities p_i, p_j holds

$$(p_i - p_j)\left(\frac{\partial G}{\partial p_i} - \frac{\partial G}{\partial p_j}\right) \leq 0. \tag{4.28}$$

The proof can be found in Ref. [73], together with more criteria. The Schur-concavity of entropy also ensures validity of the maximality axiom because it is easy show that the uniform distribution $\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ is majorized by every other probability distribution.

We shall also note that for trace-class of entropies $G(\mathcal{P}) = \sum_{i=1}^{n} g(p_i)$ is the Schur-concavity equivalent to concavity of function $g(x)$. On the other hand, other entropies, as e.g., Rényi entropy are not concave, but one can show that they are Schur-concave (see Sect. 4.3.1).

4.3 Special Classes of Entropies

This section compares three special classes of entropies. The first two classes, Rényi entropy and Tsallis entropy (also sometimes called Tsallis-Havrda-Charvát entropy, after czech mathematicians J. Havrda and F. Charvát) are popular classes extensively used by large scientific communities. In the first chapter were discussed various applications of Rényi entropy to multifractals. We present another important property of Rényi entropy commonly used in information theory in description of additive systems. On the other hand, Tsallis entropy represents a popular description of nonextensive systems and systems with long-range correlations. Finally, the last class called *hybrid entropy* combines properties of two former entropies.

For each class of entropy, we show its axiomatic definition, its actual functional form, its properties (concavity, extensivity, etc.) and calculate the MaxEnt distribution. When necessary, we mention some other interesting problems. For the last class of entropies, we broadly discuss the properties of MaxEnt distribution and briefly sketch the possible physical meaning of energy gaps present in distributions. Additionally, we also show some asymptotical expansions.

4.3.1 Rényi Entropy: Entropy of Multifractal Systems

Rényi entropy was firstly introduced in 1961 by Alfréd Rényi [75], in connection with distances for probability distributions. The main importance consists in existence of operational definition, as shown in [76]. Apart from that, the entropy has wide applications in theory of multifractals, chaotic systems and similar systems. The Rényi entropy can be axiomatized in a very similar way to Shannon entropy. It consists of four axioms:

1. *Continuity axiom*: for given n and probability distribution $\mathcal{P}$ is $\mathcal{I}_q(\mathcal{P})$ a continuous function with respect to all its arguments.

2. *Maximality axiom*: for given n takes $\mathcal{I}_q(\mathcal{P})$ the maximal value for uniform distribution.

3. *Expansibility axiom*: $\mathcal{I}_q(p_1, \ldots, p_n, 0) = \mathcal{I}_q(p_1, \ldots, p_n)$.

4. *Rényi additivity axiom*: $\mathcal{I}_q(A \cup B) = \mathcal{I}_q(A) + \mathcal{I}_q(B|A)$,
   where $\mathcal{I}_q(B|A) = g^{-1}\left[\sum_i \rho_{i,A}(q)\, g(\mathcal{I}_q(B|A = a_i))\right]$
   is conditional entropy and $\varrho_A(q) = (\rho_{1,A}(q), \ldots, \rho_{n,A}((q))$ is escort distribution corresponding to experiment A. Function g is a positive, invertible function on $[0, \infty)$.

These axioms lead to the functional form of Rényi entropy which can be expressed as

$$\mathcal{I}_q(\mathcal{P}) = \frac{1}{1-q} \ln \sum_{i=1}^{n} p_i^q \,. \tag{4.29}$$

When solving the four axioms, it is easy to show that the function $g(x)$ pertaining to the conditional entropy can be only in the form $g_q(x) = \exp[(1-q)x]$ for $q > 0$. Interestingly, the definition of conditional entropy can be done without the actual knowledge of probability distribution, i.e., only from the knowledge of unconditional entropies

$$\mathcal{I}(B|A) = \mathcal{I}(A \cup B) - \mathcal{I}(A) \,. \tag{4.30}$$

Thus the last relation is valid not only for the Shannon entropy, but also for the whole class of Rényi entropies. Indeed, Rényi entropy is generalization of Shannon entropy and $\lim_{q\to 1} \mathcal{I}_q(\mathcal{P}) = \mathcal{H}(\mathcal{P})$. Consequently, the only difference between Shannon entropy and Rényi entropy is that the conditional entropy is defined in a slightly different way. Both entropies are additive and share many common properties.

Some authors, including Rényi, used an alternative axiomatic approach [17], which differs particularly in the presence of escort distribution $\rho(q)$, which are not present in their definitions. The escort distribution has been originally used in description of dynamical systems [49]. The escort distribution $\rho_i(q) = p_i^q / \sum_j p_j^q$ is also sometimes called "zooming distribution", because the parameter q serves as a magnifier which accentuates different parts of distribution for different values o parameter q. Therefore, the escort distribution has a clear interpretation.

At this place, it is necessary to mention the recent discussion on definition of conditional Rényi entropy (see e.g., Ref. [77]). Apart from the definition arising from the aforementioned axioms, there are several other definitions of conditional entropy [78]. Nevertheless, we have to note that the definition of conditional entropy is inherently connected with the axiomatic definition and different definitions of conditional entropies lead generally to different properties of the entropy. The importance of previously mentioned definition of conditional entropy is in the relation to unconditional entropies. This is important from both theoretical and practical reasons. The conditional entropy can be in this case measured without an actual knowledge of probability distribution, it can be measured only on the basis of unconditional entropy measurements as a difference between entropy of the whole system and the subsystem.

Regarding the observability of Rényi entropy, it has been shown that it is a measurable quantity [79]. This is closely related to the existence of an operational informational definition. Campbell [80] showed that Rényi entropy represents the minimal average price of a message code when the prior occurrences are described by the probability distribution $\mathcal{P}$ and the price is an exponential function of message code-length.

Following points summarize the most important properties of Rényi entropy:

- $\mathcal{I}_q(\mathcal{P})$ is a symmetric function of all its arguments
- $\mathcal{I}_q(\{1, 0, \ldots, 0\}) = 0$
- $\max_{\mathcal{P}} \mathcal{I}_q(\mathcal{P}) = \mathcal{I}_q(\{1/n, \ldots, 1/n\}) = \ln n$
- $\mathcal{H}(\mathcal{P}) \leq \mathcal{I}_q(\mathcal{P})$ for $q \leq 1$ and $\mathcal{I}_q(\mathcal{P}) \leq \mathcal{H}(\mathcal{P})$ for $q \geq 1$
- $\mathcal{I}_q(\mathcal{P})$ is a strictly decreasing function of q for every distribution. This can be easily seen from

$$\frac{\partial \mathcal{I}_q(\mathcal{P})}{\partial q} = \frac{1}{1-q} \left(\langle \ln \mathcal{P} \rangle_q + \mathcal{I}_q(\mathcal{P}) \right) \leq 0 \tag{4.31}$$

- $\mathcal{I}_q(\mathcal{P})$ is a concave function for $q \leq 1$
- $\mathcal{I}_q(\mathcal{P})$ is a Schur-concave function for $q \geq 0$. This is easy to show with help of criterion presented in Sect. 4.2.3, so

$$(p_i - p_j) \left(\frac{\partial \mathcal{I}_q(\mathcal{P})}{\partial p_i} - \frac{\partial \mathcal{I}_q(\mathcal{P})}{\partial p_j} \right) = \left(\frac{p_i - p_j}{\sum_k p_k^q} \right) \left(\frac{p_i^{q-1} - p_j^{q-1}}{q-1} \right) \leq 0 \,. \tag{4.32}$$

In Fig. 4.1 is depicted Rényi entropy for several values of q for binary system. We can observe several aforementioned properties. Mainly the concavity issue and q-monotonicity. Now, we turn the attention to the MaxEnt distribution obtained by maximization of Rényi entropy under constraints.

We discuss the MaxEnt distribution under two types of constraints: first, classic linear average energy $\sum_j p_j E_j = \langle E \rangle$, and second, the q-average energy in terms of escort distribution $\sum_j \rho_j(q) E_j = \langle E \rangle_q$. The Lagrange function can be written in form

$$\mathcal{L}_{\mathcal{I}_q}(\mathcal{P}) = \mathcal{I}_q(\mathcal{P}) - \alpha \sum_i p_i - \beta \sum_i \rho_i(r) E_i \tag{4.33}$$

where r is either equal to 1 or q depending on chosen averaging. In the case of linear averaging, one obtains the equation:

$$\frac{q}{1-q} \frac{p_i^{q-1}}{Z(q)} - \alpha - \beta E_i = 0 \,, \tag{4.34}$$

where $Z(q) = \sum_i p_i^q$ is the partition function. The parameter α can be deduced from normalization condition and we get $\alpha = \frac{q}{1-q} - \beta \langle E \rangle$, so end with the probability distribution

$$p_i = \frac{1}{Z(q)^{1/(1-q)}} \left[1 + (q-1) \frac{\beta}{q} \left(E_i - \langle E \rangle \right) \right]^{1/(q-1)} \,. \tag{4.35}$$

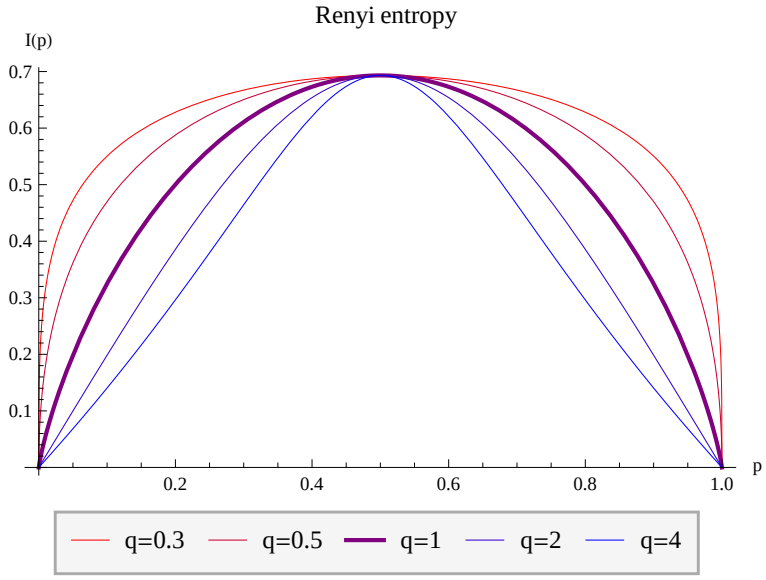


Figure 4.1: Rényi entropy for a binary system with probability distribution $\mathcal{P} = (p, 1-p)$. We can recognize several important properties. Namely, the entropy is concave only for $q < 1$ and it is a decreasing function of q.

In the case of q-averaging we obtain very similar equation

$$\frac{q}{1-q}\frac{p_i^{q-1}}{Z(q)} - \alpha - \frac{qp_i^{q-1}}{Z(q)}(E_i - \langle E\rangle_q) = 0\,, \tag{4.36}$$

resulting into $\alpha = \frac{q}{1-q}$. The distribution can be expressed as

$$p_i = \frac{1}{Z(q)^{1/(1-q)}}\left[1 + (q-1)\,\beta\,(E_i - \langle E\rangle)\right]^{1/(q-1)}\,. \tag{4.37}$$

The distribution is called q-Gaussian distribution. It is a generalization of Gaussian distribution (or Boltzmann distribution in case of energy) and has power-law decay. The distribution was described in connection with nonextensive systems [81]. The analog of inverse temperature (inversely proportional to standard deviation, when it exists) is for $r = 1$ equal to $\Omega_1 = \frac{\beta}{q}$ and for $r = q$ is equal to $\Omega_q = \beta$. The connection between linear averaging and q-averaging is therefore established by rescaling of the inverse temperature parameter. The functional form of the distribution remains the same.

We have seen that the Rényi entropy is a powerful tool in the analysis of many systems, from multifractal systems to systems with power-law decays. Although it does not belong to popular classes of entropies as e.g., class of trace entropies (defined in the beginning of this chapter) or class of f-entropies (widely discussed e.g. in [82]), it has many common properties with these two classes. Apart from that, it possesses

many other important properties as e.g., additivity or Schur-concavity. The next section is devoted to another generalization of Shannon entropy, i.e., Tsallis entropy, which represents an approach to nonextensive systems with long-range correlations and confined state space.

4.3.2 Tsallis Entropy: nonextensive Thermodynamics and Long-range Correlations

Tsallis entropy (also called Tsallis-Havrda-Charvát entropy [50]) is another generalization of Shannon entropy. It was firstly introduced in connection with theory of divergences by Havrda and Charvát [18]. The entropy remained for some time unknown to physicists until the pioneering work of Tsallis [19]. The entropy was used for the description of nonextensive thermodynamics. Since that, there have been found many other applications of Tsallis entropy, as systems with long-range interactions, granular systems or financial markets. From classification point of view it belongs to class of trace entropies and also to f-entropies.

The difference from Shannon entropy lies in the generalization of additivity axiom. Tsallis entropy $\mathcal{S}_q(\mathcal{P})$ is defined by these four axioms:

1. *Continuity axiom*: for given n and probability distribution $\mathcal{P}$ is $\mathcal{S}_q(\mathcal{P})$ a continuous function with respect to all its arguments.

2. *Maximality axiom*: for given n takes $\mathcal{S}_q(\mathcal{P})$ the largest value for uniform distribution.

3. *Expansibility axiom*: $\mathcal{S}_q(p_1, \ldots, p_n, 0) = \mathcal{S}_q(p_1, \ldots, p_n)$.

4. *Tsallis additivity axiom*: $\mathcal{S}_q(A \cup B) = \mathcal{S}_q(A) + \mathcal{S}_q(B|A) + (1-q)\mathcal{S}_q(A)\mathcal{S}_q(B|A)$, where $\mathcal{S}_q(B|A) = \sum_i \rho_i(q)\, \mathcal{S}_q(B|A = a_i)$.

Tsallis entropy can be expressed as

$$\mathcal{S}_q(\mathcal{P}) = \frac{1}{1-q}\left(\sum_i p_i^q - 1\right) \tag{4.38}$$

for $q > 0$. There is a close relation between Tsallis entropy and Rényi entropy, because

$$\mathcal{I}_q(\mathcal{P}) = \frac{1}{1-q}\ln\left(1 + (1-q)\mathcal{S}_q(\mathcal{P})\right)\,. \tag{4.39}$$

Naturally, for $\mathcal{S}_q(\mathcal{P})$ close to zero ($\mathcal{P}$ is close to pure state) is $\mathcal{I}_q \approx \mathcal{S}_q$.

Contrary to Rényi entropy, the conditional entropy cannot be simply expressed as a difference of entropies, but we obtain different relation

$$\mathcal{S}_q(B|A) = \frac{\mathcal{S}_q(A \cup B) - \mathcal{S}_q(A)}{1 + (1-q)\mathcal{S}_q(A)} \,. \tag{4.40}$$

Tsallis entropy is closely connected to so-called q-deformed calculus. Defining the operation of q-addition as

$$x \oplus_q y = x + y + (1-q)xy \tag{4.41}$$

we obtain that the entropy is q-additive, so $\mathcal{S}_q(A \cup B) = \mathcal{S}_q(A) \oplus_q \mathcal{S}_q(B|A)$. It is possible also to define q-analogs of elementary functions as q-exponential

$$e_{\{q\}}(x) = [1 + (1-q)x]^{1/(1-q)} \tag{4.42}$$

and its inverse function, i.e. q-logarithm

$$\ln_{\{q\}}(x) = \frac{x^{1-q} - 1}{1-q} \,. \tag{4.43}$$

For $q \to 1$ we obtain the ordinary functions. Tsallis entropy can be expressed in an elegant way in terms of q-deformed calculus:

$$\mathcal{S}_q(\mathcal{P}) = -\sum_i p_i \ln_{\{q\}}(p_i) = \sum_i p_i \ln_{\{q\}}\left(\frac{1}{p_i}\right) \,. \tag{4.44}$$

We summarize the main properties of Tsallis entropy. The entropy is depicted in Fig. 4.2. To the main properties belong:

- $\mathcal{S}_q(\mathcal{P})$ is a symmetric function of all its arguments
- $\mathcal{S}_q(\{1, 0, \ldots, 0\}) = 0$
- $\max_{\mathcal{P}} \mathcal{S}_q(\mathcal{P}) = \mathcal{S}_q(\{1/n, \ldots, 1/n\}) = \ln_{\{q\}} n$
- $\mathcal{H}(\mathcal{P}) \leq \mathcal{I}_q(\mathcal{P}) \leq \mathcal{S}_q(\mathcal{P})$ for $q \leq 1$ and $\mathcal{S}_q(\mathcal{P}) \leq \mathcal{I}_q(\mathcal{P}) \leq \mathcal{H}(\mathcal{P})$ for $q \geq 1$
- $\mathcal{S}_q(\mathcal{P})$ is a strictly decreasing function of q for every distribution, because

$$\frac{\partial \mathcal{S}_q(\mathcal{P})}{\partial q} = \frac{1}{1-q}\left(\mathcal{S}_q(\mathcal{P}) + \sum_i p_i^q \ln p_i\right) \leq 0 \tag{4.45}$$

- $\mathcal{S}_q(\mathcal{P})$ is a concave function for all $q > 0$. Because Tsallis entropy belongs to class of trace-entropies, we only investigate the function $g(p) = \frac{p^q - p}{1-q}$. Its second derivative is

$$\frac{\mathrm{d}^2 g(p)}{\mathrm{d}p^2} = -qp^{q-2} \leq 0 \,. \tag{4.46}$$

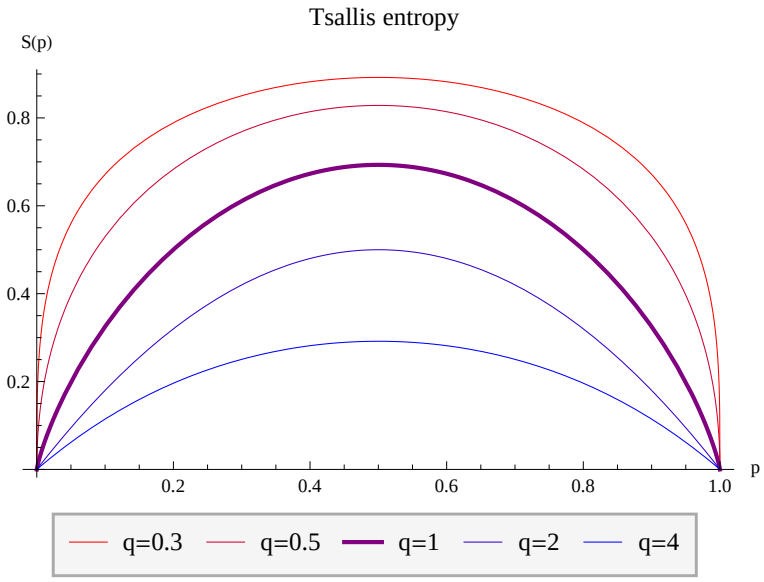


Figure 4.2: Tsallis entropy for a system with probability distribution $(p, 1-p)$. We can observe the concavity of the entropy and the fact that the maximal valueis equal to $\ln_{\{q\}}(2)$. Moreover Tsallis entropy is a monotonic function of q.

Similarly to the previous section, we turn the attention to the MaxEnt distribution. We derive the distribution again under the constraint of linear energy averaging and under q-average. The Lagrange function is

$$\mathcal{L}_{\mathcal{S}_q}(\mathcal{P}) = \mathcal{S}_q(\mathcal{P}) - \alpha \sum_i p_i - \beta \sum_i \rho_i(r) E_i \,. \tag{4.47}$$

In the case of linear averaging, we get the relation for probability distribution

$$\frac{q}{q-1} p_i^{q-1} = -\alpha - \beta E_i \,. \tag{4.48}$$

By elimination of α we get the MaxEnt distribution in the form

$$p_i = \frac{1}{Z(q)^{1/(1-q)}} \left[1 + (q-1) \frac{\beta}{qZ(q)} \left(E_i - \langle E \rangle \right) \right]^{1/(q-1)} . \tag{4.49}$$

In case of q-averaging, we get the similar relation

$$\frac{q}{q-1} p_i^{q-1} = -\alpha - \frac{\beta q p_i^{q-1}}{Z(q)} \left(E_i - \langle E \rangle_q \right) \tag{4.50}$$

resulting into the probability distribution in the form

$$p_i = \frac{1}{Z(q)^{1/(1-q)}} \left[1 + (q-1) \frac{\beta}{Z(q)} \left(E_i - \langle E \rangle_q \right) \right]^{1/(q-1)} . \tag{4.51}$$

In both cases is the distribution very similar to MaxEnt distributions of Rényi entropy, i.e., it is expressible in terms of q-gaussian (or deformed Boltzman) distribution. The only difference is that inverse temperature is equal to $\Omega_1 = \frac{\beta}{qZ(q)}$, resp. $\Omega_q = \frac{\beta}{Z(q)}$. Thus, the inverse temperature is the same as in case of Rényi, but it is divided by partition function. When the entropy is influenced by some particular properties of the probability distribution itself, the distribution is called *self-referential* and has some interesting properties in connection with shifts in energy spectrum [83].

One important application of Tsallis entropy is nonextensive thermodynamics. The term "nonextensive" refers to the fact that systems which are usually described are not extensive, (i.e., there exist correlations in the system, such that the state space grows polynomially [84]). The overview and various applications of nonextensive thermodynamics based on Tsallis entropy can be found in Refs. [4]. The cornerstone of the thermodynamical approach is the definition of partition function, which is in case of Tsallis statistics equal to sum of q-deformed Boltzmann factors

$$\mathcal{Z}(q) = \sum_{j=1}^{n} e_q(\Omega(E_j - \langle E \rangle_r)) \tag{4.52}$$

and the probability distribution is equal to q-gaussian distribution

$$p_i = \frac{e_q(\Omega(E_i - \langle E \rangle_r))}{\mathcal{Z}(q)} \,. \tag{4.53}$$

As discussed in Sect.4.2.2, the Legendre structure of thermodynamics remains unchanged and therefore remain valid all relations that are derived from partition function, including

$$\frac{\partial S_q(U_q, T)}{\partial U_q} = \frac{1}{T} \,. \tag{4.54}$$

In the next section, we combine the axiomatics of Rényi and Tsallis entropy and obtain the new class of entropies with interesting properties.

4.3.3 Hybrid Entropy: Overlap between Nonadditivity and Multifractality

In previous sections we have met two generalizations of Shannon entropy. In both cases were the generalizations obtained by adjusting the additivity axiom. The Rényi additivity axiom changes the definition of conditional entropy which is defined in terms of escort distributions. On the other hand, Tsallis additivity axiom changes the whole addition rule of entropies. When taken into account both nonadditivity and multifractal conditionality, a new class of entropies arises. The entropy is called *hybrid entropy* ($\mathcal{D}_q(\mathcal{P})$) and was firstly introduced in [85] by following four axioms:

1. *Continuity axiom*: for given n and probability distribution $\mathcal{P}$ is $\mathcal{D}_q(\mathcal{P})$ a continuous function with respect to all its arguments.

2. *Maximality axiom*: for given n takes the $\mathcal{D}_q(\mathcal{P})$ the largest value for uniform distribution.

3. *Expansibility axiom*: $\mathcal{D}_q(p_1,\ldots,p_n,0) = \mathcal{D}_q(p_1,\ldots,p_n)$.

4. *J.-A. additivity axiom*: $\mathcal{D}_q(A\cup B) = \mathcal{D}_q(A)+\mathcal{D}_q(B|A)+(1-q)\mathcal{D}_q(A)\mathcal{D}_q(B|A)$, where $\mathcal{D}_q(B|A) = f^{-1}\left[\sum_i \rho_i(q)\, f(\mathcal{D}_q(B|A=a_i))\right]$ and f is a positive and invertible function on $[0,\infty)$.

The generalized additivity axiom combines both nonadditivity and generalized conditioning with the same parameter q, which corresponds also to the parameter of escort distribution. In Appendix E is shown the derivation of a functional form of the entropy. There are also discussed the allowable forms of function f and the uniqueness of the solution. Hybrid entropy can be expressed in the form

$$\mathcal{D}_q(\mathcal{P}) = \frac{1}{1-q}\left(e^{-(1-q)^2\frac{\mathrm{d}\mathcal{I}_q(\mathcal{P})}{\mathrm{d}q}}\left(\sum_{j=1}^{n} p_j^q\right) - 1\right) =$$

$$= \frac{1}{1-q}\left(e^{-(1-q)\sum_{j=1}^{n}\rho_j(q)\ln p_j} - 1\right) = \ln_{\{q\}} e^{-\langle \ln \mathcal{P}\rangle_q}\,. \tag{4.55}$$

We can also recognize that the functional form of entropy consists of parts typical for Tsallis entropy (q-logarithm) and expressions typical for Rényi entropy (q-average). First, it is necessary to discuss, which values of parameter q are obeying all axioms. Particularly important is the maximality axiom which is used in the proof only in special cases (see Appendix E). We have to verify if the uniform distribution actually maximizes $\mathcal{D}_q$. Because $\ln_{\{q\}}(x)$ and $\exp(x)$ are monotonic functions, we can treat only $\langle \ln \mathcal{P}\rangle_q$. For sake of simplicity, let us make the discussion for the case of binary distribution $\mathcal{P} = (p, 1-p)$. The stationary points are determined by the equation

$$\frac{\partial\langle \ln \mathcal{P}\rangle_q}{\partial p} = \frac{1}{Z(q)^2}\left[p^{2q-1} - (1-p)^{2q-1} + p^{q-1}(1-p)^q - p^q(1-p)^{q-1}\right.$$

$$\left. -qp^q(1-p)^{q-1}\ln\left(\frac{1-p}{p}\right) + qp^{q-1}(1-p)^q \ln\left(\frac{p}{1-p}\right)\right] \;=\; 0\,. \tag{4.56}$$

After substitution $y = \frac{p}{1-p}$ and a few straightforward transformations we end with

$$\Psi_q(y) \;\equiv\; q\ln y - \frac{1 - y^{q-1} + y^q - y^{2q-1}}{y^q + y^{q-1}} \;=\; 0\,. \tag{4.57}$$

In Ref. [24] is shown that the equation $\Psi_q(y) = 0$ has for $q \geq \frac{1}{2}$ the only one solution, i.e., $y = 1$ leading to $p = \frac{1}{2}$. On the other hand, for $q < \frac{1}{2}$ there exist two other solutions leading to two other stationary points. From the nature of the entropy is apparent that these two points are local maxima, while $p = \frac{1}{2}$ is the local minimum. Consequently, for $q < \frac{1}{2}$, the entropy does not fulfill the maximality axiom.

We summarize the main properties of hybrid entropy in the following points:

- $\mathcal{D}_q$ is a symmetric function of all arguments
- $\mathcal{D}_q(\{1, 0, \ldots, 0\}) = 0$
- $\mathcal{D}_q(\mathcal{P}) \geq 0$
- $\lim_{q\to 1} \mathcal{D}_q = \lim_{q\to 1} \mathcal{I}_q = \lim_{q\to 1} \mathcal{S}_q = \mathcal{H}$
- $\max_{\mathcal{P}} \mathcal{D}_q(\mathcal{P}) = \mathcal{D}_q(\{1/n, \ldots, 1/n\}) = \ln_{\{q\}} n$ for $q \geq \frac{1}{2}$
- $\mathcal{H}(\mathcal{P}) \leq \mathcal{I}_q(\mathcal{P}) \leq \mathcal{S}_q(\mathcal{P}) \leq \mathcal{D}_q(\mathcal{P}) \leq \ln_{\{q\}} n$ for $q \leq 1$ and $\mathcal{D}_q(\mathcal{P}) \leq \mathcal{S}_q(\mathcal{P}) \leq \mathcal{I}_q(\mathcal{P}) \leq \mathcal{H}(\mathcal{P}) \leq \ln n$ for $q \geq 1$
- $\mathcal{D}_q$ is a strictly decreasing function of q, i.e., $\frac{\partial \mathcal{D}_q}{\partial q} \leq 0$
- $\mathcal{D}_q$ is a concave function for $q \in [\frac{1}{2}, 1]$. Because $\ln_{\{q\}}$ is concave and nondecreasing function for all $q > 0$, we have to treat only

$$\frac{\partial^2}{\partial p_i^2}\left(e^{-\langle \ln \mathcal{P}\rangle_q}\right) \;=\; e^{-\langle \ln \mathcal{P}\rangle_q}\left[\left(\frac{\partial \langle \ln \mathcal{P}\rangle_q}{\partial p_i}\right)^2 - \frac{\partial^2 \langle \ln \mathcal{P}\rangle_q}{\partial p_i^2}\right]. \tag{4.58}$$

  It can be shown that in the interval $q \in [\frac{1}{2}, 1]$ is the second derivative always negative, but for $q \geq 1$ there are regions, where it is positive.

- $\mathcal{D}_q$ is a Schur-concave function for all $q \geq \frac{1}{2}$. Shi et al. have shown [86] that a subset of functions called Gini means, which can be expressed in the form

$$G(q; x, y) = \exp\left(\frac{x^q \ln x + y^q \ln y}{x^q + y^q}\right), \tag{4.59}$$

  is Schur-concave for $x > 0$ and $y > 0$ in the interval $q \geq \frac{1}{2}$. As a consequence, the hybrid entropy is Schur-concave in the whole region, where it fulfills the maximality axiom.

The entropy is depicted in Fig. 4.3 for several values of q. Let us turn our attention to the derivation of MaxEnt distribution. Similarly to previous sections, we treat two

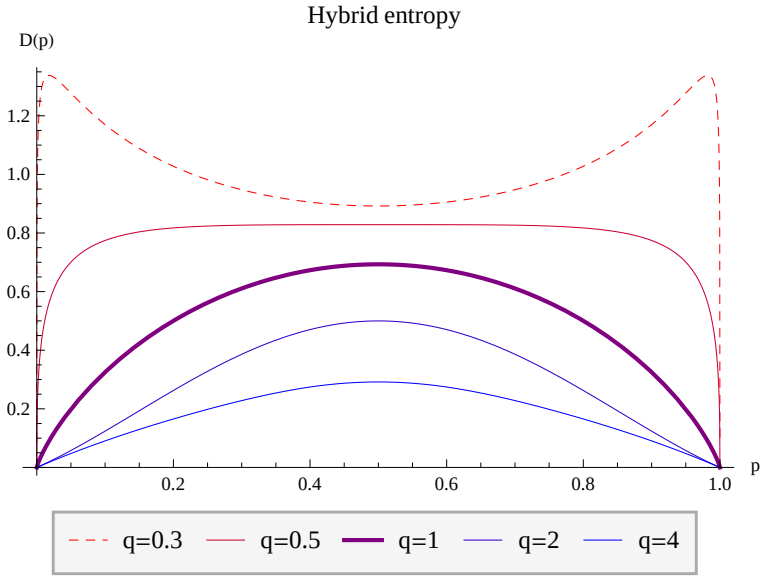


Figure 4.3: Hybrid entropy for a binary system with probability $\mathcal{P} = (p, 1-p)$. The entropy does not fulfill the maximality axiom for $q = 0.3$. We observe two local maxima equal to 1.34 given by probability $p \doteq 0.98$ and a local minimum equal to $\ln_{\{0.3\}}(2) \doteq 0.89$ for $p = \frac{1}{2}$.

types of energy averaging, i.e. linear average and q-average. The Lagrange function reads:

$$\mathcal{L}_{\mathcal{D}_q}(\mathcal{P}) = \mathcal{D}_q(\mathcal{P}) - \alpha \sum_i p_i - \beta \sum_i \rho_i(r) E_i \,. \tag{4.60}$$

For the case of linear averaging we obtain that the MaxEnt distribution is the solution of the equation

$$\alpha \;=\; e^{(q-1)\langle \ln \mathcal{P} \rangle_q} \left[q \left(\langle \ln \mathcal{P} \rangle_q - \ln p_i \right) - 1 \right] \frac{(p_i)^{q-1}}{\sum_k (p_k)^q} \;-\; \beta (E_i - \langle E \rangle) \,. \tag{4.61}$$

The Lagrange parameter α can be determined by multiplying the previous equation by p_i and summing up over i. We get that $\alpha = -e^{(q-1)\langle \ln \mathcal{P} \rangle_q}$. Plugging back into the original equation, we obtain

$$Z(q) p_i^{1-q} \;=\; \frac{\alpha}{\alpha + \beta (E_i - \langle E \rangle)} \left[q \ln p_i - \frac{q \ln(-\alpha)}{q-1} + 1 \right] . \tag{4.62}$$

The equation can be solved in terms of Lambert W–function (The main properties of Lambert function are summarized in Appendix F). The resulting distribution is equal to

$$p_i = \left[\frac{q\alpha}{(q-1)Z(q)(\alpha+\beta\Delta E_i)}\ W\left(-\frac{Z(q)\,(q-1)e^{\left(\frac{q-1}{q}\right)}}{\alpha q}\ \left(1+\frac{\beta}{\alpha}\Delta E_i\right)\right)\right]^{1/(1-q)}$$
$$= \frac{1}{(-\alpha)^{1/(1-q)}e^{1/q}}\ \exp\left\{\frac{1}{(q-1)}W\left(-\frac{Z(q)\,(q-1)e^{\left(\frac{q-1}{q}\right)}}{\alpha q}\ \left(1+\frac{\beta}{\alpha}\Delta E_i\right)\right)\right\}. \quad (4.63)$$

The second representation is a direct consequence of definition of Lambert W-function. A few comments should be noted now. First, from the properties of Lambert function is clear that the probability is always positive. This is apparent from the second representation. Second, for the limit $q \to 1$, the original Boltzmann-Gibbs distribution is recovered. Third, contrary to the previous cases, Eq. (4.62) does not have solution for all energies, which is reflected in the fact that Lambert W–function is not defined on the whole domain of real numbers, but only on the interval $[-e^{-1}, \infty)$.

In the case of $r = q$, we obtain very similar equation defining the MaxEnt distribution:

$$\alpha p_i^{1-q} Z(q) = e^{(q-1)\langle \ln \mathcal{P}\rangle_q}\left[q\left(\langle \ln \mathcal{P}\rangle_q - \ln p_i\right) - 1\right] - q\beta(E_i - \langle E\rangle_q)\,. \quad (4.64)$$

Similarly to linear averaging, α can be determined as $\alpha = -e^{(q-1)\langle \ln \mathcal{P}\rangle_q}$ and the previous equation becomes

$$Z(q) \ = \ (p_i)^{q-1}\left[q\ln p_i + \left(1 - \frac{q\ln(-\alpha)}{q-1} - \frac{q\beta}{\alpha}\,(E_i - \langle E\rangle_q)\right)\right]\,. \quad (4.65)$$

This equation can be again solved in terms of Lambert W–function and eventually we arrive at a very similar distribution as in the first case:

$$p_i = \left[\frac{q}{Z(q)(q-1)}\,W\left(\frac{Z(q)(q-1)}{q}\,e^{\frac{q-1}{q}\left\{1-\frac{q\ln(-\alpha)}{q-1}-\frac{q\beta}{\alpha}(E_i-\langle E\rangle_q)\right\}}\right)\right]^{1/(1-q)}$$
$$= \exp\left\{\frac{W\left(\frac{Z(q)(q-1)}{q}\,e^{\frac{q-1}{q}\mathcal{E}_i}\right)}{q-1} - \frac{\mathcal{E}_i}{q}\right\} \quad (4.66)$$

where $\mathcal{E}_i \ = \ 1 - \frac{q\ln(-\Phi)}{q-1} - \frac{q\Omega}{\Phi}\,(E_i - \langle E\rangle_q)$.

Particularly interesting is the case, when the system is additionally fulfilling *multifractal scaling*. In this case we have some typical multifractal spectrum. When the scale s goes to zero, all scaling exponents α_i are dominated by the most likely value $\langle\alpha\rangle_q$, while contributions of other scaling exponents have zero fractal dimension, i.e., have probability equal to zero. This is a consequence of *curdling theorem* [87]. As

shown in the Section 2.5.2, the inverse temperature is equal to q and energy is equal to $E_i(s) = -\alpha_i \ln s$. Now, applying the multifractal formalism to the equation (4.62), we obtain the interesting relation

$$\varepsilon^{\tau(q)+\alpha_i(1-q)} \;\sim\; 1 \;+\; q\,[\alpha_i - \langle\alpha\rangle_q(\varepsilon)] \left(1 + \frac{\widetilde{\beta}}{\widetilde{\alpha}}\right) \ln \varepsilon\,, \tag{4.67}$$

where $\widetilde{\alpha}$, resp. $\widetilde{\beta}$ are the Lagrange multipliers (the tilde is used to distinguish them from the scaling exponents) and $\langle\alpha\rangle_q = \sum_j \rho_j(q)\alpha_j$ is related to q-mean of logarithm of probability distribution, so

$$\langle \ln \mathcal{P}\rangle_q = \sum_j \rho_j(q) \ln p_j = \sum_j \rho_j(q)\alpha_j \ln \varepsilon = \langle\alpha\rangle_q \ln \varepsilon\,. \tag{4.68}$$

Similarly to conventional thermodynamics, also in multifractal case are the fluctuations proportional to square root of characteristic scale, so $|a_i - \langle a\rangle_q| \sim \frac{1}{\sqrt{-\ln \varepsilon}}$. The only non-trivial scaling is obtained when

$$\left| q \left(1 + \frac{\widetilde{\beta}}{\widetilde{\alpha}}\right) \right| \;<\; \frac{1}{\sqrt{-\ln \varepsilon}}\,. \tag{4.69}$$

In this case, the probability distribution can be recast as

$$p_i \;\sim\; [1 + (1-q)(a_i - \langle a\rangle_q) \ln \varepsilon]^{1/(1-q)}\;. \tag{4.70}$$

In connection with multifractals is more conventional to estimate the total probability of a phenomenon with scaling exponent α, which is proportional to

$$P_i(\alpha, s) \propto s^{-f(\alpha_i)+\alpha_i}\,, \tag{4.71}$$

which is in the quadratic expansion equal to

$$P_i \;\propto\; \left[1 \;-\; (1-q)\frac{(\alpha_i - \langle\alpha\rangle_q)^2}{2(\Delta\alpha)^2}\right]^{1/(1-q)}\,, \tag{4.72}$$

where $\Delta\alpha$ is the standard deviation around the mean value. As a consequence of previous relations, we mention that the relation between Lagrange multipliers is $\widetilde{\beta} = q|\widetilde{\alpha}|$. More details can be found in Ref. [24]. In the case of linear averaging we obtain after a straightforward derivation very similar results, but in terms of $\langle\alpha\rangle_1$. The inverse temperature is given as $\widetilde{\beta} = (q-1)|\widetilde{\alpha}|$.

Apart from multifractal systems, it is also important to study the high-temperature and low-temperature regimes. In Appendix F are thoroughly discussed the asymptotical expansions of Lambert W–function. For so-called "high-temperature" expansion, i.e., when $\widehat{\beta} \ll 1$, it is possible to use the Taylor expansion of Lambert W–function and exponential function and obtain that

$$p_i \approx \frac{1}{Z}\left[1-(1-q)\beta_r^*\Delta_r E\right]^{1/(1-q)} \tag{4.73}$$

with $\beta_1^* = -\frac{q}{q-1}\frac{\widehat{\beta}W(x)}{\widehat{\alpha}(W(x)+1)}$, resp. $\beta_q^* = -\frac{\widehat{\beta}}{\widehat{\alpha}(W(x)+1)}$ where $x = -\frac{Z(q)(q-1)}{\widehat{\alpha}q}\exp\left(\frac{q-1}{q}\right)$. The partition function is

$$Z = \sum_k \left[1 - (1-q)\beta_r^*\Delta_r E_k\right]^{1/(1-q)} = \left[\frac{q}{\kappa(q-1)}W(x)\right]^{1/(q-1)}. \tag{4.74}$$

The distribution obtained by the high-temperature expansion is the q-gaussian distribution similar to distribution obtained by Rényi or Tsallis entropy. Similarly to Tsallis entropy, the temperature is depending on the probability distribution, i.e., *self-referential*.

The situation with "low-temperature" expansion is much more complicated. Depending on parameter q, constraint parameter r and the sign of $\Delta_r E$, there can arise several different situations. The argument of Lambert function can be either close to zero, infinity, can have two possible solutions because of existence of two real branches of complex W–function, or does not have to exist at all. Also the resulting distribution approximations can have form of exponential functions, q-Gaussian distributions or even more complicated forms of distributions. The whole discussion is realized again in Ref. [24].

It is a challenging task to find some systems, where the hybrid entropy can be successfully applied. Lambert W–function can be found in connection with such systems as Lotka–Volterra models, Tonks gas or quantum systems. These systems are possible candidates for application of hybrid entropy. Also multifractal systems with long-range interactions can be an example of system driven by hybrid entropy. All these aspects provides interesting directions for further research.

Chapter 5

Applications in Financial Markets

In this chapter are presented applications of models discussed in previous chapters to financial markets. The connection between physical models and financial markets has been observed first in the beginning of twentieth century. In the work of L. Bachelier [88], Brownian motion was firstly applied to prediction of financial markets. Still, for many decades, the practitioners from finance did not take enough attention to other disciplines and more sophisticated mathematical models. The situation changed in early 90's, because there arose a new scientific field combining models familiar from physics, which found new applications is financial markets. The branch is called *econophysics*. The main contributions to that field are summarized in the classic books by E. Stanley and R. Mantegna [27], J. P. Bouchard and M. Potters [2] or W. Paul and J. Baschangel [1].

We focus on some specific applications connected to previous theoretical chapters. Firstly, the applications of multifractal models into financial markets have become a hot topic. We apply the methods introduced in Chapter 2 to measure multifractal scaling exponents of financial series and try to compare advantages and problems of each particular method. Secondly, we discuss the possibility of applications of diffusion models based on double-fractional diffusion to option pricing. We compare it with classic Black-Scholes model and the Lévy fractional model based on totally asymmetric stable distributions and briefly discuss the possibility of applications to other types assets.

5.1 Estimation of Multifractal Spectra of Financial Time Series

Our aim is to compare methods presented in Sections 2.3, 2.4 and discuss their applicability to real complex time series. We divide the discussion into several steps. First, we illustrate the necessity of precise estimation of underlying histograms in case of Multifractal diffusion entropy analysis (MFDEA) in order to obtain relevant δ-spectrum. We

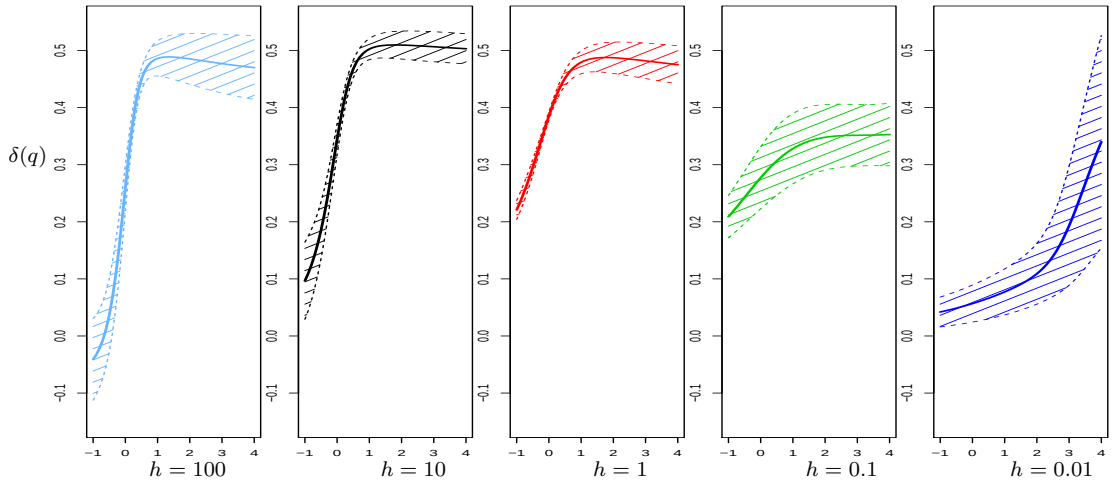


Figure 5.1: Estimated δ-spectra (central line) and 99% confidence intervals (shaded regions) of daily time series of S&P 500 for different values of bin-width h. For bin-width far from the optimal width is the spectrum diminished and the confidence intervals get wider. Particularly, for under-fitted histograms the error is most dramatic for small q's, for over-fitted histograms the error visible for large q's.

method	optimal bin-width for $q = 1$	in multiples of $u = 3 \times 10^{-4}$
Scott	0.00470	14.10
Freedman–Diaconis	0.00384	12.81

Table 5.1: Optimal values of h_1^* for different methods, h_q^* can be easily obtained from Eq. (2.68). The results are also converted to the same units like in Fig. 2.3, so that the reader can easily compare the results with previous values.

test the two methods, i.e., Scott and Freedman-Diaconis method, on the sample time series of daily returns of S&P 500 in time period 1950-2013 (approximately 16000 entries). This is a good example of complex series with multiple scaling exponents. The procedure of histogram estimation is depicted in Fig. 2.1. The necessity of proper bin-width estimation is presented in Figs 2.3, resp. 2.4. The resulting δ-spectra estimated with different bin-widths are depicted in Fig. 5.1. The under-fitted histograms (too large bin-width) contain not enough information about the time series, while the over-fitted histograms (too small bin-width) cannot also describe properly the dynamics of the series due to the finite amount of measured data. For not enough data it is probable that the histogram is disintegrated into a normalized count function and does not recover the proper nature of the series. This is reflected in the estimated spectra, indeed. We observe that the spectra differ and also the confidence intervals are different for different values of q. This supports the necessity of q-dependent bin-width. Especially for too

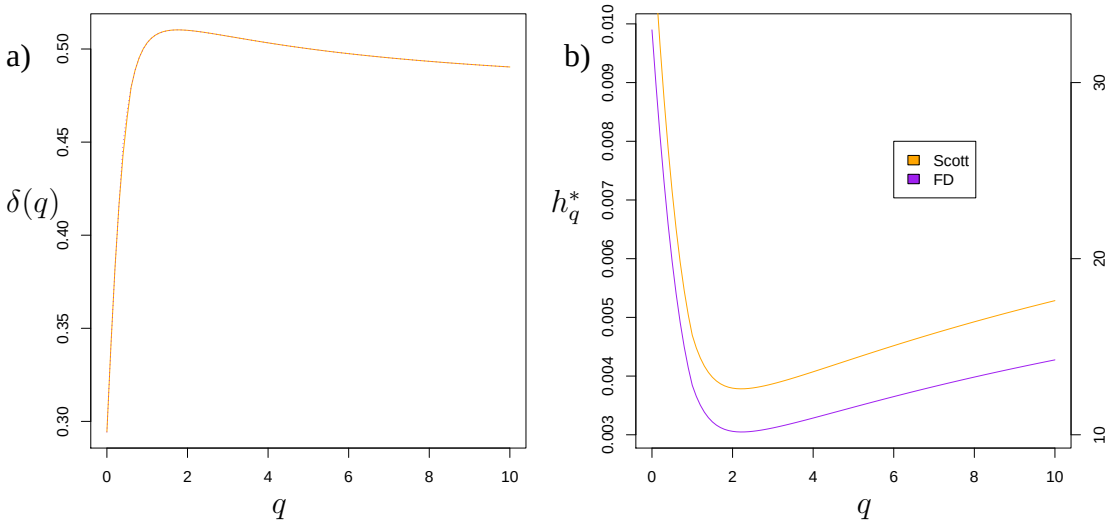


Figure 5.2: Left: δ-spectrum for bin-widths estimated by Scott rule and Freedman-Diaconis rule. Spectra for both methods coincidence. Right: Optimal bin-widths $\hat{h}_q^*$ for both methods. Left y-axis displays natural units, right y-axis compares the width to multiples of $u = 3 \times 10^{-4}$ for comparison with Fig. 5.1

small bin-width is the spectrum deformed from the optimal case and the confidence intervals are large. We have calculated the optimal bin-width for both presented methods. The values $q = 1$, corresponding to classic Scott, resp. Freedman-Diaconis method are listed in Tab. 5.1. The optimal bin-width function depending on q and optimal spectra for both methods are shown in Fig. 5.2. Although the Scott method estimates the optimal bin-width slightly larger than Freedman-Diaconis method, the spectra practically coincide. This is caused by the fact that the prices are rounded to dollars and cents and therefore the data have finite precision. Hence, the small change in bin-width does not necessarily change estimated histograms.

The second part of the analysis is to apply the methods used for estimation of multifractal exponents to several kinds of financial assets on different time scales. We want to test the robustness of each methods, discuss their limitations and to find optimal policy when analyzing the multifractal properties of time series. The main results of the analysis are presented in Ref. [22]. We have done the comparison mainly between f-spectrum obtained from Multifractal detrended fluctuation analysis (MFDFA, defined in Sect. 2.3.3) and δ-spectrum obtained from Multifractal diffusion entropy analysis (MFDEA, defined in Sect. 2.3.4). There were chosen four assets for the analysis, namely the stock index Nikkei 225 (main index of Tokyo stock exchange), ASE Composite index (main index of Athens stock exchange), IBM stock and the VIX index (implied

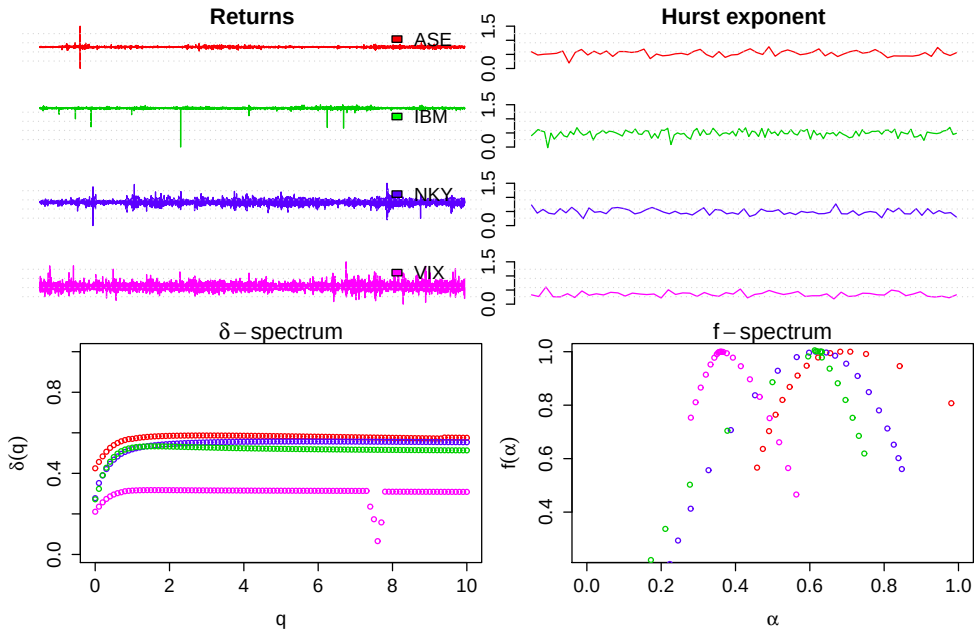


Figure 5.3: Multifractal spectra of daily time series. We observe that multifractal exponents of VIX differ from spectra of other series. This is caused by different nature of volatility. Consequently, the Hurst exponent of VIX series is noticeably lower than for other series.

volatility of S&P index), all recorded on the daily basis and the high-frequency basis. High-frequency data are from year 2013 and have approximately 100000 records and daily data are recoded during the last 10-20 years (depending on particular index) and contain 5000-10000 entries. We also estimate Hurst exponent, f-spectrum obtained from MFDFA and δ-spectrum obtained from MFDEA. Fig. 5.3 shows the spectra for daily series and Fig. 5.4 depicts the spectra for high-frequency series. It is possible to observe discontinuities in all spectra which can be caused by the presence of correlations or power-laws in financial series. The discontinuities are observed in the case of daily data of VIX and in several cases of high-frequency data. This is quite natural, because VIX has different characteristic scaling from the other series, which is given by the nature of the index. In the case of high-frequency data, we often work with illiquid data with calm periods and sudden jumps. Generally, we see that the high-frequency data possess a richer structure of scaling exponents, while in case of aggregated daily data, some scaling exponents disappear, which is expected. Because all methods are based on linear regression, it is always better to combine several methods in order to obtain a real picture of multifractal nature of our system.

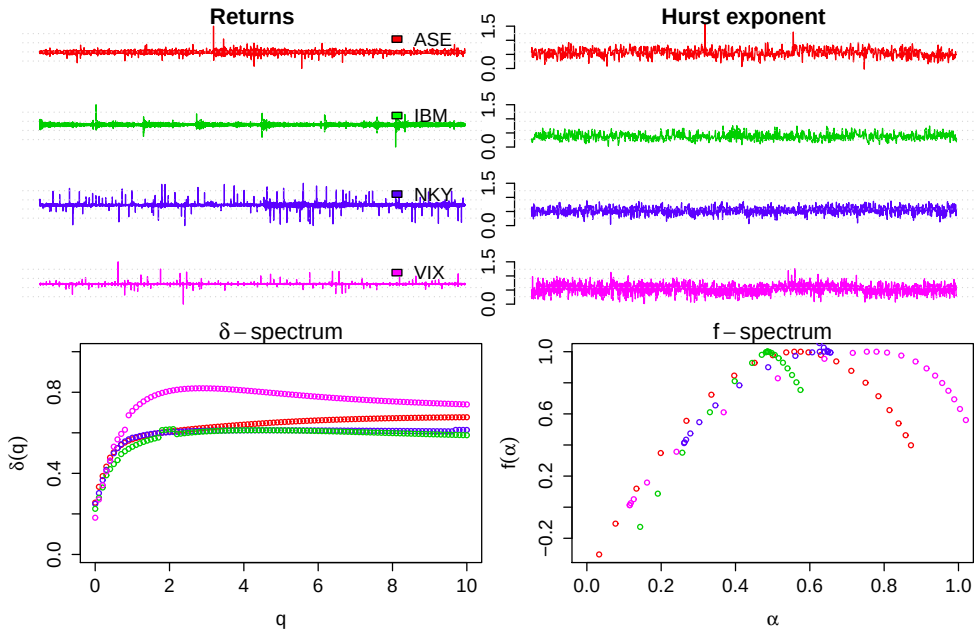


Figure 5.4: Multifractal spectra of high-frequency data. The data are illiquid and exhibit power-law behavior. This is reflected in discontinuities in both spectra, mainly in the multifractal f-spectrum.

5.2 Option Pricing Based on Double-Fractional Diffusion

The first mathematically rigorous option pricing model, based on Brownian motion, was published in 1973 [89] by F. Black and M. Scholes. Scholes together with Merton received later the Nobel prize in economics. The model became very popular and most of the finance community still use the Black-Scholes model for option pricing. Nevertheless, as discussed in chapter about diffusion, the classic Brownian motion does not reflect the complex behavior of financial markets, including large jumps, long-range correlations or regime switching and leads to improper risk estimation. This motivated many scientists to generalize the Black-Scholes model and to invent more sophisticated option pricing models which are able to model the risk more realistically. Among others, let us mention models based on Lévy distributions [90], truncated Lévy distributions [91], multifractal volatility [11], jump processes [15], fractional Brownian motion [92, 93] and double-stochastic equation [94]. We focus on the approach based on stable distributions, because for systems without long-range correlations are distributions limiting distributions of diffusion processes [3]. As discussed in Sect. 3.3.2, Lévy flights are solutions of (spatially) fractional diffusion equations. We generalize Lévy option pricing to double-fractional model, which brings about some more complex behavior and also

All options			
parameter	Black-Scholes	Lévy stable	Double-fractional
α	-	1.493(0.028)	1.503(0.037)
γ	-	-	1.017(0.019)
σ	0.1696(0.027)	0.140(0.021)	0.143(0.030)
agg. error	8240(638)	6994(545)	6931(553)
Call options			
parameter	Black-Scholes	Lévy stable	Double-fractional
α	-	1.563(0.041)	1.585(0.038)
γ	-	-	1.034(0.024)
σ	0.140(0.021)	0.118(0.026)	0.137(0.020)
agg. error	3882(807)	3610(812)	3550(828)
Put options			
parameter	Black-Scholes	Lévy stable	Double-fractional
α	-	1.493(0.031)	1.508(0.036)
γ	-	-	1.047(0.017)
σ	0.193(0.039)	0.163(0.034)	0.163(0.037)
agg. error	3741(711)	3114(591)	2968(594)

Table 5.2: Estimated mean values and standard deviations of model parameters (α, γ, σ) and aggregated errors of three considered models, i.e., Black-Scholes model, Lévy stable model and Double-fractional model. The results are presented for three cases: estimation done for all options and for calls and puts separately. We see that the mean value of γ is very close to one for all options. On the other hand, in the case of separate estimation for call options and put options is γ larger than one.

the aforementioned regime switching between kernel mode and long-memory mode. We test the model on European options of index S&P 500 traded in November 2008. The price of European call option can be determined as

$$C_{(\alpha,\gamma,\kappa)}(S_t, K, \tau) = e^{-r\tau} \int_{\mathbb{R}} \mathrm{d}y \ \left[S_t e^{\tau(r-q+\mu)+y} - K\right]^+ g^{DF}(y, \tau) = \tag{5.1}$$

$$= e^{-r\tau} \int_{\mathbb{R}} \mathrm{d}y \ \left[S_t e^{\tau(r-q+\mu)+y} - K\right]^+ \frac{\Gamma(\kappa)}{2\alpha\pi i y} \int_{c-i\infty}^{c+i\infty} \frac{\Gamma\left(\frac{s}{\alpha}\right)\Gamma\left(1-\frac{s}{\alpha}\right)\Gamma(1-s)}{\Gamma\left(\kappa-\frac{\gamma}{\alpha}s\right)\Gamma\left(\frac{(\alpha-\theta)s}{2\alpha}\right)\Gamma\left(1-\frac{(\alpha-\theta)s}{2\alpha}\right)} \left[\frac{y}{(-\mu\tau^\gamma)^{1/\alpha}}\right]^s \mathrm{d}s.$$

The corresponding put price can be calculated from the *put-call partity* relation

$$P_{(\alpha,\gamma,\kappa)}(S_t, K, \tau) = C_{(\alpha,\gamma,\kappa)}(S_t, K, \tau) - S_t e^{-q\tau} + K e^{-r\tau}. \tag{5.2}$$

Green functions and corresponding option prices are shown in Fig. 5.5. For $\alpha = 2$ and $\gamma = 1$ we recover the Black-Scholes model. The parameters play the role of risk redistribution parameters. In the case of α the lower the parameter means higher

probability of drops and higher price of those options, for which $K < S_t e^{-q\tau}$ is higher. On the other hand, the price is lower for options, for which $K > S_t e^{-q\tau}$. Similarly, the parameter γ influences the temporal redistribution. For $\gamma < 1$, the options with short expiration become more expensive and the options with long expiration become cheaper, and vice versa for $\gamma > 1$. This usually reflects the situation, when the actual market evolution is hard to predict, on the other hand, the long-term dynamics is not affected by actual evolution.

The calibration of the model is done on the example of S&P 500 options traded in November 2008. Following related works [90], we use out-of-the-money options to find the set of parameters, which minimize the *aggregated error*:

$$(\alpha_O, \gamma_O, \sigma_O) = \arg \min_{(\alpha,\gamma,\sigma)} \sum_{\tau \in \mathcal{T}, K \in \mathcal{K}} |\mathcal{O}_{\alpha,\gamma,\kappa} - \mathcal{O}_{\text{market}}| \,. \tag{5.3}$$

The optimization is done for each trading day for three models: Black-Scholes model [89], Lévy stable model [90] and double-fractional model [23]. Because all values of γ are close to one, the differences between particular types of derivatives are negligible, so we use Caputo derivatives in the whole analysis. The results are presented in Tab. 5.2. For comparison, for each method is also presented the aggregated error. It is obvious that in comparison with Black-Scholes model, the latter two models represent a substantial improvement. In the case of Double-fractional model the improvement of parameters for all options is not significant. Nonetheless, in the case of separate calibration of call options and put options, the improvement is more significant. In Fig. 5.6 are presented daily estimates of parameters α, γ and the ratio $\Omega = \gamma/\alpha$, which is the scaling exponent of Green function. Apart from the parameters is also presented the ratio between aggregated errors of Lévy stable model and double-fractional model. In some particular days is the difference between error of Lévy model and double-fractional model quite large. The second finding is that while parameters α and γ fluctuate, the scaling parameter Ω is more stable.

The main advantage of the double-fractional model is the presence of temporal risk redistribution, which allows to distinguish between actual, short-term risk and long-term evolution. From the theoretical point of view, the presence "superstatistical", slow diffusion regime with two distinguishable time scales, and fast-diffusion mode with long-range memory allows to describe different situations. This regime-switching approach is applicable in other scientific fields a could be possibly combined with other similar models.

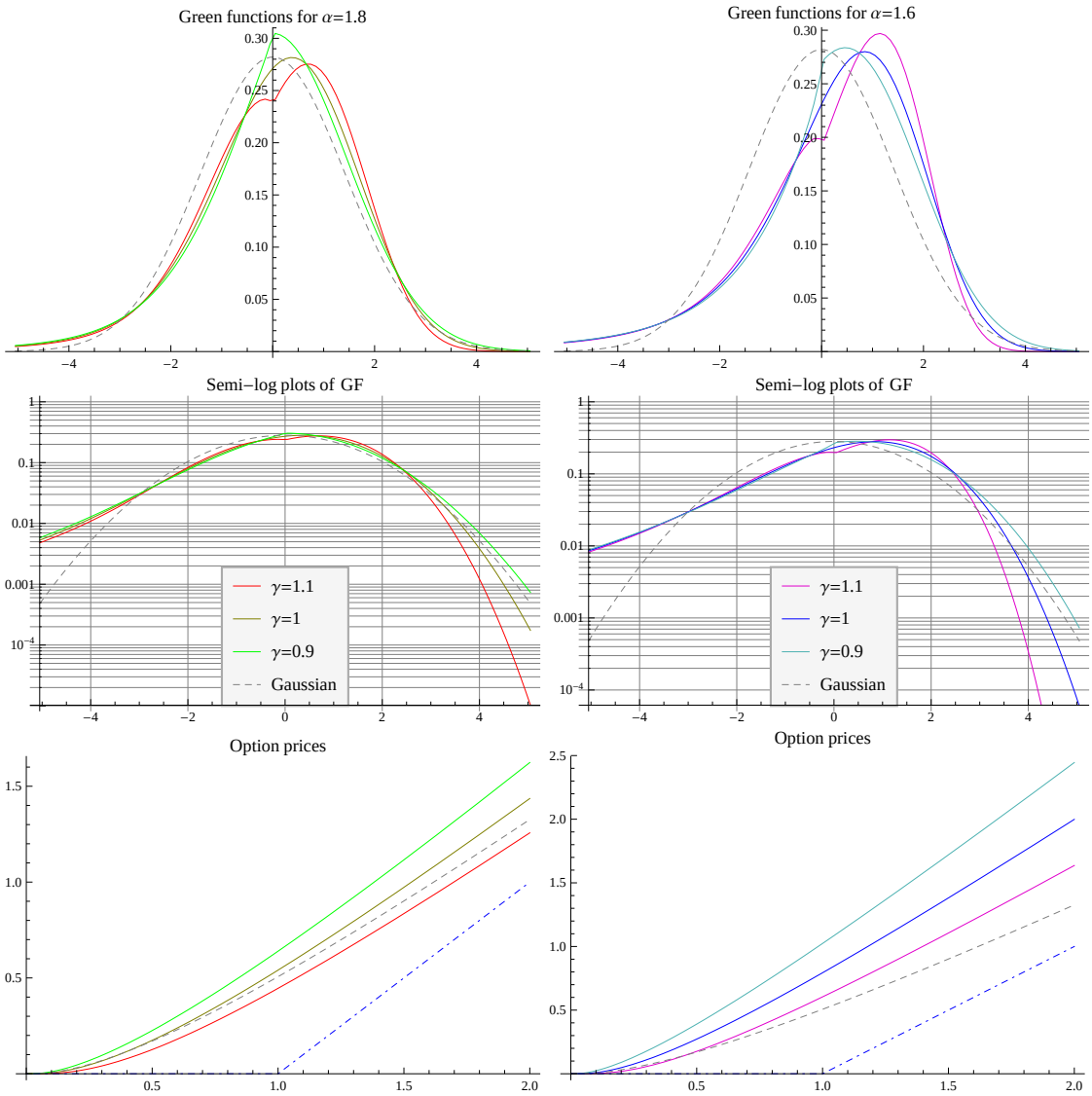


Figure 5.5: linear plots of Green functions (top), semi-log plots of Green functions (center) and corresponding option prices (bottom) for $\alpha = 1.8$ (left) and $\alpha = 1.6$ (right) and comparison with the Black-Scholes model (grey dashed lines). There exist some particular choices of parameters for which are the option prices cheaper than BS model and vice versa.

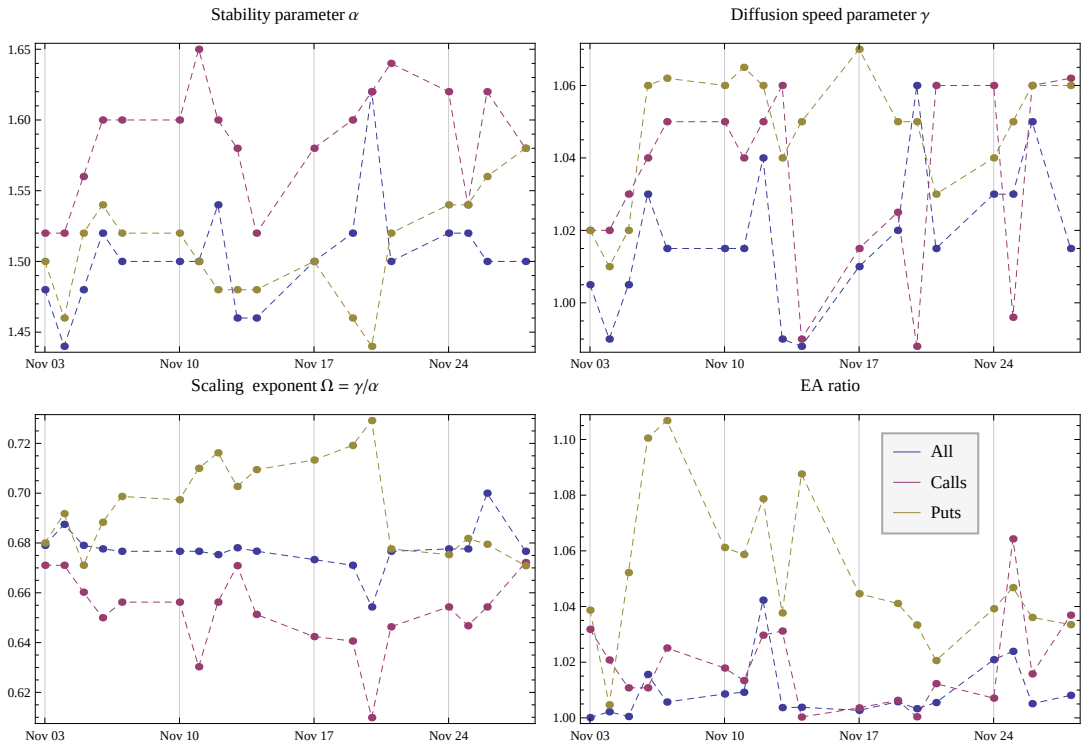


Figure 5.6: Estimated values of stability parameter α, diffusion speed parameter γ, scaling exponent Ω and the ratio of aggregated errors between Lévy model and Double-fractional model for each particular day. The calibration is done for all options and for calls and puts separately. We see that for call and put options treated separately is the improvement of the Double-fractional model more significant. The parameter Ω measures the ratio between γ and α and corresponds to the temporal scaling exponent, so $g(x,t) \sim t^{\Omega}$. For BS model ($\Omega = \frac{1}{2}$) corresponds to Hurst exponent of Brownian motion. The graphics shows that Ω exhibits more stable behavior than α and γ.

Chapter 6

Conclusions and Perspectives

The thesis presented several subject matter related to the currently broadly discussed topic of complex systems. All of these models are based on very universal ideas of scaling, similarity, additivity and generalized statistics. We discussed both their theoretical aspects and practical applications mainly to financial markets and thermodynamics. Nevertheless, the universality of the presented models predestines them for further applications in physics, biology or chemistry.

The main aim in the case of multifractal analysis was to compare several methods for multifractal spectrum estimation. Among others, to the most discussed models belong Detrended fluctuation analysis and Diffusion entropy analysis. We have compared their effectiveness in the matter of heavy-tailed data. We have also discussed technical details of both methods and compared both methods on the real financial time series for daily data and high-frequency data. In the case of Diffusion entropy analysis we have pointed out that the crucial point for the proper calculation of scaling exponents is the estimation of histogram bin-width. Too large or too small bin-width (i.e. too many boxes or too few boxes) does not describe the underlying distribution properly. We have also derived the formula for the optimal bin-width depending on the Rényi entropy parameter q.

In the chapter on generalized diffusion we have compared several existing models of anomalous diffusion. Some of them include long-term memory (fractional Brownian motion) or are based on heavy-tailed distributions (Lévy flights). We have also presented a generalization of these models based on the diffusion equation with derivatives with non-natural (or fractional) orders. The main part of the chapter was dedicated to derivation and properties of Green functions for Double-fractional diffusion equation. We have discussed several representations including Mellin-Barnes integral representation and smearing-kernel representation for $\gamma < 1$. It is also possible to introduce a novel option pricing model based on the double-fractional diffusion which generalizes the Lévy-stable option pricing and introduces risk-redistribution also for the time scale.

The concept of entropy is very important in statistical physics and thermodynamics. By introduction of generalized entropies it is possible to deal with non-extensive sys-

tems and systems with heavy-tailed distributions. The two most popular examples are represented by Rényi entropy and Tsallis entropy. We have shown that there is a possibility of obtaining a completely new class of entropies by combining of axioms of Rényi entropy and Tsallis entropy. This class is called Hybrid entropy. The respective MaxEnt distribution can be expressed in terms of Lambert W-function. Because Lambert W-function is defined only on a subinterval of real numbers, we come to a conclusion that there exist energy regions which remain unoccupied. Therefore, Hybrid entropy has a potential to describe systems with energy gaps.

There are still interesting questions on the issues remaining. In the case of multifractal analysis,there exist many sophisticated models based on multifractal analysis [11]. These models can be a good inspiration for further applications in other fields. It is therefore interesting to compare these models and to find some common properties of them. Similarly, double-fractional diffusion represents a promising model for many applications in description of biological processes or in cosmological models. In the case of generalized entropies, one of possible directions is to generalize the entropy classification [67, 84] to canonical ensembles and/or more generalized form of entropies. Further applications of nonextensive thermodynamics would probably shed some light on sources of nonextensivity.

Apart from the topics discussed in the thesis, there are even more closely related topics that are extremely interesting and worth investigating. Let us mention, for instance, two other applications of Rényi entropy. The first is *Rényi transfer entropy* [95]. Transfer entropy is a model-free measure of information transfer between two time series and can be used in forecasting of evolution in various systems with multiple time series. The second is *Point information entropy* [96], a measure used in image recognition and classification, which is uses multifractal analysis to decode the information hidden in images. All previously mentioned topics represent a potentially fruitful background for new ideas and for applications in new scientific fields.

Appendix A

Basic Properties of Stable Distributions

We summarize basic properties and representations of stable distributions. Sometimes are the distributions called Lévy distributions (or Lévy-stable distributions), after mathematician Paul Lévy, who studied some special examples of stable distributions [97]. Gnedenko and Kolmogorov [12] studied the infinite sums of random variables. In the case of i.i.d. random variables X_n, probability distributions of infinite sums

$$S = \sum_{n=1}^{\infty} a_n X_n - b_n \tag{A.1}$$

belong to a special class of distributions. In the case, when the variance of X_n is finite, the *Central limit theorem* determines that the resulting distribution is Gaussian [98]. When we assume that the variance is not necessarily finite, we obtain the class of limiting distributions. This class has one important property: they are form-invariant under the operation of convolution. The convolution of two probability distributions

$$p(x) = p_1(z) \star p_2(z) = \int_{\mathbb{R}} p_1(z) p_2(x-z) \mathrm{d}z \tag{A.2}$$

is nothing else than the probability distribution of sum of two random variables $X = X_1 + X_2$ with probability distributions p_1, resp p_2. Therefore, the probability is stable, if the convolution of a distribution with itself does not change the functional form of the distribution, i.e.,

$$p(a_1 z + b_1) * p(a_2 z + b_2) = \int_{-\infty}^{\infty} p\left(a_1[x-z] + b_1\right) p(a_1 z + b_1) \mathrm{d}l = p(ax + b). \tag{A.3}$$

The last relation can be nicely reformulated in Fourier transform image. In fact, the convolution is in Fourier image represented as product

$$\mathcal{F}[f * g](k) = \mathcal{F}[f](k) \cdot \mathcal{F}[g](k)\,. \tag{A.4}$$

Thus, Fourier images of stable distributions are invariant under multiplication. From this is possible to determine the functional form of stable distributions $L_{\alpha,\beta}(x)$ as a four-parametric class of distribution with parameters $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\bar{\sigma} \geq 0$, $\bar{x} \in \mathbb{R}$ in the *stable Hamiltonian representation*:

$$\mathcal{H}_{\alpha,\beta;\bar{x},\bar{\sigma}}(p) \equiv \ln\langle e^{ipx}\rangle = i\bar{x}p - \bar{\sigma}^{\alpha}|p|^{\alpha}\left(1 - i\beta\,\mathrm{sign}(p)\omega(p,\alpha)\right), \tag{A.5}$$

$$\omega(p,\alpha) = \begin{cases} \tan(\pi\alpha/2) & \text{if} \quad \alpha \neq 1 \\ \frac{2}{\pi}\ln|p| & \text{if} \quad \alpha = 1. \end{cases} \tag{A.6}$$

As already discussed in Sect. 3.3.2, each parameter has its particular interpretation. Parameters $\bar{x}$ and $\bar{\sigma}$ are location, resp. scale parameters and equal to mean value, resp. standard deviation, if they exist.

Parameter α is called *stability parameter*. It influences the shape of the distribution, the degree of tail decay and also existence of fractional moments. At the end of the section we show that the distribution decays as $1/|x|^{\alpha+1}$, except for extremely asymmetric cases. Parameter β is *asymmetry parameter*, because it influences the skewness of the distribution. From relation

$$L_{\alpha,\beta}(x) = L_{\alpha,-\beta}(-x) \tag{A.7}$$

is obvious that for $\beta = 0$ is the distribution symmetric around the location parameter. In extreme cases, i.e., when $\beta = -1$, the right tail (for $\beta = 1$ the left tail) does not decay polynomially. Namely, for $\alpha > 1$, it decay subexponentially

$$L_{\alpha,-1}(x) \sim \frac{1}{2(\alpha-1)}\left(\frac{x}{\alpha c_{\alpha}}\right)^{\frac{\alpha}{2(\alpha-1)}-1} \exp\left[-(\alpha-1)\left(\frac{x}{\alpha c_{\alpha}}\right)^{-\frac{\alpha}{(\alpha-1)}}\right] \quad \text{for } x \to +\infty\,. \tag{A.8}$$

For $\alpha < 1$, the support of the distribution is bounded to $(-\infty, \bar{x})$ for $\beta = -1$.

There exists an alternative representation of stable Hamiltonian, which is useful in some applications. It can be expressed as

$$\mathcal{H}_{\alpha,\theta;\bar{x},c}(p) \equiv \ln\langle e^{ipx}\rangle = i\bar{x}p - c|p|^{\alpha}e^{i\,\mathrm{sign}(p)\,\theta\,\frac{\pi}{2}}, \tag{A.9}$$

where c and θ are uniquely determined by parameters α, β and $\bar{\sigma}$ [99]. Parameter θ plays analogous role as β and it is bounded by condition $|\theta| \leq \min\{\alpha, 2-\alpha\}$. The region of accessible values in (α,θ)-plane is sometimes called *Feller-Takayasu diamond*. The boundary of the diamond corresponds to $\beta = \pm 1$.

For the purposes of financial applications, it is important to calculate the log-Lévy distribution, which is the distribution of random variable $\exp(X)$, where X is stable random variable. It is equal to two-sided Laplace transform of $L_{\alpha,\beta}$, which exists only for $\beta = 1$. Hence, for $\Re(\lambda) > 0$ (see Ref. [100]) the logarithm of Laplace transform can be expressed as

$$\ln\langle e^{-\lambda x}\rangle = -\lambda\bar{x} - \lambda^{\alpha}\bar{\sigma}^{\alpha}\sec\frac{\pi\alpha}{2}\,. \tag{A.10}$$

Finally, we derive the asymptotic expansion of Lévy distribution. We present the case when $\beta = 0$ because of simplicity, the other cases can be derived analogously. The probability distribution is given by the inverse Fourier transform of its characteristic function

$$L_{\alpha,0}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}p\, e^{-\gamma|p|^\alpha} e^{ikx} = \frac{1}{2\pi}\int_{0}^{\infty} \mathrm{d}p\ e^{-\gamma p^\alpha}\left(e^{ipx} + e^{-ipx}\right)$$
$$= \frac{1}{2\pi}\int_{0}^{\infty} \mathrm{d}p\ e^{-\gamma p^\alpha} 2\Re\left(e^{ipx}\right) = \frac{1}{\pi}\Re\left[\int_{0}^{\infty} \mathrm{d}p\ e^{-\gamma p^\alpha} e^{ipx}\right].$$

We expand the exponential in the integral to the power series, integrate and express the integral in terms of Γ function. We therefore obtain

$$\frac{1}{\pi}\Re\left[\sum_{n=0}^{\infty}\frac{(-\gamma)^n}{n!}\int_0^\infty \mathrm{d}p\, p^{\alpha n} e^{ipx}\right] = \frac{1}{\pi}\Re\left[\sum_{n=0}^{\infty}\frac{(-\gamma)^n}{n!}\frac{\Gamma(\alpha n+1)}{(-ix)^{\alpha n+1}}\right].$$

The real part can be easily expressed with help of the identity

$$\Re\left((\pm i)^{\alpha n+1}\right) = -\sin\left(\frac{\pi\alpha n}{2}\right), \tag{A.11}$$

so we end with the final expansion

$$L_\alpha(x) = -\frac{1}{\pi}\sum_{n=1}^{\infty}\frac{(-\gamma)^n}{n!}\frac{\Gamma(\alpha n+1)}{|x|^{\alpha n+1}}\sin\left(\frac{\pi\alpha n}{2}\right). \tag{A.12}$$

From the previous expansion is also clear that the tails decay as $1/|x|^{\alpha+1}$.

Appendix B

Mellin Transform

Mellin transform is an integral transform useful in many applications in physics, number theory and theory of asymptotic expansions. Mellin transform is also used for the Mellin-Barnes integral representations [101], which can be advantageous, e.g., from computational reasons. More details can be found in Ref. [102].

The Mellin transform is defined as follows:

$$\mathcal{M}[f](z) := \int_0^\infty x^{z-1} f(x) \mathrm{d}x \,. \tag{B.1}$$

Conversely, inverse transform is given by the formula

$$\mathcal{M}^{-1}[f](x) := \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-z} f(z) \mathrm{d}z \,, \tag{B.2}$$

where c is given by the *Mellin inversion theorem* [102].

Mellin transform is closely related to Fourier transform and two-sided Laplace transform:

$$\mathcal{L}[f](z) = \mathcal{M}[f \circ (-\ln x)](z) \tag{B.3}$$
$$\mathcal{F}[f](z) = \mathcal{M}[f \circ (-\ln x)](-iz) \,. \tag{B.4}$$

In other words, Mellin transform can be considered as a multiplicative version of two-sided Laplace transform. As a consequence, the main properties of Mellin transform are

$$\mathcal{M}[f(ax)](s) = a^{-s} F(s) \tag{B.5}$$
$$\mathcal{M}[x^a f(x)](s) = F(s+a) \tag{B.6}$$
$$\mathcal{M}[f(x^a)](s) = |a|^{-1} F(s/a) \tag{B.7}$$
$$\mathcal{M}[\log x^n f(x)](s) = F^{(n)}(s). \tag{B.8}$$

As an example, the Mellin transform of exponential is for $\Re(s) > 0$ equal to Gamma function

$$\mathcal{M}[e^{-x}](s) = \int_0^\infty e^{-x} x^{s-1} \mathrm{d}x = \Gamma(s)\,. \tag{B.9}$$

As a consequence, for $\exp(-x^n)$ we obtain

$$\mathcal{M}[e^{-x^n}](s) = \frac{1}{n}\Gamma\left(\frac{s}{n}\right). \tag{B.10}$$

Interestingly, many functions can be expressed in terms of Gamma function in Mellin image. This is a motivation for introduction of so-called *Mellin-Barnes integrals*. The resulting function is called *H-function*. In the most general form it is defined as

$$H_{p,q}^{m,n}(z) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{\Gamma(\alpha_1 + a_1 s)\ldots\Gamma(\alpha_m + a_m s)}{\Gamma(\beta_1 + b_1 s)\ldots\Gamma(\beta_n + b_n s)} \frac{\Gamma(\gamma_1 - c_1 s)\ldots\Gamma(\gamma_p - c_p s)}{\Gamma(\delta_1 - d_1 s)\ldots\Gamma(\delta_q - d_q s)} z^{-s}\mathrm{d}s\,. \tag{B.11}$$

The poles of $\Gamma(\alpha_i + a_n s)$ are separated from poles of $\Gamma(\beta_i - b_n s)$. The integration is taken between the poles in the common strip of analycity. In this class of integrals are included hypergeometric functions, confluent hypergeometric functions [103] or Mittag-Leffler functions, as shown in Appendix C. More details about H-function can be found e.g., in book [104].

Appendix C

Mittag-Leffler Function

Mittag-Leffler function is a class of special functions discovered by Swedish mathematician G. M. Mittag-Leffler. It is a generalization of several classes of functions. It includes exponentials, hyperbolic functions, trigonometric functions and several other functions. The Mittag-Leffler function is most commonly defined as the infinite series

$$E_{\alpha,\beta}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\beta + \alpha k)} \tag{C.1}$$

for $\alpha, \beta \in \mathbb{C}$, $\Re(\alpha) > 0$, $\Re(\beta) > 0$ and for complex z. Particularly important is the case when $\beta = 1$. In this case we use notation $E_{\alpha,1}(z) = E_\alpha(z)$. The Mittag-Leffler function incorporates several important functions, for example

$$E_0(z) = \sum_{n=0}^{\infty} z^n = \frac{1}{1-z} \tag{C.2}$$

$$E_1(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} = e^z \tag{C.3}$$

$$E_2(z) = \sum_{n=0}^{\infty} \frac{z^n}{(2n)!} = \cosh(\sqrt{z}). \tag{C.4}$$

Moreover, the relation between Mittag-Leffler functions of a double parameter is given by the relation

$$E_{2\alpha}(z^2) = \frac{1}{2}[E_\alpha(z) + E_\alpha(-z)] \tag{C.5}$$

which is a nice generalization of a relation between $\cosh$ function and exponentials. In Ref. [105] are presented even more relations and it is also shown there the relation to hypergeometric functions.

One of the most important properties is the fact that the Mittag-Leffler functions are eigenfunctions of Caputo derivative operators:

$$ {}^{*}\mathcal{D}_{x}^{\nu} E_{\nu}(\lambda x^{\nu}) = \lambda E_{\nu}(\lambda x^{\nu}). \tag{C.6} $$

Another important property of Mittag-Leffler function is its Laplace a Mellin transform. The Laplace transform of Mittag-Leffler function is important theory of integro-differential equations, which is also case of fractional derivative operators [106]. It is possible to show that the transforms are [54]

$$ \mathcal{L}[E_{\alpha,\beta}(\lambda z^{\alpha})](s) = \frac{s^{\alpha-\beta}}{s^{\alpha}-\lambda}, \tag{C.7} $$

$$ \mathcal{M}[E_{\alpha,\beta}(z)](s) = \frac{\Gamma(s)\Gamma(1-s)}{\Gamma(\beta-\alpha s)}. \tag{C.8} $$

The second relation can be used for the Mellin-Barnes intregral representation, as discussed in Appendix B.

Appendix D

Properties of Smearing Kernels

In this appendix we compare smearing kernels of Green functions obtained as a solution of double-fractional diffusion equations for Caputo derivatve and Riesz-Feller derivative when $\gamma < 1$. The Green function is equal to

$$g_{\alpha,\gamma}(x,t) = \int_0^\infty \mathrm{d}l\, \psi_K(t,l) \frac{1}{l^{1/\gamma}} L_{\gamma,1}\left(\frac{t}{l^{1/\gamma}}\right) g_\alpha(x,l)\,, \tag{D.1}$$

where $\psi_K(t,l)$ differs according to derivative

$$\psi_K(t,l) = \begin{cases} \frac{\Gamma(\gamma)}{t^{\gamma-1}} & \text{for Riesz derivative,} \\ \frac{\tau}{l^\gamma} & \text{for Caputo derivative.} \end{cases} \tag{D.2}$$

We are interested in the asymptotic behavior of smearing kernel for small and large values. First, for small values, i.e., when $l \to 0$ and constant t, the argument of the stable distribution goes to infinity. Thus, we can use the asymptotic expansion similar to (A.12)

$$\frac{1}{l^{1/\gamma}} L_{\gamma,1}\left(\frac{t}{l^{1/\gamma}}\right) \sim \frac{\Gamma(\gamma+1)\sin(\pi\gamma)}{\cos\left(\frac{\pi\gamma}{2}\right)} \frac{l}{t^{\gamma+1}} \qquad \text{for } l \to 0\,. \tag{D.3}$$

Hence, For Riesz-Feller derivative is the kernel

$$g_1^{RF}(t,l) \sim \frac{l}{\tau^{2\gamma}} \frac{\Gamma(\gamma)\Gamma(\gamma+1)\sin(\pi\gamma)}{\cos\left(\frac{\pi\gamma}{2}\right)} \qquad \text{for } l \to 0\,. \tag{D.4}$$

On the other hand, for Caputo derivative we obtain a non-zero value for $l = 0$, namely

$$g_1^C(t,0) = \left(\frac{1}{t^\gamma}\right) \frac{\Gamma(\gamma)\sin(\pi\gamma)}{\cos\left(\frac{\pi\gamma}{2}\right)}\,. \tag{D.5}$$

In case of asymptotic expansion for $l \to \infty$, the argument of the stable distribution goes to zero. According to Ref. [107], we have

$$L_{\gamma,1}(x) \sim A_\gamma x^{-1-\frac{\lambda_\gamma}{2}} \exp\left(-B_\gamma x^{-\lambda_\gamma}\right) \quad \text{for } x \to 0^+\,, \tag{D.6}$$

with $\lambda_\gamma = \frac{\gamma}{1-\gamma}$ and γ-dependent constants A_γ, B_γ. Thus, the asymptotic behavior can be described as

$$g_1^{RF}(\tau, l) \sim C^{RF}(\tau) A_\gamma l^{\frac{1}{2(1-\gamma)}} \exp\left(-B_\gamma D(\tau) l^{\frac{1}{1-\gamma}}\right) \qquad \text{for } l \to +\infty, \tag{D.7}$$

respectively

$$g_1^{C}(\tau, l) \sim C^{C}(\tau) A_\gamma l^{\frac{1}{2(1-\gamma)}-1} \exp\left(-B_\gamma D(\tau) l^{\frac{1}{1-\gamma}}\right) \qquad \text{for } l \to +\infty. \tag{D.8}$$

Normalization factors $C^{RF}(\tau)$, resp. $C^{C}(\tau)$ can be determined from previous expressions. Both kernels are depicted in Fig. 3.1.

Appendix E

Derivation of Hybrid Entropy from J.-A. Axioms

In this appendix we present the main steps of derivation of hybrid entropy from J.-A. axioms. The proof was firstly published in [85], together with broad discussion. The proof follows the Khinchin machinery firstly used by himself to derive the functional form of Shannon entropy.

Let us denote $\mathcal{D}(1/r,\ldots,1/r) = \mathcal{L}(r)$. From expansibility axiom and maximality axiom we immediately obtain

$$\mathcal{L}(r) = \mathcal{D}\left(\frac{1}{r},\ldots,\frac{1}{r}\right) = \mathcal{D}\left(\frac{1}{r},\ldots,\frac{1}{r},0\right) \leq \mathcal{D}\left(\frac{1}{r+1},\ldots,\frac{1}{r+1}\right) = \mathcal{L}(n+1)\,, \tag{E.1}$$

i.e., $\mathcal{L}(n)$ is a non-decreasing function of n. By repeated application of the additivity axiom to i.i.d. variables $A^{(m)}$ with uniform distribution $(1/r,\ldots,1/r)$, we obtain that

$$\begin{aligned}\mathcal{D}(A^{(1)}\cup A^{(2)}\cup\ldots\cup A^{(m)}) = \mathcal{L}(r^m) &= \sum_{k=1}^{m}\binom{m}{k}(1-q)^{k-1}\mathcal{D}^k(A^{(k)}) \\ &= \frac{1}{(1-q)}\left[(1+(1-q)\mathcal{L}(r))^m - 1\right]\end{aligned} \tag{E.2}$$

The equation can be extended for $m\in\mathbb{R}^+$. Afterwards, we take the derivative of both sides w.r.t. m and set $m=1$, so

$$\frac{(1-q)\,d\mathcal{L}}{(1+(1-q)\,\mathcal{L})\left[\ln\left(1+(1-q)\,\mathcal{L}\right)\right]} = \frac{dr}{r\ln r}\,. \tag{E.3}$$

The general solution of this equation can be found in the form

$$\mathcal{L}(r) \equiv \mathcal{L}_q(r) = \frac{1}{1-q}\left(r^{c(q)}-1\right)\,, \tag{E.4}$$

where integration constant $c(q)$ has to be determined. Because for $q = 1$, the $\mathcal{L}(r^m)$ is equal to $m\mathcal{L}(r)$, and therefore $c(1) = 0$. Furthermore, monotonicity of $\mathcal{L}(r)$ requires condition $c(q)/(1-q) \geq 0$. It is clear that the functional form of $\mathcal{L}(r)$ corresponds to the form of Tsallis entropy. Indeed, in microcanonical ensemble description gives the hybrid entropy the same description as Tsallis entropy. The difference is in the different definition of conditional entropy, i.e. the canonical ensemble description.

In order to proceed, let us consider a special example of two experiments with outcomes $A = (a_1, \ldots, a_n)$ and distribution $P_A = (p_1, \ldots, p_n)$ and $B = (b_1, \ldots, b_m)$ and distribution $Q_B = (q_1, \ldots, q_m)$. Let us assume that p_k are rational numbers, so $p_k = \frac{g_k}{g}$, where $g = \sum_{k=1}^{n} g_k$. We assume that $m = g$, so the experiment B has g possible outcomes. The dependence of B to A is chosen so that if a_i happens, then all outcomes k-th group b_k happen with equal probability $1/g_k$ and other outcomes have zero probability. Therefore,

$$\mathcal{D}(B|A = a_k) = \mathcal{D}(1/g_k, \ldots, 1/g_k) = \mathcal{L}_q(g_k)\,, \tag{E.5}$$

and the additivity axiom implies that

$$\mathcal{D}(B|A) = f^{-1}\left(\sum_{k=1}^{n} \varrho_k(q) f(\mathcal{L}_q(g_k))\right)\,. \tag{E.6}$$

On the other hand, the entropy for joint experiment $A \cup B$ can be easily determined, because the joint probability distribution is

$$R = \{p_k q_{l\,|\,k}\} \;=\; \{\underbrace{\frac{p_1}{g_1}, \ldots, \frac{p_1}{g_1}}_{g_1\times}, \underbrace{\frac{p_2}{g_2}, \ldots, \frac{p_2}{g_2}}_{g_2\times}, \ldots, \underbrace{\frac{p_n}{g_n}, \ldots, \frac{p_n}{g_n}}_{g_n\times}\} \;=\; \{1/g, \ldots, 1/g\}\,, \tag{E.7}$$

So $\mathcal{D}(A \cup B) = \mathcal{L}_q(g)$. It is now straightforward to compare both representations of joint entropy given by additivity axiom and plug in the form obtained in Eq. (E.4). Consequently, we obtain the functional equation

$$\begin{aligned}\mathcal{D}(A)\left(1 + (1-q) f^{-1}\left(\sum_k \varrho_k(q) f(\mathcal{L}_q(p_k)[1 + (1-q)\mathcal{L}_q(g)] + \mathcal{L}_q(g))\right)\right)& \\ = \;\mathcal{L}_q(g) - f^{-1}\left(\sum_k \varrho_k(q) f(\mathcal{L}_q(p_k)[1 + (1-q)\mathcal{L}_q(g)] + \mathcal{L}_q(g))\right)\,.&\end{aligned} \tag{E.8}$$

Defining $f_{(\alpha,\beta)}(x) = f(-\alpha x + \beta)$ and $a = [1 + (1-q)\mathcal{L}_q(g)]$, we get

$$\mathcal{D}(A) = \frac{f^{-1}_{(a,\mathcal{L}(g))}\left(\sum_k \varrho_k(q) f_{(a,\mathcal{L}_q(g))}(-\mathcal{L}(p_k))\right)}{1-(1-q)f^{-1}_{(a,\mathcal{L}(g))}\left(\sum_k \varrho_k(q) f_{(a,\mathcal{L}_q(g))}(-\mathcal{L}(p_k))\right)}. \tag{E.9}$$

When we reformulate the entropy in terms $\mathcal{L}(1/p_k)$, which represents an elementary information of a_k

$$\mathcal{L}_q(p_k) = -\frac{\mathcal{L}_q(1/p_k)}{1+(1-q)\mathcal{L}_q(1/p_k)}\,. \tag{E.10}$$

When we define

$$g(x) = f_{(a,\mathcal{L}(g))}\left(\frac{x}{1+(1-q)x}\right)\,, \tag{E.11}$$

the entropy can be written as

$$\mathcal{D}(A) = g^{-1}\left(\sum_k \varrho_k(q) g(\mathcal{L}_q\left(1/p_k\right))\right)\,. \tag{E.12}$$

Moreover, if we set in the definition of conditional entropy $A = B$, then we get

$$\mathcal{D}(A) = f^{-1}\left(\sum_k \varrho_k(q) f(\mathcal{L}_q\left(1/p_k\right))\right)\,. \tag{E.13}$$

. Because the left-hand sides are the same, so have to the right-hand sides. According to [75], the two quasi-linear means are the same if and only if their Kolmogorov-Nagumo functions are linearly related

$$g(x) = f\left(\frac{-x+y}{1+(1-q)x}\right) = \theta_q(y) f(x) \;+\; \vartheta_q(y)\,. \tag{E.14}$$

By defining $\varphi(x) = f(x) - f(0)$, we end with

$$\varphi\left(\frac{-x+y}{1+(1-q)x}\right) = \theta_q(y)\varphi(x) \;+\; \varphi(y). \tag{E.15}$$

In Ref. [24] is shown that the only non-trivial class of solutions is

$$\varphi(x) = \frac{1}{\alpha}\ln\left[1+(1-q)x\right]\,. \tag{E.16}$$

α is a free parameter. When inserted back into E.13, α is canceled and we end with

$$\mathcal{D}_q(A) \;=\; \frac{1}{1-q}\,\left(e^{-c(q)\sum_k \varrho_k(q)\ln p_k} - 1\right) = \frac{1}{1-q}\left(\prod_k (p_k)^{-c(q)\varrho_k(q)} - 1\right) \tag{E.17}$$

From additivity axiom we finally obtain that $c(q) = 1-q$. As discussed before, we have used the maximality axiom only in certain cases, i.e. for the function $\mathcal{L}(r)$, and it is necessary to verify additionally the validity of maximality axiom for each q.

Appendix F

Lambert W-function

Lambert W-function is defined as the complex inverse of ze^z. It has been firstly defined by Lambert in eighteenth century. Since that time it has found many applications in pure mathematics, hydrodynamics, quantum theory and many other fields. The Lambert W-function has many interesting properties in both real and complex domain and we discuss some of them in the next lines. The Lambert W-function $W(z)$ is defined from equation

$$z = W(z)e^{W(z)} \qquad \text{for } z \in \mathbb{C}\,. \tag{F.1}$$

In the complex plane has the previous equation an infinite number of solutions $W_k(z)$ for every $z \neq 0$. Nevertheless, for real arguments we observe only two branches of real solutions. From the theory of branch cuts (more details can be found e.g. in Ref. [108]) we have the principal cut $W_0(x)$, which exists on the interval $[-1/e, \infty)$ and the branch cut $W_{-1}(x)$, which exists on the interval $[-1/e, 0)$. The two real branches are depicted in Fig. F.1. It is easy to show that many equations combining logarithmic functions and polynomicals can be solved in terms of Lambert W-function. The solution of equation

$$\ln z + bz^c = d \tag{F.2}$$

can be expressed as

$$z = \frac{1}{bc}\, W\left(b\, c\, e^{dc}\right)^{1/c}\,. \tag{F.3}$$

Now we turn our attention to asymptotic expansions of the Lambert W-function. First, we are interested in the Taylor series of $W_0(x)$around $x_0 = 0$. This can be obtained in the form

$$W(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} n^{n-2}}{(n-1)!} x^n\,. \tag{F.4}$$

The radius of convergence is $1/e$. When we are interested in the linear expansion, i.e. very close to zero, we ge that $W(x) \approx x$. On the other hand, for $x \to \infty$ we get that

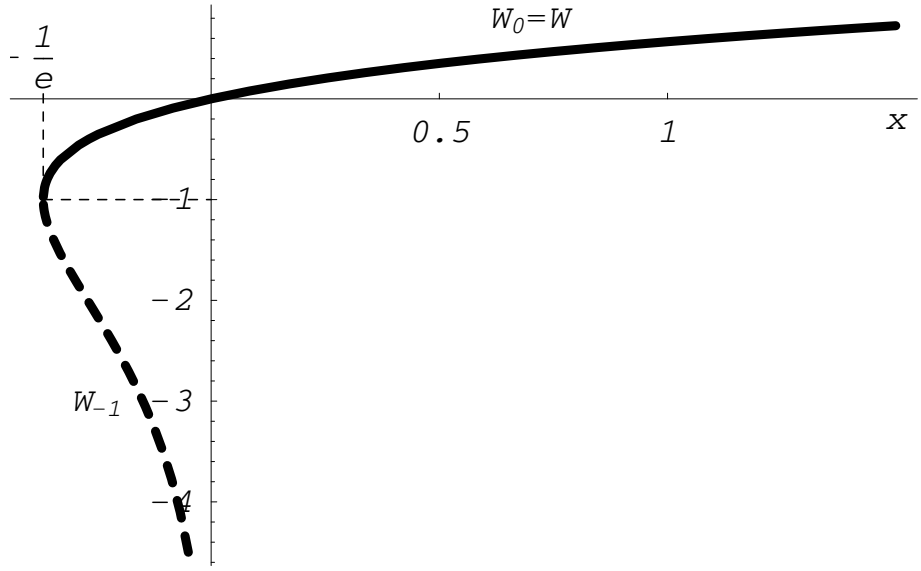


Figure F.1: Two real branches of Lambert W-function

$W_0(x)$ can be approximated by [108]

$$W(x) \approx \ln x - \ln \ln x + o(1) \qquad \text{for } x \to \infty. \tag{F.5}$$

In case of branch $W_{-1}(x)$, we are interested in behavior asymptotic expansion close to zero, because

$$\lim_{x \to 0^-} W_{-1}(x) = -\infty\,. \tag{F.6}$$

The expansion is functionally quite similar to asymptotic expansion of the principal branch

$$W_{-1}(x) \approx \ln(-x) - \ln(-\ln(-x)) + o(1) \qquad \text{for } x \to 0^-. \tag{F.7}$$

List of Author's Publications

Articles

P. Jizba and J. Korbel. Multifractal diffusion entropy analysis: Optimal bin width of probability histograms. *Physica A*, 413:438 – 458, 2014
P. Jizba and J. Korbel. Techniques for multifractal spectrum estimation in financial time series. *International Journal of Design & Nature and Ecodynamics*, 10(3):261–266, 2015
P. Jizba and J. Korbel. On q-non-extensive Statistics with Non-Tsallisian Entropy. *Physica A*, 444:808–827, 2016

Proceedings

P. Jizba and J. Korbel. Methods and techniques for multifractal spectrum estimation in financial time series. In *Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013), Mataró (Barcelona), Spain 25 - 28 June 2013*, 2014
P. Jizba and J. Korbel. Modeling financial time series: Multifractal cascades and rényi entropy. In *ISCS 2013: Interdisciplinary Symposium on Complex Systems*, pages 227–236. Springer Berlin Heidelberg, 2014
P. Jizba and J. Korbel. Applications of multifractal diffusion entropy analysis to daily and intraday financial time series. In *ISCS 2014: Interdisciplinary Symposium on Complex Systems*, pages 333–342. Springer International Publishing, 2015

Publications in submission process

H. Kleinert and J. Korbel. Option Pricing Beyond Black-Scholes Based on Double-Fractional Diffusion. *ArXiv*, 2015. (`http://arxiv.org/abs/1503.05655`), submitted to Phys. A
R. Rychtáriková, J. Korbel, P Macháček, P. Císař, J. Urban, D. Soloviov, and D. Štys. Point Information Gain, Point Information Gain Entropy and Point Information Gain Entropy Density as Measures of Semantic and Syntactic Information of Multidimensional Discrete Phenomena. *ArXiv*, 2015. (`http://arxiv.org/abs/1501.02891`), submitted to IEEE Inf. Trans. Theor.

Bibliography

[1] W. Paul and J. Baschangel. *Stochastic Processes: From Physics to Finance*. Springer, Berlin, 1999.

[2] J.-P. Bouchard and M. Potters. *Theory of Financial Risks: From Statistical Physics to Risk Management*. CUP, Cambridge, 2001.

[3] A. N. Kolmogorov. The local structure of turbulence in incompressible viscous fluids at very large reynolds numbers. In *Doklady Akademii Nauk SSSR*, volume 32, pages 16–18, 1941.

[4] S. Abe and Y. Okamoto. *Nonextensive Statistical Mechanics and Its Applications*. Springer, 2001.

[5] H. E. Hurst. Long-term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–808, 1951.

[6] H. G. E. Hentschel and I. Procaccia. The infinite number of generalized dimensions of fractals and strange attractors. *Physica D*, 8(3):435 – 444, 1983.

[7] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman. Fractal measures and their singularities: The characterization of strange sets. *Physical Review A*, 33:1141–1151, Feb 1986.

[8] B. B. Mandelbrot and J. W. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968.

[9] B. B. Mandelbrot. Multifractal measures, especially for the geophysicist. *Pure and Applied Geophysics*, 131(1-2):5–42, 1989.

[10] H. E. Stanley and P. Meakin. Multifractal phenomena in physics and chemistry. *Nature*, 335(29):405, 1988.

[11] L.E. Calvet and A.J. Fisher. *Multifractal Volatility: Theory, Forecasting, and Pricing*. Academic Press Advanced Finance. Elsevier Science, 2008.

[12] B.V. Gnedenko and A.N. Kolmogorov. *Limit Distributions for Sums of Intepen-dent Random Variables*. Adison-Wesley, 1968.

[13] N.N. Taleb. *The Black Swan: The Impact of the Highly Improbable Fragility*. Random House Publishing Group, 2010.

[14] H. Kleinert. Quantum field theory of black-swan events. *Foundations of Physics*, pages 1–11, 2013. (http://klnrt.de/409).

[15] P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC Financial Mathematics Series. Taylor & Francis, 2003.

[16] H. Haken. *Synergetics: an Introduction : Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology*. Springer-Verlag, 1977.

[17] A. Rényi. *Probability Theory*. North-Holland Series in Applied Mathematics and Mechanics. Elsevier, 1970.

[18] J. Havrda and F. Charvát. Quantification method of classification processes. con-cept of structural a-entropy. *Kybernetika*, 03(1):30–35, 1967.

[19] C. Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of Sta-tistical Physics*, 52(1-2):479–487, 1988.

[20] H. Kleinert. *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics and Financial Markets, 4th edn*. World Scientific, 2009.

[21] P. Jizba and J. Korbel. Multifractal diffusion entropy analysis: Optimal bin width of probability histograms. *Physica A*, 413:438 – 458, 2014.

[22] P. Jizba and J. Korbel. Techniques for multifractal spectrum estimation in finan-cial time series. *International Journal of Design & Nature and Ecodynamics*, 10(3):261–266, 2015.

[23] H. Kleinert and J. Korbel. Option Pricing Beyond Black-Scholes Based on Double-Fractional Diffusion. *ArXiv*, 2015. (http://arxiv.org/abs/1503.05655).

[24] P. Jizba and J. Korbel. On q-non-extensive Statistics with Non-Tsallisian Entropy. *Physica A*, 444:808–827, 2016.

[25] D. Harte. *Multifractals: Theory and Application*. Chapmann & Hall/CRC, 2001.

[26] B.B. Mandelbrot, M.L. Lapidus, and M. Van Frankenhuysen. *Fractal Geometry and Applications: Multifractals, Probability and Statistical Mechanics, Applica-tions*. American Mathematical Society, 2004.

[27] R. N. Mantegna and H. E. Stanley. *An Introduction to Econophysics*. CUP, Cambridge, 2000.

[28] V.I. Arnol'd. *Mathematical Methods of Classical Mechanics*. Graduate Texts in Mathematics. Springer, 1989.

[29] J. F. Muzy, E. Bacry, and A. Arneodo. Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method. *Physical Review E*, 47(2):875, 1993.

[30] H. E. Hurst, R. P. Black, and Y. M. Simaika. *Long-term Storage : An Experimental Study*. Constable, London, 1965.

[31] R. Morales, T. Di Matteo, R. Gramatica, and T. Aste. Dynamical generalized hurst exponent as a tool to monitor unstable periods in financial time series. *Physica A*, 391(11):3180–3189, 2012.

[32] M. Ausloos. Generalized hurst exponent and multifractal function of original and translated texts mapped into frequency and length time series. *Physical Review E*, 86:031108, Sep 2012.

[33] J. W. Kantelhardt, E. Koscielny-Bunde, H. Rego, S. Havlin, and A. Bunde. Detecting long-range correlations with detrended fluctuation analysis. *Physica A*, 295(3):441–454, 2001.

[34] W. Kantelhardt, S. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H.E. Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, 316:87 – 114, 2002.

[35] Z. Yu, L. Yee, and Y. Zu-Guo. Relationships of exponents in multifractal detrended fluctuation analysis and conventional multifractal analysis. *Chinese Physics B*, page 090507, 2011.

[36] D. Schertzer, S. Lovejoy, F. Schmitt, Y. Chigirinskaya, and D. Marsan. Multifractal cascade dynamics and turbulent intermittency. *Fractals*, 5(03):427–471, 1997.

[37] N. Scafetta and P. Grigolini. Scaling detection in time series: Diffusion entropy analysis. *Physical Review E*, 66(3):036130, 2002.

[38] J. Huang, P. Shang, and X. Zhao. Multifractal diffusion entropy analysis on stock volatility in financial markets. *Physica A*, 391(22):5739 – 5745, 2012.

[39] E. Koscielny-Bunde, J.W. Kantelhardt, P. Braun, A. Bunde, and S. Havlin. Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies. *Journal of Hydrology*, 322(1–4):120 – 137, 2006.

[40] T. Van Erven and P. Harremoës. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

[41] S. Kullback and R. A. Leibler. On Information and Sufficiency. *Annals of Mathematical Statistics*, 22:49–86, 1951.

[42] I. Csiszár. Information-type distance measures and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2(1):299, 1967.

[43] H. Sturges. The choice of a class-interval. *Jornal of American Statistical Association*, 21:65, 1926.

[44] D. W. Scott. *Multivariate Density Estimation: Theory, Practice and Visualisation*. John Willey and Sons, Inc., 1992.

[45] D. Freedman and P. Diaconis. On the histogram as a density estimator. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(4):453, 1981.

[46] P. Meakin. Diffusion-limited aggregation on multifractal lattices: A model for fluid-fluid displacement in porous media. *Physical Review A*, 36(6):2833, 1987.

[47] Y. Tessier, S. Lovejoy, and D. Schertzer. Universal multifractals: Theory and observations for rain and clouds. *Journal of Applied Meteorology*, 32:223–223, 1993.

[48] B.B. Mandelbrot. Fractal financial fluctuations; do they threaten sustainability? In *Science for Survival and Sustainable Development*. Pontificia Academia Scientiarum, 1999.

[49] C. Beck and F. Schlögl. *Thermodynamics of Chaotic Systems: An Introduction*. Cambridge University Press, 1995.

[50] P. Jizba and T. Arimitsu. The World according to Rényi: Thermodynamics of Multifractal Systems. *Annals of Physics*, 312(1):17, 2004.

[51] A. Plastino and A.R. Plastino. On the universality of thermodynamics' legendre transform structure. *Physics Letters A*, 226(5):257 – 263, 1997.

[52] I. Podlubny. *Fractional Differential Equations, Volume 198: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their ... (Mathematics in Science and Engineering)*. Academic Press, first edition, 1998.

[53] S.S.G. Samko, A.A.A. Kilbas, and O.O.I. Marichev. *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach Science Publishers, 1993.

[54] K. Diethelm. *The Analysis of Fractional Differential Equations an Application-Oriented Exposition Using Differential Operators of Caputo Type*. Springer-Verlag, Heidelberg New York, 2010.

[55] T. Abdeljawad. On riemann and caputo fractional differences. *Computers & Mathematics with Applications*, 62(3):1602–1611, August 2011.

[56] F. C. Klebaner et al. *Introduction to Stochastic Calculus with Applications*, volume 57. World Scientific, 2005.

[57] N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland Personal Library. Elsevier Science, 2011.

[58] G. Samoradnitsky and S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Stochastic Modeling Series. Taylor & Francis, 1994.

[59] R. Gorenflo and F. Mainardi. Fractional diffusion processes: Probability distribution and continuous time random walk. In *Lecture Notes in Physics 621*, pages 148–166, 2007.

[60] Ch. Beck and E.G.D. Cohen. Superstatistics. *Physica A*, 322:267–275, 2003.

[61] H. Kleinert and V. Zatloukal. Green function of the double-fractional fokker-planck equation: Path integral and stochastic differential equations. *Physical Review E*, 88:052106, Nov 2013. (`http://klnrt.de/407`).

[62] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal, The*, 27(4):623–656, 1948.

[63] A. Feinstein. On the coding theorem and its converse for finite-memory channels. *Information and Control*, 2(1):25 – 44, 1959.

[64] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106:620–630, 1957.

[65] A.I.A. Khinchin. *Mathematical Foundations of Information Theory*. Dover Books on Mathematics. Dover Publications, 1957.

[66] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. Number 11 in Illini books. University of Illinois Press, 1949.

[67] R. Hanel and S. Thurner. A comprehensive classification of complex statistical systems and an ab initio derivation ofttheir entropy and distribution functions. *Europhysics Letters*, 93:20006, 2011.

[68] P. Tempesta. Group entropies, correlation laws, and zeta functions. *Physical Review E*, 84(2):021121, 2011.

[69] T.S. Biró and G.G. Barnaf˙ New entropy formula with fluctuating reservoir. *Physica A*, 417:215 – 220, 2015.

[70] C. Tsallis, M. Gell-Mann, and Y. Sato. Asymptotically scale-invariant occupancy of phase space makes the entropy s_q extensive. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15377–15382, 2005.

[71] A. Plastino and A. R. Plastino. Tsallis Entropy and Jaynes' Information Theory Formalism. *Brazilian Journal of Physics*, 29:50 – 60, 03 1999.

[72] R. Balescu. Equilibrium and nonequilibrium statistical mechanics. *NASA STI/Recon Technical Report A*, 76:32809, May 1975.

[73] S. Wang, T. Zhang, and B. Xi. Schur convexity for a class of symmetric functions. In Chunfeng Liu, Jincai Chang, and Aimin Yang, editors, *Information Computing and Applications*, volume 243 of *Communications in Computer and Information Science*, pages 626–634. 2011.

[74] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer Series in Statistics. Springer, New York, 2011.

[75] A. Rényi. *Selected Papers of Alfréd Rényi*. Number 2 in Selected Papers of Alfréd Rényi. Akadémiai Kiadó, 1976.

[76] E. Arikan. An inequality on guessing and its application to sequential decoding. *IEEE Transactions on Information Theory*, 42(1):99–105, 1996.

[77] A. Teixeira, A. Matos, and L. Antunes. Conditional rényi entropies. *IEEE Transactions on Information Theory*, 58(7):4273–4277, July 2012.

[78] Ch. Cachin. *Entropy Measures and Unconditional Security in Cryptography*. PhD thesis, Swiss Federal Institute of Technology Zurich, 1997.

[79] P. Jizba and T. Arimitsu. Observability of rényi's entropy. *Physical Review E*, 69:026128, Feb 2004.

[80] L.L. Campbell. A coding theorem and rényi's entropy. *Information and Control*, 8(4):423 – 429, 1965.

[81] S. Umarov, C. Tsallis, and S. Steinberg. On a q-central limit theorem consistent with nonextensive statistical mechanics. *Milan Journal of Mathematics*, 76(1):307–328, 2008.

[82] F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412, 2006.

[83] S. Martínez, F. Nicolás, F. Pennini, and A. Plastino. Tsallis' entropy maximization procedure revisited. *Physica A*, 286(3–4):489 – 502, 2000.

[84] R. Hanel and S. Thurner. When do generalized entropies apply? how phase space volume determines entropy. *Europhysics Letters*, 96(5):50003, 2011.

[85] P. Jizba and T. Arimitsu. Generalized statistics: Yet another generalization. *Physica A*, 340(1):110–116, 2004.

[86] H.-N. Shi, Y.-M. Jiang, and W.-D. Jiang. Schur-convexity and schur-geometrically concavity of gini means. *Computers & Mathematics with Applications*, 57(2):266 – 274, 2009.

[87] P. Billingsley. *Ergodic Theory and Information*. Wiley Series in Probability and Mathematical Statistics. J. Wiley, 1965.

[88] M. Davis and A. Etheridge. *Louis Bachelier's Theory of Speculation: The Origins of Modern Finance*. Princeton University Press,, 2006.

[89] F. Black and M. S. Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–54, May-June 1973.

[90] P. Carr and L. Wu. Finite moment log stable process and option pricing. *Journal of Finance*, 58(2):753–777, 2003.

[91] H. Kleinert. Option pricing from path integral for non-gaussian fluctuations. natural martingale and application to truncated lévy distributions. *Physica A*, 312(1–2):217–242, 2002. (`http://arxiv.org/abs/cond-mat/0202311`).

[92] W. Wyss. The fractional black-scholes equation. *Fractional Calculus & Applied Analysis*, 3(51), 2000.

[93] C. Necula. Option pricing in a fractional brownian motion environment. Advances in Economic and Financial Research - DOFIN Working Paper Series 2, Bucharest University of Economics, Center for Advanced Research in Finance and Banking - CARFIB, January 2008.

[94] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.

[95] P. Jizba, H. Kleinert, and M. Shefaat. Rényi's information transfer between financial time series. *Physica A*, 391(10):2971 – 2989, 2012.

[96] R. Rychtáriková, J. Korbel, P Macháček, P. Císař, J. Urban, D. Soloviov, and D. Štys. Point Information Gain, Point Information Gain Entropy and Point Information Gain Entropy Density as Measures of Semantic and Syntactic Information of Multidimensional Discrete Phenomena. *ArXiv*, 2015. (`http://arxiv.org/abs/1501.02891`).

[97] B. B. Mandelbrot. The pareto-lévy law and the distribution of income. *International Economic Review*, 1(2):79–106, 1960.

[98] William F. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, 1968.

[99] K.-I. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.

[100] V.M. Zolotarev. *One-dimensional Stable Distributions*. Translations of Mathematical Monographs. American Mathematical Society, 1986.

[101] R.B. Paris and D. Kaminski. *Asymptotics and Mellin-Barnes Integrals*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2001.

[102] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.

[103] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products, Fifth Edition*. Academic Press, 5th edition, 1994.

[104] A.M. Mathai, R.K. Saxena, and H.J. Haubold. *The H-Function: Theory and Applications*. Springer New York, 2009.

[105] H. J. Haubold, A.M Mathai, and R. K. Saxena. Mittag-leffler functions and their applications. *Journal of Applied Mathematics*, 2011, 2011.

[106] F. Mainardi and R. Gorenflo. On mittag-leffler-type functions in fractional evolution processes. *Journal of Computational and Applied Mathematics*, 118(1–2):283 – 299, 2000.

[107] A.V. Skorohod. Asymptotic formulas for stable distribution laws. *Selected Translations in Mathematical Statistics and Probability*, 1:157–161, 1961.

[108] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the lambert w function. In *Advances in Computational Mathematics*, pages 329–359, 1996.

[109] P. Jizba and J. Korbel. Methods and techniques for multifractal spectrum estimation in financial time series. In *Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013), Mataró (Barcelona), Spain 25 - 28 June 2013*, 2014.

[110] P. Jizba and J. Korbel. Modeling financial time series: Multifractal cascades and rényi entropy. In *ISCS 2013: Interdisciplinary Symposium on Complex Systems*, pages 227–236. Springer Berlin Heidelberg, 2014.

[111] P. Jizba and J. Korbel. Applications of multifractal diffusion entropy analysis to daily and intraday financial time series. In *ISCS 2014: Interdisciplinary Symposium on Complex Systems*, pages 333–342. Springer International Publishing, 2015.