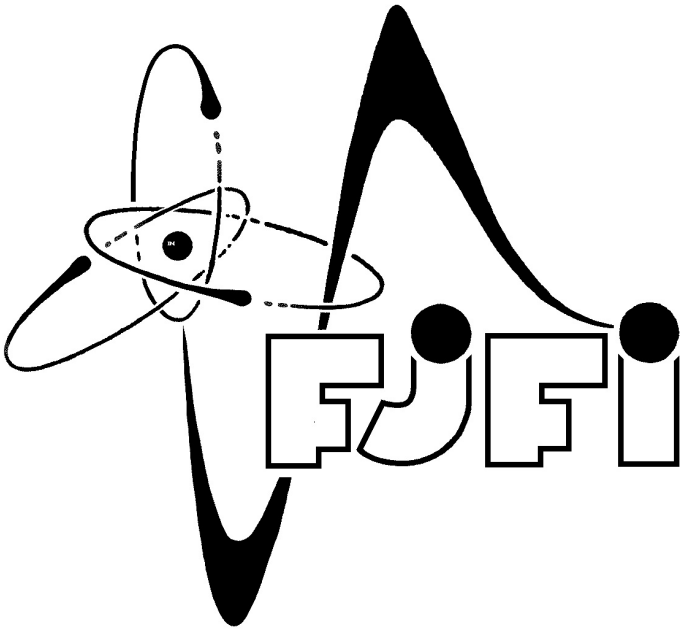
CZECH TECHNICAL UNIVERSITY IN PRAGUE
Faculty of Nuclear Sciences and Physical Engineering




# DIPLOMA THESIS

## Applications of Multifractals in Financial Markets

Author: Bc. Jan Korbel
Supervisor: Ing. Petr Jizba, PhD.
Year: 2012

## Acknowledgement

I would like to thank to my supervisor Ing. Petr Jizba PhD., for all consultations, possibilities to study the topic of econophysics from many different points of view and particulary for all valuable advices during my studies. I would also like to thank to Karolína Somrová for careful reading of my thesis and for corrections.

## Poděkování

Rád bych poděkoval mojí rodině a přátelům za podporu během mého celého studia.

# Prohlášení

Prohlašuji, že jsem svou diplomovou práci vypracoval samostatně a použil jsem pouze literaturu uvedenou v přiloženém seznamu.

Nemám závažný důvod proti užití školního díla ve smyslu §60 Zákona č. 212/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

# Declaration

I declare, I wrote my diploma thesis independently and exclusively with the use of cited bibliography.

I agree with the usage of this thesis in the purport of the §60 Act 121/2000 (Copyright Act).

V Praze dne ..... .....

*Název práce:* **Použití multifraktálů ve finančních trzích**
*Autor:* Bc. Jan Korbel
*Obor:* Matematické inženýrství
*Druh práce:* Diplomová práce
*Vedoucí práce:* Ing. Petr Jizba, PhD.
Katedra fyziky, Fakulta jaderná a fyzikálně inženýrská

**Abstrakt**

Nový vědecký obor, který spojuje fyzikální metody a finanční trhy, se jmenuje ekonofyzika. Tato práce ukazuje možné aplikace matematických a fyzikálních metod ve finančních trzích a diskutuje hlavní omezení nejznámějších modelů používaných pro analýzu a předpovědi na finančních trzích. Především se zabývá přítomností paměti a velkých změn v časových řadách, posléze je zaveden koncept multifraktálních deformací, který je přirozenou a elegantní cestou k popsání komplexního chování cen. Nakonec jsou prezentovány praktické výsledky z reálných trhů, které podporují naše teoretické argumenty.

*Klíčová slova:* multifraktály, finanční trhy, multifraktální časové deformace, stochastické processy, autokorelace, škálování

*Title:* **Applications of Multifractals in Financial Markets**
*Author:* Jan Korbel

**Abstract**

A new scientific branch that connects physical methods and financial markets is called econophysics. This work shows some possible applications of mathematical and physical concepts in financial markets and discusses main limitations of the most popular model used for analysis and forecasts on financial markets. Foremost are discussed phenomena of memory and large jumps in time series and it is introduced a concept of multifractal deformations that is a natural and elegant way, how to describe complex behavior of asset prices. At the end, we show some practical results from real markets that support our theoretical arguments.

*Keywords:* multifractals, financial markets, multifractal time deformations, stochastic processes, autocorrelation, scaling

# Contents

**List of Symbols** **vi**

**List of Figures and Tables** **vii**

**1 Introduction** **1**

**2 Physical Motivation: Langevin Equation** **3**
2.1 Langevin Equation: Introduction . . . . . 3
2.2 Overdamped Langevin Equation . . . . . 6
2.3 Smoluchovski Equation and Diffusion Equation . . . . . 6
2.4 General Langevin Equation and Klein-Kramers Equation . . . . 9

**3 Mathematical Background: Stochastic Processes** **11**
3.1 Basic notes . . . . . 11
3.2 Gaussian Processes . . . . . 11
3.2.1 Wiener Process . . . . . 12
3.2.2 Ornstein-Uhlenbeck Process . . . . . 14
3.3 Martingales . . . . . 15
3.4 Markov Processes . . . . . 16
3.4.1 Chapman-Kolmogorov Equation . . . . . 17
3.4.2 Markov Chains . . . . . 17
3.4.3 Master Equation . . . . . 18
3.4.4 Fokker-Planck Equation . . . . . 19
3.5 Stochastic Calculus . . . . . 21
3.5.1 Stochastic Integration . . . . . 22
3.5.2 Itō lemma . . . . . 23
3.5.3 Itō Integration versus Stratonovich Integration . . . . . 25
3.5.4 Geometric Brownian Motion . . . . . 27

**4 Memory in Physical and Economical Processes** **28**
4.1 Correlations and Stationarity . . . . . 28
4.2 Power Spectrum . . . . . 30
4.2.1 Wiener-Kchinchin Relation . . . . . 30
4.2.2 1/f Noise . . . . . 31
4.3 Fractional Stochastic Processes . . . . . 32
4.3.1 Fractional Integration and Differentiation . . . . . 32
4.3.2 Mittag-Leffler Functions . . . . . 34
4.3.3 Langevin Equation with Memory . . . . . 35

4.4 Fractional Brownian Motion . . . . . 37

**5 Scaling Properties of Random Processes** **39**
5.1 Scaling and Memory . . . . . 39
5.2 A Way beyond Central Limit Theorem . . . . . 40
5.2.1 Lévy Flights . . . . . 40
5.2.2 Fractional Diffusion Equation . . . . . 42
5.2.3 Truncated Lévy Distributions . . . . . 43
5.3 Fractals and Self-similarity . . . . . 44
5.4 Fractal Properties of Stochastic Processes . . . . . 47
5.5 Hurst Exponent . . . . . 48
5.5.1 Rescaled Range Analysis . . . . . 48
5.6 Detrended Fluctuation Analysis . . . . . 49

**6 Multifractal Stochastic Processes** **50**
6.1 Multifractal Formalism . . . . . 50
6.2 Trading Time and Clock Time . . . . . 52
6.3 Modeling of Time Deformations . . . . . 53
6.4 Multifractal Measures . . . . . 54
6.4.1 Binomial Measure . . . . . 54
6.4.2 Multiplicative Measure . . . . . 55

**7 Models of Price Evolution** **59**
7.1 Classical Models . . . . . 59
7.1.1 ARMA . . . . . 59
7.1.2 ARCH, GARCH . . . . . 61
7.1.3 Regime Switching Models . . . . . 62
7.2 Multifractal Models . . . . . 63
7.2.1 Multifractal Model of Asset Returns . . . . . 63
7.2.2 Markov Switching Multifractal . . . . . 64
7.2.3 Continuous Time MSM . . . . . 64

**8 Real Financial Markets: Memory, Jumps** **66**

**9 Conclusion** **71**

**A Stable distributions** **73**

# List of Symbols

| | |
|---|---|
| $\#$ | number of elements |
| $a.s.$ | almost surely |
| $c.d.f.$ | cumulative distribution function |
| $\mathrm{Cov}(x,y) = \mathrm{E}(xy) - \mathrm{E}(x)\mathrm{E}(y)$ | covariance |
| $\stackrel{d}{=}$ | equal in distribution |
| $\dim_B(F)$ | Box counting dimension |
| $\dim_H(F)$ | Hausdorff dimension |
| $\mathrm{E}(.)$ | mean, expectation value |
| $\eta(t)$ | white noise |
| $\mathcal{F}[f](k)$ | Fourier transform of f |
| $\mathcal{F}^{-1}[f](k)$ | inverse Fourier transform of f |
| $f(\alpha)$ | multifractal spectrum |
| $\lfloor x \rfloor = \max\{z \in \mathbb{Z} \vert z \leq x\}$ | floor function |
| $\mathcal{H}^s(F)$ | Hausdorff measure |
| $i.i.d.$ | independent, identically distributed |
| $\mathcal{L}[f](s)$ | Laplace transform of f |
| $L_\alpha(t)$ | Lévy $\alpha$-stable process |
| $\underset{x \to x_0}{l.i.m.}$ | mean-square limit |
| $\mathcal{N}(\mu, \sigma^2)$ | normal distribution with mean $\mu$ and variance $\sigma^2$ |
| $\stackrel{p.v.}{=}$ | equal in principal value |
| $\tau(q)$ | scaling function |
| $\mathrm{Var}(x) = \mathrm{Cov}(x,x)$ | variance |
| $Y(t)$ | Ornstein-Uhlenbeck process |
| $W(t)$ | Wiener process |
| $W_H(t)$ | Fractional Brownian Motion |

# List of Figures and Tables

## List of Figures


1.1 Index S&P 500, its traded volume and volatility in the year 2008 2

2.1 White noise . . . . . 8

3.1 Wiener process . . . . . 13
3.2 Geometric Brownian motion . . . . . 27

4.1 Power spectrum of White noise and Wiener process . . . . . 31
4.2 Fractional Brownian motion . . . . . 37

5.1 Lévy distribution and its properties . . . . . 41
5.2 Wiener process and Lévy flight in 2D . . . . . 42
5.3 Koch curve and Random Koch curve . . . . . 47

6.1 Wiener pattern . . . . . 54
6.2 Multifractal pattern . . . . . 55
6.3 Construction of time deformation . . . . . 56
6.4 Comparison of Geometric Brownian motion and GBM with multifractal deformation . . . . . 57
6.5 Hurst exponent and multifractal spectrum of time deformation . 58

7.1 Simulation of ARMA(2,1) . . . . . 60
7.2 Simulation of GARCH(1,1) . . . . . 62

8.1 Value of S&P 500 index in period 1985-2010 . . . . . 67
8.2 Daily returns of S&P 500 in period 1985-2010 . . . . . 67
8.3 Autocorrelation function of daily return and absolute value of daily returns, distribution of daily returns and power spectrum of absolute value of daily returns for S&P500 . . . . . 70


## List of Tables


8.1 Estimated parameters of Lévy distribution for some indices . . . 66
8.2 Estimated parameters of Lévy distribution and estimated Hurst exponent for S&P 500 calculated annually for period 1985-2010 . 68

# Chapter 1

# Introduction

Financial markets is a field that is for many decades studied by economists, analysts, speculators and many other people, but quite recently also natural scientists and especially physicists have begun to be interested in the markets and created a new scientific field - *econophysics*, which is an application of physical methods to financial markets. The reason was that models that had been developed before, were often too simple, or worked with idealized assumptions that were not completely fulfilled in real situations, so it was very difficult to analyze the market and to make more precise predictions. Many models used by practitioners from finance are based on simple ideas of coin tossing and random walks. That does not mean, that these models are bad, but they have their own limitations that can be in some situations crucial to right prediction and forecast of future evolution. They are usually based on the assumption of random walk and its continuous limit called the Brownian motion. The problem of this model is that because of Central limit theorem (theorem A.1 in appendix) is the distribution of returns similar to Gaussian distribution. The disadvantage of Gaussian distribution lies in the behavior for large values. We know that Gaussian distribution has the form

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{1.1}$$

which means that for large values the probability decays extremely rapidly. The probability, that we observe a value larger than two $\sigma$, is smaller than 5%; for $6\sigma$ we get a probability similar to $10^{-6}$%. Nevertheless, these large fluctuations can be observed much more often than is predicted. On figure 1.1 is possible to see the evolution of S&P 500, the most important stock index of New York stock exchange, in 2008. In the second half of the year the financial crisis broke out and the index lost about 40% of its value. That would be almost impossible within the exponential decay. Another assumption of Central limit theorem that can be misleading is assumption of full independence between different times. That means that there cannot be any memory in the market, which is also not true in real markets. The target of this work is to analyze the presence of these phenomena, to understand, why they arise and to present models that take these situations into account. We will introduce an approach discovered by B. Mandelbrot [23, 25, 27] based on the existence of two kinds of time - the real, clock time and the other time that runs virtually in the market. The transformation

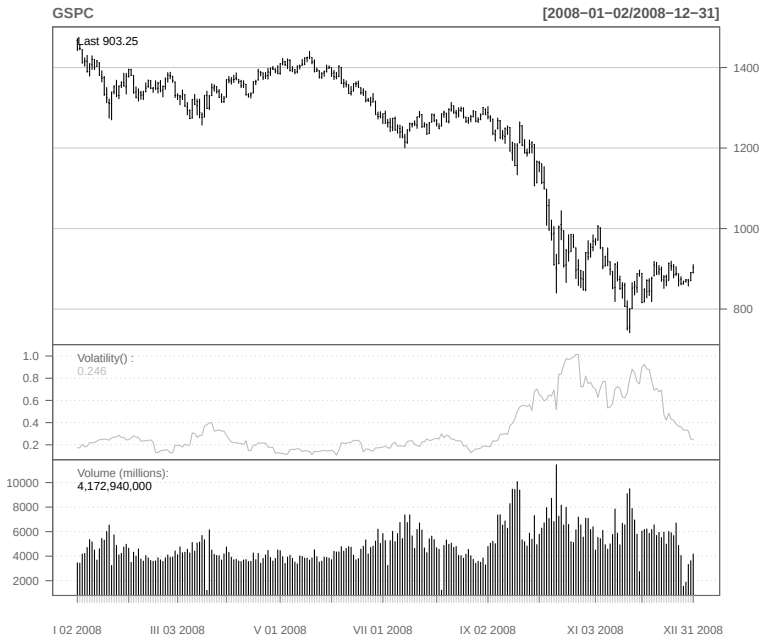



Figure 1.1: Evolution of index Standard and Poor's 500 in the year 2008. We can observe that with the begin of financial crisis the value of index fell down, the volatility was a few times higher than before, and the traded volume was also very high. In order to simulate and forecast this kind of behavior we need models that enables volatility fluctuations, rapid changes and existence of economic cycles. The graphics was generated by R package *quantmod* [35].

between them is given by multifractal deformations, which are used to generate processes that produce periods of large fluctuations, memory effects and other kinds of behavior observed on the real markets. The multifractal background provides a rich basis for these transformations with many applications. We also introduce a final result of theory of multifractal deformations, a Markov Switching Multifractal that was designed by B. Mandelbrot, L. Calvet and A. Fisher and that provides a complex simulation of real financial series through the model with variable volatility. The work is divided into following chapters: chapter two provides a physical motivation for studying random systems by generalizing the second Newton's law to its stochastic version - *Langevin equation* and shows alternative ways to describe these systems. In the third chapter we introduce a cornerstone of the whole work, *stochastic processes*, define special classes of stochastic processes, show the connection to Langevin equation and introduce differential and integral calculus for these processes based on ideas of K. Itō and R. Stratonovich. Chapter four is devoted to the *memory* and correlations of random processes, in chapter five we look at *scaling* properties and *self-similarity*. Chapter six is the most important chapter of the thesis, we discuss *multifractal formalism* and generation of *multifractal deformations*. In the penultimate chapter we discuss models used for simulation and forecasting of price returns including multifractal models. Finally, in the last chapter are shown some concepts discussed in previous chapters on the example of one concrete time series of index S&P 500.

# Chapter 2

# Physical Motivation: Langevin Equation

In the first chapter we look at the description of random processes from the physical point of view. The description of physical phenomena in the statistical way arose from the observation of systems with many degrees of freedom, like gases or liquids, so that it is impossible to describe them deterministically. One example that will accompany us through the whole work is a phenomenon called the Brownian motion. We will find the way to describe these processes and show the generalization of Newton's second law for random (or so called stochastic) processes.

## 2.1 Langevin Equation: Introduction

We consider a system in contact with a heat bath. It can be a particle (a molecule, or in larger scale a pollen grain, for example) flowing in a liquid or any other thinkable system. There exist fluctuations that cause a movement. They are caused by collisions of our system with other particles. We cannot describe in the deterministic way describe every single collision. How can we therefore describe our system without exact knowledge of all information about it? We cannot use the statistical description of equilibrium physical systems, because the particle interacts with the environment and it is not generally true, that the whole system stays in equilibrium. We have to build up a microscopic description, and we have to use some description, that includes statistical methods, anyway.

Let us begin in one dimension. This assumption is according to later application not restrictive. We start from the very beginning, that means from classical mechanics. The equation of motion for the single particle is classical Newton equation, e.g.

$$m\ddot{x}(t) = -U'(x) + F_{ext} \tag{2.1}$$

where $U$ is the potential and $F_{ext}$ is an external force. The external force conveys the binding with the outer system. We can divide the external forces into two groups: deterministic friction force and random force. Friction force

acts against the movement of the particle and grows with the velocity and is deterministic, whereas random force causes random shifts of the particle.

We can rewrite the previous equation into the form

$$m\ddot{x}(t) + U'(x) + \gamma\dot{x}(t) = F_{ran} \tag{2.2}$$

where $\gamma$ is a friction coefficient and $F_{ran} \equiv \eta$ is a random force. With the description of the force we have to partially help ourselves with the experiments and everyday experience. Our acquaintance of the behavior of the grain in the water tells us that the random force is not preferred in any particular direction and it does not depend on the position of the particle. We also can observe that the force depends on the past behavior of the particle, neither.

Because the force is random, we can only express this reality by statistical quantities. These a few basic observations can be formed into these relations:

$$\mathrm{E}(\eta(t)) = 0 \tag{2.3}$$

$$\mathrm{Cov}(\eta(t), \eta(t')) = \mathrm{E}(\eta(t)\eta(t')) = 2D\delta(t - t'). \tag{2.4}$$

With the substitution $v(t) = \dot{x}(t)$ and together with conditions on the random force $\eta$, we can now rewrite the equation (2.2) into the famous form, named after Paul Langevin, the discoverer of the equation. It is:

$$m\dot{v}(t) + \gamma v(t) + U'(x) = \eta(t). \tag{2.5}$$

Both parts of the external force, i.e. random noise and friction have their importance, as we can demonstrate on pathological cases with no noise, respectively with no friction. Further, we assume in this section the Langevin equation for free particle (i.e. $U'(x) = 0$).

**No friction:** In case when no friction is present, the equation becomes an easy form

$$m\dot{v} = \eta(t) \tag{2.6}$$

and the general solution is

$$v(t) = v_0 + \frac{1}{m}\int_0^t \mathrm{d}t'\eta(t'). \tag{2.7}$$

It is easy to see that the mean value is $\mathrm{E}(v(t)) = v_0$, since $\mathrm{E}(\eta(t)) = 0$. More interesting is to calculate the mean square velocity. With aid of (2.3) and (2.4) we get

$$\mathrm{E}(v^2(t)) = v_0^2 + \frac{2v_0}{m}\int_0^t \mathrm{d}t' \underbrace{E(\eta(t'))}_{=0} + \frac{1}{m^2}\int_0^t \mathrm{d}t' \int_0^t \mathrm{d}t'' \underbrace{\mathrm{E}(\eta(t')\eta(t''))}_{=2D\delta(t'-t'')} =$$

$$= v_0^2 + \frac{2Dt}{m^2}. \tag{2.8}$$

This means that variance of the velocity grows with time to infinity.

**No fluctuations:** We assume that $\eta(t) = 0$. The equation is reduced then to form

$$m\dot{v} + \gamma v = 0 \tag{2.9}$$

with the solution in the form

$$v(t) = v_0 \exp\left(-\frac{\gamma t}{m}\right). \tag{2.10}$$

We now send time to infinity, we get that $v(t) \to 0$. On the other hand, because of the absence of fluctuations the system is in equilibrium and we can derive the average value of $v^2$ from equipartition theorem, namely $\mathrm{E}(v^2) = \frac{k_B T}{m}$, where $k_B$ is the Boltzmann constant. As we have seen, both cases bring some unwanted properties, therefore it is important to take both the noise and friction into account. The general solution of Langevin equation describing free particle is

$$v(t) = v_0 \exp\left(-\frac{\gamma t}{m}\right) + \frac{1}{m}\int_0^t \mathrm{d}t' \eta(t') \exp\left(-\frac{\gamma(t-t')}{m}\right) \tag{2.11}$$

Because the force $\eta(t)$ is random, we can be interested in some statistical quantities to get some idea about the solution. The mean and average of the square velocity can be with (2.3) and (2.4) expressed as

$$\mathrm{E}(v(t)) = v_0 e^{-\frac{\gamma t}{m}} \tag{2.12}$$

$$\mathrm{E}(v^2(t)) = v_0^2 e^{-\frac{2\gamma t}{m}} + \frac{1}{m^2}\int_0^t \mathrm{d}t' \int_0^t \mathrm{d}t'' e^{-\frac{\gamma(t-t')}{m}} e^{-\frac{\gamma(t-t'')}{m}} \mathrm{E}(\eta(t')\eta(t'')) =$$

$$= v_0^2 e^{-\frac{2\gamma t}{m}} + \frac{D}{\gamma m}\left(1 - e^{-\frac{2\gamma t}{m}}\right) \tag{2.13}$$

and for variance of the velocity holds

$$\mathrm{Var}(v(t)) = \frac{D}{\gamma} m \left(1 - e^{-\frac{-2\gamma t}{m}}\right). \tag{2.14}$$

We can see that for long times, i.e. relaxation to equilibrium state, the average value of square of velocity tends to the equilibrium value $\frac{D}{\gamma m} = \frac{k_B T}{m}$. For small times we get up to $\mathcal{O}(t^2)$ the expression $\mathrm{Var}(v(t)) = \frac{2Dt}{m^2}$. This means that for small times the velocity varies very little and its practically constant. This means that the displacement for small time is $\mathrm{E}(x(t) - x(0)) \simeq v_0 t$. This is the behavior of the free particle. Nevertheless, the initial velocity is hampered by the friction and the behavior of the particle considerably changes after some time.

**Variance of displacement for long times:** Let us assume that $x(0) = 0$. By integration of the velocity we can get a solution for the position. The mean square displacement is (the result is taken from [8]):

$$\mathrm{Var}(x(t)) = \mathrm{E}(x^2(t)) = \int_0^t \mathrm{d}t' \int_0^t \mathrm{d}t'' \mathrm{E}(v(t')v(t'')) \overset{\text{large } t}{\longrightarrow} 2Dt. \tag{2.15}$$

We observe that for long times the behavior of the particle is different, it is now driven by diffusion and the effect of initial velocity is unrecognizable. This is the famous result of Albert Einstein [12] of effect of fluctuations and explains the principle of Brownian motion.

## 2.2 Overdamped Langevin Equation

In this section we consider a Langevin equation with the potential, but assume that the mass of the particle is small and therefore for large times $t \gg \frac{m}{\gamma}$ we can neglect the first term $m\ddot{x}(t)$ and we get so called Overdamped Langevin equation:

$$\gamma\dot{x}(t) + U'(x) = \eta(t) \tag{2.16}$$

with usual conditions (2.3) and (2.4). The equation describes the situation, when friction dominates and movement is caused especially by fluctuations. Unfortunately, there exists the analytic solution only for very few potentials. Besides free particle exists a solution also for harmonic potential $U(x) = \frac{kx^2}{2}$. This process is called Ornstein-Uhlenbeck process [29]. The equation becomes as follows:

$$\gamma\dot{x} + kx = \eta(t). \tag{2.17}$$

This is linear differential equation of the first order, so we can write down the solution

$$x(t) = x_0 e^{-\frac{kt}{\gamma}} + \int_0^t \mathrm{d}t' \eta(t') e^{-\frac{k(t-t')}{\gamma}}. \tag{2.18}$$

The mean and the variance of $x(t)$ have the same form as (2.12) and (2.14), so it is possible to write them down immediately

$$\mathrm{E}(x(t)) = x_0 e^{-\frac{2kt}{\gamma}} \tag{2.19}$$

$$\mathrm{Var}(x(t)) = \frac{D}{\gamma k}\left(1 - e^{-\frac{2kt}{\gamma}}\right). \tag{2.20}$$

As we could expect, the particle becomes very quickly to the area of a minimum of the potential and moves around the minimum driven by thermal fluctuations in the equilibrium regime after sufficiently long time.

## 2.3 Smoluchovski Equation and Diffusion Equation

So far, we were interested only in trajectories of the particle and their statistics. We would also like to find another description, which is maybe for some people more familiar. We would like to find the probability distribution $p(x,t)$ of the trajectory $x(t)$. This duality has some parallels in classical mechanics (Lagrange description versus Liouville description). In order to calculate the probability distribution, we first look at the definition of mean value.

The mean value of some function $f$ of a random process $x(t)$ is defined as:

$$\mathrm{E}(f(x(t)) = \int_\Omega \mathrm{d}z f(z) p(z,t), \tag{2.21}$$

where $\Omega$ is the set on which a random process $x(t)$ is defined. If we formally plug in for the probability distribution the expression $p(x,t) = \mathrm{E}(\delta(x - x(t))$, we get again the expectation value of $f$:

$$\begin{aligned}\int_\Omega \mathrm{d}z f(z)p(z,t) &= \int_\Omega \mathrm{d}z f(z)\mathrm{E}(\delta(z - x(t)) \\ &= \mathrm{E}\left(\int_\Omega \mathrm{d}z f(z)\delta(z - x(t))\right) = \mathrm{E}(f(x(t)).\end{aligned} \tag{2.22}$$

This relation has also an intuitive explanation: $p(x,t)\mathrm{d}x$ is the probability that the particle is found in interval $[x, x+\mathrm{d}x]$ in time $t$, i.e. number of all trajectories, that go through the interval at time $t$. At first, we formally integrate the equation (2.16)

$$x(t) = x_0 + \frac{1}{\gamma}\int_0^t U'(x(t'))\mathrm{d}t' + \frac{1}{\gamma}\int_0^t \eta(t')\mathrm{d}t' \tag{2.23}$$

and rewrite the delta function through the Fourier transform as

$$\mathrm{E}(\delta(x - x(t)) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k\ \mathrm{E}\left(e^{-ik(x-x(t))}\right) =$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k\ e^{-ik\left(x - x_0 - \frac{1}{\gamma}\int_0^t U'(x(t'))\mathrm{d}t'\right)}\mathrm{E}\left(e^{\frac{-ik}{\gamma}\int_0^t \eta(t')\mathrm{d}t'}\right). \tag{2.24}$$

Now we look closer at the last exponential term

$$\mathrm{E}\left(\exp\left(-ik\frac{1}{\gamma}\int_0^t \eta(t')\mathrm{d}t'\right)\right).$$

Up till now was the definition of $\eta(t)$ not complete. We did not define higher point correlations, because for studied behavior it was not necessary. Now we have two good reasons to refine the definition of random force a bit. Firstly, it helps us to continue with the calculation, secondly, the random force $\eta$ will be closely related to other processes that will be introduced in following chapters.

**Definition 2.1** *The random process $X(t)$ is called Gaussian noise, if $X(t) \sim \mathcal{N}(0,\sigma)$ and $\mathrm{E}(X(t_1)\dots X(t_n)) = 0$ for different $t_1,\dots,t_n$.*

We assume that $\eta(t)$ is a Gaussian noise. We have a look to the moments of $\eta$. Because it is a Gaussian variable, then

$$\mathrm{E}(\eta^{2n}(t)) = (2D)^n(2n-1)!!, \tag{2.25}$$

$$\mathrm{E}(\eta^{2n+1}(t)) = 0. \tag{2.26}$$

We now expand the exponential function into the Taylor series and from linearity we apply the mean operation on every term:

$$\mathrm{E}\left(e^{\frac{-ik}{\gamma}\int_0^t \eta(t')\mathrm{d}t'}\right) = 1 - \frac{ik}{\gamma}\int_0^t E(\eta(t'))\mathrm{d}t' - \frac{k^2}{2\gamma^2}\int_0^t\int_0^t \mathrm{E}(\eta(t')\eta(t''))\mathrm{d}t'\mathrm{d}t'' + \dots \tag{2.27}$$

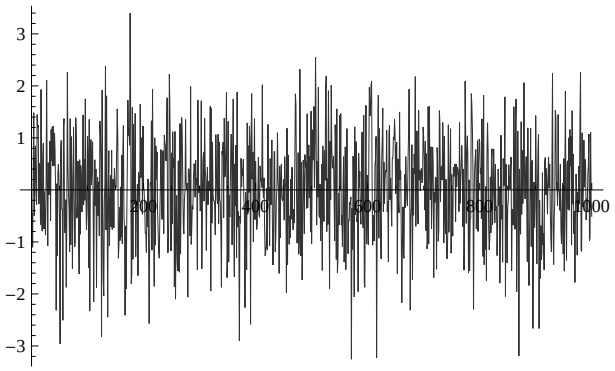



Figure 2.1: 1000 steps of Gaussian (or white) noise.

With the definition of the moments the odd terms vanish and the series becomes an elegant form:

$$\mathrm{E}\left(e^{\frac{-ik}{\gamma}\int_0^t \eta(t')\mathrm{d}t'}\right) = \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{-k^2 Dt}{\gamma^2}\right)^n = e^{-\frac{k^2 Dt}{\gamma^2}}. \tag{2.28}$$

The whole expression for the $p(x,t)$ can be expressed as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k\; e^{-ik\left(x-x_0-\frac{1}{\gamma}\int_0^t U'(x(t'))\mathrm{d}t'\right)} e^{-\frac{k^2 Dt}{\gamma^2}}. \tag{2.29}$$

We become the final result, when we calculate $\frac{\partial p(x,t)}{\partial t}$:

$$\frac{\partial p(x,t)}{\partial t} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k\left[\frac{-ik}{\gamma}U'(x)-\frac{k^2 D}{\gamma^2}\right] e^{-ik\left(x-x_0-\frac{1}{\gamma}\int_0^t U'(x(t'))\mathrm{d}t'\right)} e^{-\frac{k^2 Dt}{\gamma^2}}. \tag{2.30}$$

In the terms we can recognize Fourier images of first, resp. second derivative, so finally:

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}\left[\frac{1}{\gamma}U'(x)p(x,t)\right] + \frac{D}{\gamma^2}\frac{\partial^2 p(x,t)}{\partial x^2}. \tag{2.31}$$

This equation is called Smoluchowski equation and describes the probability distribution $p(x,t)$ of particles at time $t$ in the overdamped limit. There are few interesting remarks:

- **free particle**: For $U'(x) = 0$ the equation (2.31) becomes Diffusion equation

$$\frac{\partial p(x,t)}{\partial t} = \mathcal{D}\frac{\partial^2 p(x,t)}{\partial x^2}, \tag{2.32}$$

  where $\mathcal{D} = \frac{D}{\gamma^2} = \frac{k_B T}{\gamma}$. The solution is known,

$$p(x,t) = \frac{1}{\sqrt{4\pi\mathcal{D}t}}\exp\left(-\frac{(x-x_0)^2}{4\mathcal{D}t}\right). \tag{2.33}$$

  The variance is $\mathrm{Var}(x(t)) = 2\mathcal{D}t$. This is a classical result observed by Einstein in 1905 [12].

- **stationary solution**: if we assume that the probability is time-independent, then the equation (2.31) simplifies to form

$$\frac{\mathrm{d}p(x)}{\mathrm{d}x} = \frac{1}{\gamma\mathcal{D}}U'(x)p(x). \tag{2.34}$$

This is the equation for Maxwell-Boltzmann distribution

$$p(x) = \mathcal{N}\exp\left(-\frac{U(x)}{k_B T}\right), \tag{2.35}$$

which is in perfect agreement with equilibrium theory.

## 2.4 General Langevin Equation and Klein-Kramers Equation

In this section we look at general non-overdamped Langevin equation. We introduce a transformation $\dot{x}(t) = v(t)$ and rewrite the equation as a system of differential equations of the first order, concretely:

$$\begin{aligned} \dot{x}(t) &= v(t) && (2.36)\\ \dot{v}(t) &= -\frac{U'(x)}{m} - \gamma v(t) + \frac{F(t)}{m} && (2.37)\\ \mathrm{E}(F(t)F(t')) &= 2D\delta(t-t'). && (2.38) \end{aligned}$$

One example of system described by these equations is the RLC electric circuit with random electric force. This system was described by Johnson and Nyquist in 1928 [28]. The electric charge plays role of the position in this equation.

We look also at the analogical probability distribution belonging to the full Langevin equation. The derivation of corresponding probability distribution is similar, but we have to make the transformation in two variables, $x$ and $v$: $p(x,v,t) = \mathrm{E}[\delta(x-x(t))]\mathrm{E}[\delta(v-v(t))]$. The corresponding solution has this form [34]:

$$\partial_t p(x,v,t) + v\partial_x p(x,v,t) - \frac{U'(x)}{m}\partial_v p(x,v,t) = \frac{\gamma}{m}\partial_v\left[vp(x,v,t)\right] + \frac{\mathcal{D}^2}{m^2}\partial_{vv}p(x,v,t). \tag{2.39}$$

This equation is called *Klein-Kramers* equation and it contains the full information about evolution of the system in the phase space. On the left side of the equation we can identify the total derivative with respect to t, i.e.:

$$\frac{\mathrm{d}p(x(t),\dot{x}(t),t)}{\mathrm{d}t} = \frac{\partial p(x(t),\dot{x}(t),t)}{\partial x}\dot{x} + \frac{\partial p(x(t),\dot{x}(t),t)}{\partial \dot{x}}\ddot{x} + \frac{\partial p(x(t),\dot{x}(t),t)}{\partial t} \tag{2.40}$$

which represents the classical-mechanical density in the phase space and in the deterministic regime is equal to zero. On the right side we can recognize terms that correspond to friction and noise. It is similar to the recognition of particular terms in Langevin equation.

In this chapter we introduced the concept of the non-equilibrium processes. In following chapters we will see that there is a connection between physical processes, mathematical description and processes in financial markets. All of

them will produce similar behavior and it is therefore very suitable to find the common language between all of these branches. The next chapter will provide a sufficient theoretical background to describe it.

# Chapter 3

# Mathematical Background: Stochastic Processes

In the first chapter we described the behavior of the system with random fluctuation in the phenomenological way, by using physical laws. In this part of the work a theoretical concept of description of randomness, so called *stochastic processes*, will be introduced. Although the formalism looks at first sight quite different from the one commonly used in physics (and in chapter 2.1), at the end of the chapter one emerges strong relationship between physical and mathematical point of view.

## 3.1 Basic notes

We begin with the concept of stochastic process. This is the keyword of the description of time evolution in the probability theory.

**Definition 3.1** *A stochastic process is an ordered set of random variables $X(t)$, where $t \in \mathcal{I} \subset \mathbf{R}$.*

The parameter $t$ has usually meaning of time and can be either discrete or continuous. Then we talk about *discrete* or *continuous* stochastic processes. We can look at the stochastic process in two different ways. For a given time is $X(t)$ a random variable with a probability distribution $p(x,t)$. We can then understand $p(x,t)$ as evolution of the probability in time. On the other hand, one realization of the process $X_\omega(t)$ represents one concrete possible evolution of the system. The mapping $t \mapsto X_\omega(t)$ is called *sample path* and we can think about $X(t)$ as about a probability density of sample paths.

We saw this dual description also in the first chapter, where the Langevin equation was an equation for sample paths, whereas Smoluchovski equation resp. Klein-Kramers equation were equations for the probability distribution. This duality is favorable and in the next we will combine both possible descriptions.

## 3.2 Gaussian Processes

One important class of stochastic processes are stable processes. The reason of its importance comes from the property of stability of the process. The

term *stable distribution* is discussed in appendix A. The main importance of stable distributions is the property, that they are limiting distributions of sum of random variables. That means that every process compound of many random variables can be for sufficiently long times well approximated by the process with the limiting distribution. This favorable property can be transferred to the stochastic processes.

**Definition 3.2** *Stochastic process is stable if all its finite dimensional joint distributions are stable.*

The most popular and used distribution of stable distributions is normal distribution. We define important class of stochastic processes are *Gaussian processes*. The class of Lévy processes will be discussed in chapter 5.

**Definition 3.3** *Gaussian process is a stochastic process for that for every arbitrary finite subset of times $t_i \in \mathcal{I}$ has the vector $(X(t_1), X(t_2), \ldots, X(t_n))$ multinomial Gaussian distribution, so*

$$p(x_1, t_1; \ldots; x_n, t_n) = \sqrt{\frac{|\det \mathbb{A}|}{(2\pi)^n}} \exp\left( -\frac{1}{2} \sum_{i,j=1}^{n} (x_i - E(X(t_i))\mathbb{A}_{ij}(x_j - E(X(t_j)) \right)$$

*where $\mathbb{A}^{-1}$ is the covariance matrix: $\mathbb{A}^{-1}_{ij} = \mathrm{Cov}(X(t_i), X(t_j))$*

Because of the stability of the normal distribution we can immediately formulate a corollary about Gaussian processes.

**Corollary 3.4** *The process $X(t)$ is Gaussian, if every finite linear combination $\sum_i a_i X(t_i)$ is identically zero or has a Gaussian distribution.*

Every Gaussian process is fully defined with two quantities: $\mathrm{E}(X(t))$ and $\mathrm{Cov}(X(t), X(s))$. We have already met one example of the Gaussian process. It is the Gaussian noise with $\mathrm{E}(\eta(t)) = 0$ and $\mathrm{Cov}(\eta(t), \eta(s)) = 2D\delta(t-s)$. We aim in the next section on one very important example of Gaussian processes.

### 3.2.1 Wiener Process

Wiener process is formal mathematical concept of Brownian motion. The Wiener process describes therefore the random movement of the particle. The movement has no preferred direction and no memory. Before we define Wiener process, we discuss one important property of stochastic processes.

**Definition 3.5** *A stochastic process is $X(t)$ called (strictly) stationary, if for an arbitrary vector $(X(t_1), \ldots, X(t_n))$ holds:*

$$(X(t_1 + \tau), \ldots, X(t_n + \tau)) \overset{d}{=} (X(t_1), \ldots, X(t_n)).$$

**Definition 3.6** *A stochastic process $W(t)$ (for $t \in [0, \infty]$) is called Wiener process, if*

- $W(0) \overset{a.s.}{=} 0$
- *for every $t, s$ are increments $W(t) - W(s)$ stationary process with distribution*

$$W(t) - W(s) \sim \mathcal{N}(0, |t - s|).$$

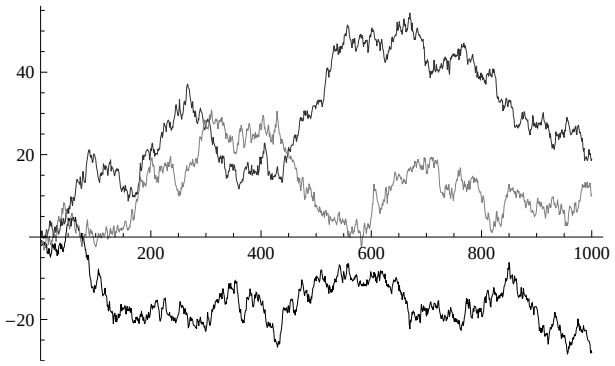



Figure 3.1: Three sample paths of Wiener process for 1000 steps.

- *for different values are increments not correlated.*

The Wiener process is therefore a Gaussian process with the probability distribution

$$p_W(x,t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right). \tag{3.1}$$

Because every Gaussian process is described by mean and covariance, we calculate at first the mean:

$$\mathrm{E}(W(t)) = \mathrm{E}(W(t) - W(0)) = 0 \tag{3.2}$$

For covariance we assume that $s > t$:

$$\begin{aligned}\mathrm{Cov}(W(s)W(t)) = \mathrm{E}(W(s)W(t)) = \mathrm{E}([(W(s) - W(t)) + W(t)]W(t)) = \\ = \mathrm{E}((W(s) - W(t)(W(t) - W(0))) + \mathrm{E}(W(t)W(t)) = \\ = \mathrm{Cov}(W(s) - W(t), (W(t) - W(0))) + t = t = \min(s,t). \end{aligned} \tag{3.3}$$

Therefore is Wiener process the Gaussian process with zero mean vector and covariance matrix $\min(s,t)$. The conditional probability of Wiener process can be therefore written in form

$$p_W(x_2,t_2|x_1,t_1) = \frac{1}{\sqrt{2\pi(t_2-t_1)}} \exp\left(-\frac{(x_2-x_1)^2}{2(t_2-t_1)}\right). \tag{3.4}$$

The important property of every stochastic process is its continuity and differentiability. With the conditional probability distribution we check firstly the continuity of the process. For an arbitrary $C > 0$:

$$\begin{aligned}\lim_{h\to 0} P[|W(t+h) - W(t)| > C] = \lim_{h\to 0} P[|(W(h) - W(0))| > C] = \\ = \lim_{h\to 0} P[|(W(h)|W(0))| > C] = 2\lim_{h\to 0}\int_C^\infty \mathrm{d}x \frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{x^2}{2h}\right) = \\ = 2\int_C^\infty \mathrm{d}x \lim_{h\to 0} \frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{x^2}{2h}\right) = 2\int_C^\infty \delta(x)\mathrm{d}x \ = 0. \end{aligned} \tag{3.5}$$

So Wiener process is almost everywhere continuous. This property was already given to the definition of the process. With similar calculations we get that sample paths of Wiener process are almost nowhere differentiable:

$$\lim_{h\to 0} P\left[\left|\frac{W(t+h)-W(t)}{h}\right| > C\right] = 2\lim_{h\to 0}\int_{Ch}^{\infty} \mathrm{d}x \frac{1}{\sqrt{2\pi h}}\exp\left(-\frac{x^2}{2h}\right) =$$
$$\int_{-\infty}^{\infty} \mathrm{d}x \underbrace{\lim_{h\to 0}\frac{1}{\sqrt{2\pi h}}\exp\left(-\frac{x^2}{2h}\right)}_{=\delta(x)} - \underbrace{2\lim_{h\to 0}\int_0^{Ch}\frac{1}{\sqrt{2\pi h}}\exp\left(-\frac{x^2}{2h}\right)\mathrm{d}x}_{\to 0} = 1. \quad (3.6)$$

The process is neither differentiable in the mean-square meaning:

$$l.i.m._{h\to 0^+}\frac{W(t+h)-W(t)}{h} = \lim_{h\to 0^+}\frac{h}{h^2} = \infty. \quad (3.7)$$

The fact of non-differentiability of Wiener process will be discussed in next sections and will lead us to the definition of stochastic differentiation and integration, which is slightly different from the deterministic integration as we know from Riemann integration.

The next parts will be devoted to other classes of processes that are somehow related to important properties of some Gaussian processes, especially Wiener process.

### 3.2.2 Ornstein-Uhlenbeck Process

Another example of Gaussian process is Ornstein-Uhlenbeck process. We have met it already as an example of process driven by overdamped Langevin equation. Now we define it as a stochastic process. We define this process in other way than Wiener process, to be more specific, we define mean and covariance function and from that we derive other properties.

**Definition 3.7** *A stationary Gaussian process $Y(t)$ is called Ornstein-Uhlenbeck process if*

- $\mathrm{E}(Y(t)) = 0$
- $\mathrm{Cov}(Y(t)Y(t')) = \exp(-\upsilon|t-t'|)$

We can immediately write unconditional and conditional probability distributions given by definition 3.3

$$p_Y(x,t) = \sqrt{\frac{1}{2\pi}}\exp\left(-\frac{x^2}{2}\right) \quad (3.8)$$

$$p_Y(x_2,t_2|x_1,t_1) = \sqrt{\frac{1}{2\pi(1-\Gamma^2)}}\exp\left(-\frac{(x_2-x_1\Gamma))^2}{2(1-\Gamma^2)}\right) \quad (3.9)$$

where $\Gamma = \exp(-\upsilon(t_2-t_1))$. Contrary to the Wiener process is Ornstein-Uhlenbeck process stationary (Wiener process has only stationary increments). On the other hand, the mean value in time $t$ for given initial value $Y(0) = x_0$ is

$$\mathrm{E}(Y(t)|Y(0)) = x_0\Gamma = x_0(\exp(-t)). \quad (3.10)$$

For long times it tends to zero. This property is called *mean-reverting*, because the process converges to its unconditional mean value. Ornstein-Uhlenbeck process has also interesting physical meaning that we will see later after derivation of some properties of special classes of random processes.

## 3.3 Martingales

The concept of martingales represent processes that can be called "fair game" and comes from environment of hazard. Roughly said, in every step of the game one has the same chance to loose as to win. On average the player will have the same amount of money after every round as before the round. In this game can players show their playing qualities, the casino will not bankrupt because of large amounts that would have to be payed to players and players will probably not lose all their money. In the language of stochastic processes, the martingale process is such a process, whose average value is estimated from the last realized value. At first, we introduce the term filtration which relates to already realized values of stochastic process.

**Definition 3.8** *Let the triplet $(\Omega, \mathcal{F}, p)$ be a probability space. Every sequence of $\sigma$-algebras $\{\mathcal{F}_t\}_{t\in\mathcal{I}} \subseteq \mathcal{F}$, such that for $t \leq s$ holds the inclusion $\mathcal{F}_t \subseteq \mathcal{F}_s$ is called the filtration.*

**Definition 3.9** *The stochastic process $X(t)$ is adapted to filtration $\mathcal{F}_t$, if for every time $t_0 \in \mathcal{I}$: $X(t = t_0)$ is a $\mathcal{F}_{t_0}$-measurable random variable.*

Roughly speaking, the first definition defines a set of events that already happened and the second definition tells us that adapted process is such process that in every time contains only the information that is known up to that time. We can also say that "process does not see to the future". Now the martingale processes can be defined in the following way.

**Definition 3.10** *Martingale process with respect to filtration $\mathcal{F}_t$ is a stochastic process, for that:*

1. $\mathrm{E}(|X(t)|) < +\infty$ *for every* $t \in \mathcal{I}$,
2. $X(t)$ *is adapted to* $\mathcal{F}_t$,
3. *for every* $s, t$ *such that* $s < t$: $\mathrm{E}(X(t)|\mathcal{F}_s) = X(s)$.

We can check, if a Wiener process is a martingale:

$$\mathrm{E}(W(t)|\mathcal{F}_s) = \mathrm{E}(W(t) - W(s)|\mathcal{F}_s) + \mathrm{E}(W(s)|\mathcal{F}_s). \tag{3.11}$$

Because the first term is stationary, we can shift the whole process by $-s$. We get the term $\mathrm{E}(W(t-s) - W(0)|\mathcal{F}_0) = 0$. Because the second term is $\mathcal{F}_s$-measurable (we already know the value), so it is equal to $W_s$. From that we get that $W(t)$ is martingale. Another example of martingale process is a process $W^2(t) - t$ [5].

We define also two generalizations of martingale.

**Definition 3.11** *A stochastic process $X(t)$ is called supermartingale, resp. submartingale if obeys conditions 1,2 from definition 3.10 and for every $s, t$; $s < t$ holds*

- $\mathrm{E}(X(t)|\mathcal{F}_s) \leq X(s)$, *for supermartingale, resp.*
- $\mathrm{E}(X(t)|\mathcal{F}_s) \geq X(s)$, *for submartingale.*

If the process is both supermartingale and submartingale, then it is martingale. The next proposition is a simple corollary of Jensen's inequality and shows, how can be supermartingales and submartingales created from martingale processes.

**Proposition 3.12** *Let $X(t)$ be a martingale process w.r.t. filtration $\mathcal{F}_t$ and let $f(x)$ be a convex (resp. concave) function, such that $\mathrm{E}(|f(X(t))|) < +\infty$. Then the process $f(X(t))$ is submartingale (resp. supermartingale).*

We show the first case, while the second case is analogical. The Jensen's inequality implies:

$$f(X(s)) = f(\mathrm{E}(X(t)|\mathcal{F}_s) \leq \mathrm{E}(f(X(t))|\mathcal{F}_s). \tag{3.12}$$

Martingales are important in many applications, wherever the request on "fair games" arises. We use it later for definition of stochastic integration which will be introduced later in one of the following sections.

## 3.4 Markov Processes

Another important class in the theory of stochastic processes are processes that have no memory, i.e. the future evolution is influenced only by the present configuration of the system and not by the past values. We can imagine it as a process that "does not remember the past". For probability distribution it means:

$$p(x_{n+1}, t_{n+1}|x_n, t_n; \ldots; x_1, t_1) = p(x_{n+1}, t_{n+1}|x_n, t_n). \tag{3.13}$$

These processes, that are called *Markov processes*, are fully described only by 1-point probability $p(x,t)$ and 1-point conditional probability $p(x,t|x',t')$. Every joint probability distribution can be expressed according to the relation between conditional and unconditional probability

$$p(x,y) = p(x|y)p(y)$$

as a product of conditional probabilities:

$$p(x_n, t_n; \ldots; x_0, t_0) = \left(\prod_{i=1}^{n} p(x_i, t_i|x_{i-1}, t_{i-1})\right) p(x_0, t_0). \tag{3.14}$$

If we assume stationary Gaussian process, we can formulate a criterion, whether the process is Markovian or not.

**Proposition 3.13** *Let $X(t)$ be a a Gaussian stationary process. The process is Markov process, if*

$$E(X(t)X(s)) = \mathcal{C}\exp(-\gamma|t-s|).$$

The proof of the proposition can be found in [1].

Accordingly, the exponential decay of covariance function is typical for processes, that are not directly influenced by previous configurations. These processes are called processes with short-range memory. One example is the Ornstein-Uhlenbeck process. We will discuss the topic of the memory in the fifth chapter.

### 3.4.1 Chapman-Kolmogorov Equation

Let us focus again into the relation (3.14), now we consider this relation for three points

$$p(x_3, t_3; x_2, t_2; x_1, t_1) = p(x_3, t_3|x_2, t_2)p(x_2, t_2|x_1, t_1)p(x_1, t_1). \tag{3.15}$$

We integrate the whole equation over $x_2$ and get on the left hand side a 2-point probability distribution

$$p(x_3, t_3; x_1, t_1) = \int \mathrm{d}x_2 p(x_3, t_3|x_2, t_2)p(x_2, t_2|x_1, t_1)p(x_1, t_1). \tag{3.16}$$

On the other hand we can write the joint probability with help of conditional probability and 1-point probability, so

$$p(x_3, t_3; x_1, t_1) = p(x_3, t_3|x_1, t_1)p(x_1, t_1) \tag{3.17}$$

and therefore we plug it into the equation (3.16) and drop out terms $p(x_1, t_1)$. After that we get *Chapman-Kolmogorov equation*, which is equivalent formulation of condition on conditional probability distribution of Markov process, so

$$p(x_3, t_3|x_1, t_1) = \int \mathrm{d}x_2 p(x_3, t_3|x_2, t_2)p(x_2, t_2|x_1, t_1), \tag{3.18}$$

for $t_3 \geq t_2 \geq t_1$.

### 3.4.2 Markov Chains

Now we look at the situation, when the probability space is finite, i.e. there are only finite number of possible states that can happen. We number the states, so we can describe the state space as $\{x_i\}_{i=1}^n$. The process of evolution is given, as we know from previous section, by two probabilities, concretely by $p(x, t)$ and $p(x, t|x', t')$. We also consider discrete time steps, so $t = n \in \{1, 2, \dots\}$. We can then write down a probability of transition from state $\xi_n(x) = x_i$ to $\xi_{n+1}(x) = x_j$ in one step as

$$P[X = x_j, t = n + 1|X = x_i, t = n] = p_1(j|i)(n) \tag{3.19}$$

which can be represented as matrix $\mathbb{P}_1(n)$ with elements $p_1(j|i)$. We introduce a vector of probabilities of particular states as $v(\xi_n)_k = p(\xi_n = k)$ for $k \in \{1, \dots, n\}$. We can then calculate the probability distribution $v(\xi_{n+1})$ from the previous step as

$$v(\xi_{n+1}) = \mathbb{P}_1(n)v(\xi_n). \tag{3.20}$$

We confine ourselves on stationary processes. It means, that $\mathbb{P}_1$ does not depend on $n$, because the difference between given times is constantly one. How can be now calculated the change of the probability vector after two steps directly from $v(\xi_n)$? We can apply (3.20) on $v(\xi_{n+2})$ twice and we get:

$$v(\xi_{n+2}) = \mathbb{P}_1 v(\xi_{n+1}) = \mathbb{P}_1^2 v(\xi_n) = \mathbb{P}_2 v(\xi_n). \tag{3.21}$$

Therefore, the transition matrix for $m$ steps is equal to the $m$-th power of the transition matrix for one step:

$$\mathbb{P}_n = \mathbb{P}_1^n. \tag{3.22}$$

Markov chains ( = stationary discrete processes on finite state space) play an important role in many applications, for us will be important, when we define Markov switching multifractals.

### 3.4.3 Master Equation

Similarly to Markov chains we look at stationary processes with continuous time steps. We will be able to concretize the Chapman-Kolmogorov equation in order to get simpler equations. From definition of stationary process we can deduce that unconditional 1-point probability is time independent and the conditional probability depends only on the difference of time. In this section we will be using the following notation

$$p(x,t) =: p(x) \tag{3.23}$$

$$p(x,t|x',t') =: p(x|x';\tau) \tag{3.24}$$

where $\tau = t - t'$. We rewrite the equation (3.18) with the notation for stationary processes

$$p(x_3|x_1;t'+t) = \int \mathrm{d}x_2 p(x_3|x_2;t')p(x_2|x_1;t) \tag{3.25}$$

and treat this equation for small time $t' \to 0$. For transitional probability holds:

$$p(x_3|x_2;t') \overset{t'\to 0}{\longrightarrow} \delta(x_3 - x_2), \tag{3.26}$$

because the only possible value is the current one. We define a *transition rate*, which is the transition probability per unit time as $w(x_3,x_2;t) = \frac{p(x_3|x_2;t)}{t}$, and for small time difference $t \to 0$ we assume, that this transition rate does not depend on time in the first approximation on time, so

$$p(x_3|x_2;t') = t'w(x_3,x_2) + o(t') \tag{3.27}$$

and we formally add the term the zero term $\left(1 - t'\int \mathrm{d}x' w(x'|x_2)\right)\delta(x_3 - x_2)$, where the first bracket represents the term of probability, that the state stays up to time $t'$ in the same state and the delta function ensures that is should be the state $x_3$. Together we get

$$p(x_3|x_2;t') \simeq t'w(x_3,x_2) + \left(1 - t'\int \mathrm{d}x' w(x'|x_2)\right)\delta(x_3 - x_2). \tag{3.28}$$

We can check that for $t' \to 0$ we get directly delta function. Now we apply this expression to the term $p(x_3|x_2;t')$ in the Chapman-Kolmogorov equation

$$\begin{aligned} p(x_3|x_1;t+t') &= \int \mathrm{d}x_2 t' w(x_3,x_2)p(x_2|x_1;t) \\ &\quad + \int \mathrm{d}x_2 \delta(x_3 - x_2)p(x_2,x_1)\left[1 - t'\int \mathrm{d}x' w(x'|x_2)\right] \\ &= \int \mathrm{d}x_2 t' w(x_3|x_2)p(x_2|x_1;t) \\ &\quad + t'p(x_3|x_1;t) - t'\int \mathrm{d}x' w(x'|x_3)p(x_3|x_1;t). \end{aligned} \tag{3.29}$$

If we now want to calculate the derivative of $p(x_3|x_1;t)$ with respect to $t$, we get

$$\frac{\partial p(x_3|x_1;t)}{\partial t} = \lim_{t'\to 0} \frac{p(x_3|x_1;t+t') - p(x_3|x_1;t)}{t'}$$
$$= \int \mathrm{d}x_2 t' w(x_3|x_2) p(x_2|x_1;t) - \int \mathrm{d}x_2 w(x_2|x_3) p(x_3|x_1;t). \tag{3.30}$$

In the second integral we renamed the dummy variable $x'$ for $x_2$. This equation is called *Master equation* and it is a general equation for stationary Markov process. The benefit of the equation is that Chapman-Kolmogorov equation is non-linear equation for conditional probability, whereas at Master equation we have terms with time-independent transition rates $w(x|y)$, which are characteristic constants for given system and we suppose that are known, or measured from an experiment. Conditional probabilities itself are in both terms linear, so with given function $w$ it holds the superposition principle.

### 3.4.4 Fokker-Planck Equation

Although is Master equation a linear equation, it is still integro-differential equation and therefore very difficult to solve. We approximate Master equation in order to get simpler equation that would still describe sufficiently most of random processes. For an initial probability $p(x_1, 0)$ is initial value $x_1$ known and we introduce a new variables $r := x_3 - x_2$ and $x = x_3$, so that

$$\frac{\partial p(x;t)}{\partial t} = \int \mathrm{d}r \left[ w(x-r|r) p(x-r;t) - w(x|-r) p(x;t) \right]. \tag{3.31}$$

For the next approximations we make a few assumptions on $w$ and $p$, namely:

- $w(x-r|r)$ is function with sharp maximum at $r = 0$, so for large $r$ is $w(x-r|r) \simeq 0$,
- $w(x-r|r)$ is slowly varying function, so for small $r$ is $w(x-r|r) \simeq w(x|r)$,
- $p(x-r;t)$ is also slowly varying function,
- $p$ and $w$ are smooth functions.

We expand now the first term in terms of $(r-x)$ around the value $x = 0$ and see that first term cancels with the remaining term from the Master equation, so

$$\frac{\partial p(x,t)}{\partial t} = \int \mathrm{d}r w(x|r) p(x,t) - \int \mathrm{d}r w(x|-r) p(x,t)$$
$$- \int \mathrm{d}r r \frac{\partial}{\partial x} \left[ w(x,r) p(x,t) \right] + \frac{1}{2} \int \mathrm{d}r r^2 \frac{\partial^2}{\partial x^2} \left[ w(x,r) p(x,t) \right] + \dots$$
$$= \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \left[ a_n(x) p(x,t) \right], \tag{3.32}$$

where coefficients $a_n(x)$ are n-th moments of function $w$

$$a_n(x) = \int_{-\infty}^{\infty} \mathrm{d}r r^n w(x|r). \tag{3.33}$$

From the assumption of sharpness of the maximum we can deduce that coefficients $a_n(x)$ are negligible from some $n$, because for values around zero the term $r^n$ rapidly decreases and for large $r$ is zero the term $w(x-r|r)$. The most common choice of number of non-vanishing terms is 2, so only first and second moment of $w(x|r)$ are non-zero. With this assumption we get famous *Fokker-Planck equation*

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[a_1(x)p(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[a_2(x)p(x,t)\right]. \tag{3.34}$$

The coefficient $a_1$ is called *drift coefficient* and $a_2$ is called *diffusion coefficient*. The meaning of these coefficients is clear, because this equation is - unsurprisingly - formally equal to Smoluchowski equation (2.31). Coefficient $a_1$ is proportional to external force and therefore to the deterministic drift and $a_2$ is proportional to diffusion. The fact is that both descriptions lead to the same differential equation and the same description of the system. The corresponding Langevin equation for the Fokker-Planck equation with coefficients $a_1(x)$ and $a_2(x)$ is

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = a_1(x) + \sqrt{a_2(x)}\eta(t) \tag{3.35}$$

with $\mathrm{E}(\eta(t)\eta(t')) = \delta(t-t')$.

On the other hand for Langevin equation with coefficients

$$\dot{x}(t) = a(x) + b(x)\eta(t) \tag{3.36}$$

with delta correlated noise $\eta(t)$ holds Fokker-Planck equation

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[a(x)p(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b(x)^2 p(x,t)\right] \tag{3.37}$$

We can connect for example the Wiener process with corresponding solution of Langevin equation. In the section 3.2.1 we showed that probability distribution of Wiener process is given by relation (3.1), and therefore fulfills diffusion equation

$$\frac{\partial p_W(x,t)}{\partial t} = \frac{1}{2}\frac{\partial^2 p_W(x,t)}{\partial x^2} \tag{3.38}$$

which corresponds to Smoluchowski equation with $U'(x) = 0$ and $D = \frac{1}{2\gamma^2}$. By scaling of the whole equation we can set $\gamma$ to one so we get the Langevin equation in the following way

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \eta(t) \tag{3.39}$$

with $\mathrm{E}(\eta(t)\eta(t')) = \delta(t-t')$. This equation is easy to solve and we get

$$x(t) = \int_0^t \mathrm{d}t'\eta(t'). \tag{3.40}$$

It would seem that the process x(t) has the same Fokker-Planck equation, so that $x(t)$ and $W(t)$ should be the same processes, respectively process $\eta(t)$ should be a derivative of $W(t)$. But we showed that $W(t)$ is not differentiable. This holds, because $\eta(t)$ is a delta function and that is why neither $x(t)$ is differentiable. The way how to improve current state and connect the solution of Langevin equation with the stochastic process, whose probability distribution fulfills Fokker-Planck equation will be discussed in the next section.

## 3.5 Stochastic Calculus

We stand in front of the problem, how to interpret the equation $\frac{\mathrm{d}W(t)}{\mathrm{d}t} = \eta(t)$ and how to assign to the solution of Langevin equation the appropriate stochastic process. In the first chapter we were using Langevin equation more or less formally and it was useful for us to derive some moments of the process. On the other hand, the interpretation of this equation is problematic itself, as we saw in the first chapter, where we assumed no friction in the system. This is the physical reason. From mathematical point of view $\eta(t)$ is a rapidly varying function, in other words even for small intervals the variance does not tend to zero. Both of these arguments confirm that direct interpretation of this equation is impossible. Another way is to consider a Ornstein-Uhlenbeck process $Y(t)$ for which the Fokker-Planck equation has for initial condition $Y(0) = 0$ the following form

$$\frac{\partial p_Y(x,t)}{\partial t} = \frac{\partial}{\partial x}\left[x p_Y(x,t)\right] + \frac{\partial^2 p_Y(x,t)}{\partial x^2}, \tag{3.41}$$

which gives us corresponding Langevin equation

$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = -Y(t) + \eta(t). \tag{3.42}$$

This process has covariance function $\exp(-|t-t'|)$. The idea is to replace the rapidly oscillating gaussian noise $\eta(t)$ by some process with continuous covariance function, which in limit tends to delta function. As an example we can take for example Ornstein-Uhlenbeck process with covariance function

$$\mathrm{Cov}(Y(t)Y(t')) = \frac{1}{\tau_c}\exp\left(-\frac{|t-t'|}{\tau_c}\right) \tag{3.43}$$

For $\tau_c \to 0$ we get delta function. The parameter $\tau_c$ is considered as a correlation time. Thus, the Langevin equation has now perfect meaning and the limit must be performed after calculating some measurable quantities. The problem arises when we consider a process driven by an equation with nonconstant noise term, namely nonconstant term at $\eta(t)$

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = a(x(t)) + b(x(t))\eta(t), \tag{3.44}$$

which can be formulated as a stochastic differential

$$\mathrm{d}x(t) = a(x(t))\mathrm{d}t + b(x(t))\eta(t)\mathrm{d}t = a(x(t))\mathrm{d}t + b(x(t))\mathrm{d}W(t) \tag{3.45}$$

or equivalently in integral representation

$$x(t) = x_0 + \int_0^t a(x(t'))\mathrm{d}t' + \int_0^t b(x(t'))\eta(t')\mathrm{d}t' \tag{3.46}$$

$$= x_0 + \int_0^t a(x(t'))\mathrm{d}t' + \int_0^t b(x(t'))\mathrm{d}W(t') \tag{3.47}$$

The first term is the normal Riemann integral. The interpretation of the second term remains unclear, because $\mathrm{d}W(t)$ is a function for that cannot be regular definition of a classical calculus used. The problems and the solution will be discussed in the next section.

### 3.5.1 Stochastic Integration

We can grasp the term $\int b(x(t))\mathrm{d}W(t)$ as a stochastic version of Stieltjes integral. We see, where this definition is ambiguous and show two possible interpretations.

**Definition 3.14** *Let us assume a division of the interval $[t_0, t]$ by points $t_1, t_2, \ldots, t_n$, where $t_0 \leq t_1 \leq \cdots \leq t_n \leq t$. We also choose some inner point $\tau_i$ of every interval, so $t_{i-1} \leq \tau_i \leq t_i$. Then the stochastic integral of function $G(t)$ is defined as*

$$\int_{t_0}^{t} G(t')\mathrm{d}W(t') = \underset{n\to\infty}{l.i.m.} \sum_{i=1}^{n} G(\tau_i)[W(t_i) - W(t_{i-1})]. \tag{3.48}$$

Contrary to the Riemann integral, the definition of the integral depends on the choice of points $\tau_i$. This can be demonstrated on the example of integral $\int_{t_0}^{t} W(t')\mathrm{d}W(t')$. The partial sum of the integral has form

$$\sum_{i=1}^{n} W(\tau_i)[W(t_i) - W(t_{i-1})] \tag{3.49}$$

We calculate the mean value of this integral, so the partial sum looks like

$$\sum_{i=1}^{n}[\mathrm{E}(W(\tau_i)W(t_i)) - \mathrm{E}(W(\tau_i)W(t_{i-1}))] = \sum_{i=1}^{n}(\tau_i - t_{i-1}) \tag{3.50}$$

because covariance of Wiener process is $E(W(s)W(t)) = \min(s, t)$. For $\tau_i = \alpha t_i + (1-\alpha)t_{i-1}$ we get

$$\sum_{i=1}^{n} \alpha(t_i - t_{i-1}) = \alpha(t - t_0). \tag{3.51}$$

Here we see that the choice of inner points influences the result of integration. We follow the definition of Itō integration, for which $\alpha = 0$. This integral is called *Itō stochastic integral* and is defined as

$$\int_{t_0}^{t} G(t')\mathrm{d}W(t') = \underset{n\to\infty}{l.i.m.} \sum_{i=1}^{n} G(t_{i-1})[W(t_i) - W(t_{i-1})]. \tag{3.52}$$

This definition is the most usual definition of stochastic integration. The main advantage of this definition is that Itō integrals are martingales. This can be shown straightforwardly. Let us assume a process given by a stochastic integral

$$X(t) = \int_{t_0}^{t} g(t')\mathrm{d}W(t'). \tag{3.53}$$

Than the expression

$$\mathrm{E}(X(t)|X(s)) = \underset{n\to\infty}{l.i.m.} \sum \mathrm{E}\left(g(t_{i-1})[W(t_i) - W(t_{i-1})]\right) \tag{3.54}$$

can be divided into two terms and without loss of generality we can consider a division point $t_k = s$. For points $t_i < t_k$ is

$$\mathrm{E}(g(t_{i-1})[W(t_i) - W(t_{i-1})]|X(s)) = g(t_{i-1})[W(t_i) - W(t_{i-1})] \tag{3.55}$$

Because $W(t_i)$ are directly determined by $X(s)$. For $t_i > t_k$ is $W(t_i) - W(t_{i-1})$ independent of $W(t_j)$, where $t_j < t_k$, so:

$$\mathrm{E}(g(t_{i-1})[W(t_i) - W(t_{i-1})]|X(s)) = g(t_{i-1})\mathrm{E}[W(t_i) - W(t_{i-1})] = 0. \quad (3.56)$$

That results into the fact that the integral is martingale, i.e.

$$\mathrm{E}\left(\int_{t_0}^{t} g(t')\mathrm{d}W(t') \mid X(s)\right) = \int_{t_0}^{s} g(t')\mathrm{d}W(t'). \quad (3.57)$$

We can demonstrate this property on an already mentioned example with an integral

$$\begin{aligned}
\int_0^t W(t')\mathrm{d}W(t') &= \underset{n\to\infty}{l.i.m.} \sum_{i=1}^{n} W(t_{i-1})[W(t_i) - W(t_{i-1})] \\
&= \underset{n\to\infty}{l.i.m.} \sum_{i=1}^{n} \left(W(t_{i-1})W(t_i) - W(t_{i-1})^2\right) \\
&= \underset{n\to\infty}{l.i.m.} \sum_{i=1}^{n} \left(\frac{1}{2}\left[W(t_i)^2 + W(t_{i-1})^2 - (W(t_i) - W(t_{i-1}))^2\right] - W(t_{i-1})^2\right) \\
&= \underset{n\to\infty}{l.i.m.} \sum_{i=1}^{n} \frac{1}{2}\left[W(t_i)^2 - W(t_{i-1})^2 - (W(t_i) - W(t_{i-1}))^2\right] \\
&= \frac{1}{2}W(t)^2 - \underset{n\to\infty}{l.i.m.} \frac{1}{2}\sum_{i=1}^{n}(W(t_i) - W(t_{i-1}))^2. \quad (3.58)
\end{aligned}$$

We look closer at the second term. Every term in the sum is equal to

$$\mathrm{E}\left([W(t_i) - W(t_{i-1})]^2\right) = t_i - t_{i-1}, \quad (3.59)$$

so together we get

$$\int_0^t W(t')\mathrm{d}W(t') = \frac{1}{2}(W(t)^2 - t). \quad (3.60)$$

We can see that the result is a martingale, as we showed earlier. On the other hand we see, that Itō stochastic integration gives us different results than Riemann integrals. We explain these strange properties more in the next section and show a stochastic version of change of variables.

### 3.5.2 Itō lemma

In order to derive a substitution rule for Itō integrals, we need one useful definition.

**Definition 3.15** *A function $g(t)$ is called non-anticipating function of $t$, when $g(t)$ is statistically independent of $W(s) - W(t)$ for all $(s, t)$ such that $s > t$.*

This assumption seems to be reasonable. It ensures that the function $g$ is not influenced by future values of $W(t)$. Because we mostly study models, that come from real systems, the assumption of causality should be fulfilled.

Examples of non-anticipating functions are $W(t)$, $\int W(t)\mathrm{d}W(t)$ or $\int g(t)\mathrm{d}W(t)$ for non-anticipating $g(t)$.

We now show an important property of stochastic integrals, which is the cornerstone of the whole stochastic calculus. For non-anticipating function $g(t)$ are

$$\int_{t_0}^{t} g(t')\mathrm{d}[W(t')^2] = \int_{t_0}^{t} g(t')\mathrm{d}t' \tag{3.61}$$

$$\int_{t_0}^{t} g(t')\mathrm{d}[W(t')^n] = 0 \text{ for } n \geq 3. \tag{3.62}$$

We can understand this integral as a stochastic version of Stieltjes integral. We show the first equation, the second can be calculated analogically. With notation $\Delta_i x := x(t_i) - x(t_{i-1})$ and $g_i := g(t_i)$ the expression can be showed directly from definition

$$\lim_{n\to\infty} \mathrm{E}\left[\sum_{i=1}^{n} g_{i-1}((\Delta_i W)^2 - \Delta_i t)\right]^2 =$$

$$= \lim_{n\to\infty} \mathrm{E}\left[\sum_{i=1}^{n} g_{i-1}^2((\Delta_i W)^2 - \Delta_i t)^2 + \right.$$

$$\left. +2\sum_{i=1}^{n}\sum_{j=i+1}^{n} g_{i-1}g_{j-1}((\Delta_i W)^2 - \Delta_i t)((\Delta_j W)^2 - \Delta_j t)\right].$$

Because $g$ is non-anticipating function, $\Delta_i W$ and $g_{i-1}$ are independent and the mean value is applied to every term, we get

$$= \lim_{n\to\infty} \left[\sum_{i=1}^{n} \mathrm{E}[g_{i-1}^2]\mathrm{E}[((\Delta_i W)^2 - \Delta_i t)^2] + \right.$$

$$\left. +2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathrm{E}[g_{i-1}g_{j-1}((\Delta_i W)^2 - \Delta_i t)]\mathrm{E}[((\Delta_j W)^2 - \Delta_j t)]\right].$$

In both terms of the expression we find that $\mathrm{E}[((\Delta_i W^2) - \Delta_i t)^2]$ and this term can be modified into the form

$$\mathrm{E}[((\Delta_i W)^2 - \Delta_i t)^2] = 2\Delta_i t. \tag{3.63}$$

Under some sufficient conditions (e.g. $g$ is bounded function) the whole limit vanishes. The formula for $n \geq 3$ is proved analogically, using higher moments of $W(t)$. If we rewrite the result into the corresponding differential form, we get

$$\mathrm{d}[W(t)^2] = \mathrm{d}t \tag{3.64}$$

$$\mathrm{d}[W(t)^n] = 0 \text{ for } n \geq 3. \tag{3.65}$$

With this machinery we can formulate important Itō lemma, a stochastic version of the chain rule.

**Proposition 3.16 (Itō lemma)** *Let $x(t)$ be a stochastic process fulfilling relation*

$$\mathrm{d}x(t) = a(x(t),t)\mathrm{d}t + b(x(t),t)\mathrm{d}W(t)$$

*and let $f$ be a twice differentiable function. Then*

$$\mathrm{d}[f(x(t))] = \left[a(x(t),t)f'(x(t)) + \frac{1}{2}b^2(x(t),t)f''(x(t))\right]\mathrm{d}t + b(x(t),t)f'(x(t))\mathrm{d}W(t).$$

In order to show the validity of the lemma we expand the differential of $f(x(t))$ up to second order, so

$$\begin{aligned}\mathrm{d}[f(x(t))] &= f'(x)\mathrm{d}x(t) + \frac{1}{2}f''(x)\mathrm{d}[(x(t))^2] + \mathcal{O}(\mathrm{d}[x(t)^3]) \\ &= f'(x)a(x,t)\mathrm{d}t + f'(x)b(x,t)\mathrm{d}W(t) + \frac{1}{2}f''(x)b^2(x,t)\underbrace{\mathrm{d}[W(t)^2]}_{=\mathrm{d}t} + \mathcal{O}(\mathrm{d}t^2) \\ &= [f'(x)a(x,t) + \frac{1}{2}f''(x)b(x,t)]\mathrm{d}t + f'(x)b(x,t)\mathrm{d}W(t).\end{aligned}$$

In calculations were used formulas (3.64) and (3.65). Note that the additional term $\frac{1}{2}f''(x)b(x,t)$ by the factor $\mathrm{d}t$ is caused by the stochastic nature of differentials.

### 3.5.3 Itō Integration versus Stratonovich Integration

Another possibility of defining the stochastic integral is to replace in the definition of stochastic integral the value $g(\tau)$ by an average value of $g$ in boundary points. That is the definition of Stratonovich integral

**Definition 3.17** *Let us consider an interval $[t_0,t]$ with a division $\{t_i\}_{i=1}^n$. Then Stranonovich stochastic integral is defined by following expression:*

$$\int_{t_0}^{t} G(x(t'),t')\circ\mathrm{d}W(t') = \underset{n\to\infty}{l.i.m.}\sum_{i=1}^{n} G\left[\frac{x(t_i)+x(t_{i-1})}{2}, t_{i-1}\right](W(t_i)-W(t_{i-1})).$$

We can see a little difference in the definition of Stratonovich integral. It is defined only for function of two variables $G(x,t)$. It is necessary to say that if $G$ is differentiable in $t$, then the integral does not depend on the choice of $t$ in the second argument. The reason for definition in this way is more precisely described in [15].

The advantage of Statonovich integral is that it behaves as regular deterministic integral. For example

$$\begin{aligned}\int_0^t W(t')\circ\mathrm{d}W(t') &= \underset{n\to\infty}{l.i.m.}\frac{1}{2}\sum_{i=1}^{n}(W(t_i)+W(t_{i-1}))(W(t_i)-W(t_{i-1})) = \\ &= \underset{n\to\infty}{l.i.m.}\frac{1}{2}\sum_{i=1}^{n}(W(t_i)^2 - W(t_{i-1})^2) = \frac{1}{2}W(t)^2.\end{aligned}$$

The disadvantage of Stratonovich integral is its physical interpretation. As we saw, it is not a martingale. This can be a problem for systems that have some kind of causality.

For a stochastic process obeying stochastic differential equation

$$\mathrm{d}x(t) = a(x(t),t)\mathrm{d}t + b(x(t),t)\mathrm{d}W(t) \tag{3.66}$$

or equivalently satisfying Itō integral relation

$$x(t) = x(t_0) + \int_{t_0}^{t} a(x(t'),t')\mathrm{d}t' + b(x(t'),t')\mathrm{d}W(t') \tag{3.67}$$

we find an appropriate relation in Stratonovich representation, i.e.

$$x(t) = x(t_0) + \int_{t_0}^{t} \alpha(x(t'),t')\mathrm{d}t' + \int_{t_0}^{t} \beta(x(t'),t') \circ \mathrm{d}W(t') \tag{3.68}$$

and deduce the dependence on coefficients. Firstly, we look at the Stratonovich term in the second representation. The partial sum is equal to

$$\sum_{i=1}^{n} \beta\left[\frac{x(t_i)+x(t_{i-1})}{2}, t_{i-1}\right](W(t_i) - W(t_{i-1})) \tag{3.69}$$

we expand $x(t_i)$ around $x(t_{i-1})$ up to the first term, so

$$x(t_i) = x(t_{i-1}) + \mathrm{d}x(t_{i-1}) + \mathcal{O}(\mathrm{d}x(t_{i-1})^2), \tag{3.70}$$

where $\mathrm{d}x(t_{i-1})$ fulfills the original stochastic differential equation

$$\mathrm{d}x(t_{i-1}) = a(x(t_{i-1}),t_{i-1})\Delta_i t + b(x(t_{i-1}),t_{i-1})\Delta_i W. \tag{3.71}$$

When plugged into the expression for beta, we get for $\Delta_i t \to 0$

$$\begin{aligned}\beta\left[\frac{x(t_i)+x(t_{i-1})}{2}, t_{i-1}\right] &\simeq \beta\left[x(t_{i-1}) + \frac{1}{2}\mathrm{d}x(t_{i-1}), t_{i-1}\right] = \\ = \beta(x(t_{i-1}),t_{i-1}) &+ \frac{1}{2}[\alpha(x(t_{i-1}),t_{i-1})\partial_x\beta(x(t_{i-1}),t_{i-1}) + \frac{1}{4}b^2(x(t_{i-1}),t_{i-1})]\Delta_i t \\ &+ \frac{1}{2}b(x(t_{i-1}),t_{i-1})\partial_x\beta(x(t_{i-1}),t_{i-1})\Delta_i W.\end{aligned} \tag{3.72}$$

In the integral is every term multiplied by $\Delta_i W$, so the first term gives us Itō integral, the second term vanishes, because $\mathrm{d}t\mathrm{d}W$ is dropped out, and the third term gives a contribution to deterministic integral, because $\mathrm{d}W^2(t) = \mathrm{d}t$. Together we can write

$$\int_{t_0}^{t} \beta(x(t'),t')\circ\mathrm{d}W(t') = \int_{t_0}^{t} \beta(x(t'),t')\mathrm{d}W(t') + \frac{1}{2}\int_{t=0}^{t} b(x(t'),t')\partial_x\beta(x(t'),t')\mathrm{d}t'. \tag{3.73}$$

This relation holds only for functions of argument $x(t)$ which is the solution of stochastic differential equation. We have to mention that it does not give any general relation between Itō and Stratonovich integral with arbitrary integrands. Finally we get the relation between both stochastic differential equations, namely if the stochastic process fulfills Itō differential equation

$$\mathrm{d}x(t) = a(x(t),t)\mathrm{d}t + b(x(t),t)\mathrm{d}W(t), \tag{3.74}$$

then is $x(t)$ a solution of Stratonovich differential equation

$$\mathrm{d}x(t) = [a(x(t),t) - \frac{1}{2}b(x(t),t)\partial_x b(x(t),t)]\mathrm{d}t + b(x(t),t)\circ\mathrm{d}W(t). \tag{3.75}$$

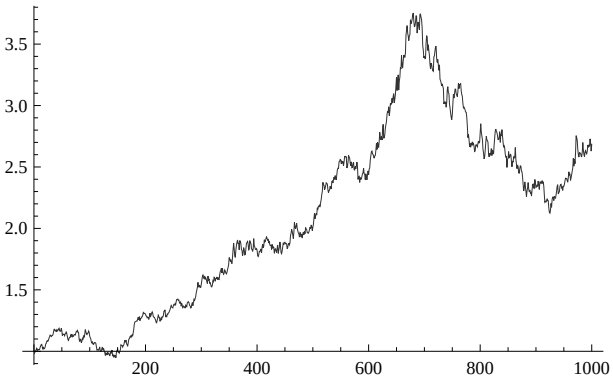



Figure 3.2: Sample path of Geommetric Brownian Motion for $\mu = 0.0015$ and $\sigma = 0.015$ for 1000 steps.

### 3.5.4 Geometric Brownian Motion

Let us return back to Itō calculus. One practical application of Itō calculus is the derivation of *Geometric Brownian motion*. We know that in the financial market financial have not series Gaussian distribution. On the other hand, the most common model assumes that logarithmical returns are normally distributed, but the mean and variance are usually not known. In many books we can find, that the price of financial series is driven by stochastic differential equation

$$\mathrm{d}P(t) = \mu P(t)\mathrm{d}t + \sigma P(t)\mathrm{d}W(t). \tag{3.76}$$

This equation comes from the fact that the strength of both terms is linearly depending on price itself, the higher the price is, the larger the growth and deviations are, so in ratio $\ln P(t_i) - \ln P(t_{i-1}) = \ln \frac{P(t_i)}{P(t_{i-1})}$ remain both quantities the same. More discussion is e.g. in [30]. We use Itō lemma and find a differential of $\ln P(t)$

$$\mathrm{d}[\ln P(t)] = \left(\mu - \frac{1}{2}\sigma^2\right)\mathrm{d}t + \sigma \mathrm{d}W(t). \tag{3.77}$$

The additional term $-\frac{1}{2}\sigma^2$ is a consequence of Itō calculus. We solve this equation and get

$$\ln P(t) = \ln P(0) + \left[\mu - \frac{1}{2}\sigma^2\right] t + \sigma W(t) \tag{3.78}$$

and finally

$$P(t) = P(0) \exp\left(\left[\mu - \frac{1}{2}\sigma^2\right] t\right) \exp(\sigma W(t)). \tag{3.79}$$

The probability distribution of $P(t)$ is log-normal with mean $\left[\mu - \frac{1}{2}\sigma^2\right] t$ and variance $\sigma^2 t$:

$$p(P,t) = \frac{1}{\sqrt{2\pi\sigma^2 P^2 t}} \exp\left(-\frac{(\ln(P/P_0) - \left[\mu - \frac{1}{2}\sigma^2\right] t)^2}{2\sigma^2 t}\right). \tag{3.80}$$

This example shows, what we saw earlier, so that manipulation with stochastic differential is not in every case intuitive and additional terms can arise.

# Chapter 4

# Memory in Physical and Economical Processes

So far, we have discussed most of the models, whose behavior depends only on the actual configuration. This represents Langevin equation with delta-correlated force, respectively Markov processes in the theory of stochastic processes. However, in the real systems (for instance physical, biological or economical ones) the very often observed behavior does not correspond to simplified behavior of these models. The behavior of large systems is often complex, exhibiting diverse properties, including memory, large fluctuations and other behavior. In this section processes with memory effects will be studied. As we will see in the next chapter, there is a close relation to scaling and large deviations.

## 4.1 Correlations and Stationarity

In practical experiments or simulations we cannot get the whole probability distribution, we can only measure statistical quantities of the ensemble. For the description of behavior of the process at one time point we have many statistical observable quantities, like mean $\mathrm{E}(x(t))$ , moments $\mathrm{E}(x(t)^n)$, central moments $\mathrm{E}(x(t)^n - \mathrm{E}(x(t))^n)$, like variance and many others. When we want to describe the dependence of two arbitrary points of a stochastic process, we introduce a well-known quantity called auto-correlation function

$$R(t_1, t_2) = \mathrm{E}(x(t_1)x(t_2)) = \int \mathrm{d}x_1 \mathrm{d}x_2 x_1 x_2 p(x_1, t_1; x_2, t_2). \tag{4.1}$$

Because there is often a non-zero mean, we define also an autocovariance function

$$C(t_1, t_2) = R(t_1, t_2) - \mathrm{E}(x(t_1))\mathrm{E}(x(t_2)). \tag{4.2}$$

We would like to focus especially on stationary processes, because it is easy to deal with those processes and we have seen that many of important stochastic processes can be modeled by stationary stochastic processes. In the previous chapter we introduced a concept of strict stationarity. The process is strictly stationary, if the probability distribution is invariant under the time shift. This

definition is unfortunately too strict, because we usually do not know the whole distribution. We introduce weaker condition for stationarity instead, but all favorable properties remain.

**Definition 4.1** *A stochastic process $x(t)$ is stationary in the wide sense, if:*

1. $\mathrm{E}(x(t)) = \mu = const.$,
2. $\mathrm{E}(x(t_1)x(t_2)) = R(|t_1 - t_2|)$.

The consequence is that variance is also time-independent, because $R(0) = \mathrm{E}(x(t)^2)$ for arbitrary $t$ and therefore $C(0) = \sigma^2$. Now we can ask the question, for which time difference are correlations too weak, so we can consider that there exists some $\tau_c$ that for sufficiently large difference $\tau_c < |t - t'|$ are variables $x(t)$ and $x(t')$ practically independent. Or if $\tau_c$ exists at all. Let us look at a discrete case. We consider a discrete stationary process $x_i$ of identically distributed variables and we also consider the sum $S_n = \sum_{i=1}^{n} x_i$. Let us assume $\mathrm{E}(x_i) = 0$. Then the variance of the sum is

$$\mathrm{E}(S_n^2) = \sum_{i=1}^{2} \mathrm{E}(x_i^2) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathrm{E}(x_i x_j) = n\mathrm{E}(x_i^2) + 2\sum_{i=1}^{n}(n-i)C(i)$$
$$= n\mathrm{E}(x_i^2) + 2n\sum_{l=1}^{n}\left(1 - \frac{l}{n}\right)C(l) \quad (4.3)$$

According to the Central limit theorem (theorem (A.1) in appendix) should the sum for independent variables grow as $N \cdot const.$ This means that from some n should the second term be practically constant, so $C(l)$ should scale as $l^{-\nu}$, where $\nu > 1$. For large $n$ can be the sum approximated as

$$2\sum_{l=1}^{n}\left(1 - \frac{l}{n}\right)C(l) \approx 2\int_0^\infty C(\tau)\mathrm{d}\tau, \quad (4.4)$$

so if there exist any time scale $\tau_c$, then must be this integral finite, and if correlations cannot be neglected for any scale, then the integral is infinite. This leads us to division of random processes. Processes with *short-range correlations* are processes for which $\int_0^\infty C(\tau)\mathrm{d}\tau$ is finite and *long-range correlated* processes are such processes for which $\int_0^\infty C(\tau)\mathrm{d}\tau = +\infty$ or is indeterminable. The example of the first class is a process with auto-covariance function

$$C(\tau) = \exp\left(-\frac{\tau}{\tau_c}\right), \quad (4.5)$$

because

$$\int_0^\infty C(\tau)\mathrm{d}\tau = \int_0^\infty \mathrm{d}\tau \exp\left(-\frac{\tau}{\tau_c}\right) = \tau_c. \quad (4.6)$$

The example of this process is the Ornstein-Uhlenbeck process. If the covariance function scales like $\tau^{\nu-1}$ for $\nu > 0$ the integral diverges and the process is long-range correlated.

## 4.2 Power Spectrum

Another useful characteristics of stochastic processes that helps to recognize memory effects is a Fourier image of the stochastic process. We define a *spectral density* of a stochastic process $x(t)$ as

$$S(\omega) = \lim_{T\to\infty} \frac{1}{2\pi T} \mathrm{E}\left( \left| \int_0^T x(t) \exp(-i\omega t) \mathrm{d}t \right|^2 \right). \tag{4.7}$$

The dependence on $\omega$ helps us to recognize the strength of the memory. The meaning of power spectrum arise clearly for stationary processes.

### 4.2.1 Wiener-Kchinchin Relation

The choice of power spectrum was not accidental. The relation to correlations and memory effects is remarkable for stationary processes, for which a relation between power spectrum and correlation function can be formulated.

**Proposition 4.2 (Wiener-Kchinchin)** *For a wide sense stationary process is*

$$S(\omega) = \mathcal{F}[R(\tau)](\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\tau) \exp(-i\omega\tau) \mathrm{d}\tau. \tag{4.8}$$

We can show the proposition straightforwardly from definition, because

$$\begin{aligned}
&\lim_{T\to\infty} \frac{1}{2\pi T} \mathrm{E}\left( \int_0^T \mathrm{d}t \int_0^T \mathrm{d}t' x(t)x(t') \exp(-i\omega(t'-t)) \right) = \\
&= \lim_{T\to\infty} \frac{1}{2\pi T} \int_0^T \mathrm{d}t \int_0^T \mathrm{d}t' \mathrm{E}(x(t)x(t')) \exp(-i\omega(t'-t)) \\
&= \lim_{T\to\infty} \frac{1}{2\pi T} \int_0^T \mathrm{d}t \int_0^T \mathrm{d}t' R(t'-t) \exp(-i\omega(t'-t)).
\end{aligned}$$

We integrate over a square $(0,T)\times(0,T)$ where we can integrate firstly over constant $t'-t=\tau$ and then over $\tau$. We therefore change the variables for $\tau = t'-t$ and $z = \frac{1}{2}(t+t')$, so we get

$$\lim_{T\to\infty} \frac{1}{2\pi T} T \int_{-T}^{T} \mathrm{d}\tau R(\tau) \exp(-i\omega\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\tau R(\tau) \exp(-i\omega\tau).$$

The original formula is very difficult to calculate, whereas the latter one can be performed easily. This allows us to calculate some important power spectrums.

- Gaussian (white) noise: we know that Gaussian noise is the stationary process with correlation function $D_0\delta(\tau)$. The power spectrum is therefore

$$S_\eta(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} D_0\delta(\tau) \exp(-i\omega\tau) = \frac{D_0}{2\pi}. \tag{4.9}$$

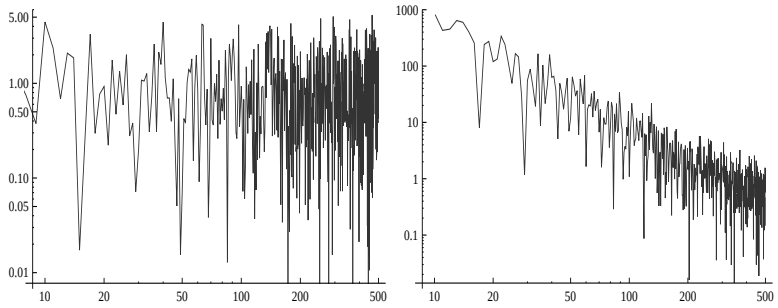



Figure 4.1: Log-Log plot of power spectrum of Gaussian noise (left) and Wiener process (right). The power spectrum of Gaussian noise behaves as constant, power spectrum of Wiener process decays as $\omega^{-2}$.

- Short-range correlated process: If we have a process with a exponential decay in correlations $\exp\left(-\frac{|\tau|}{\tau_c}\right)$ we get a power spectrum

$$S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\tau \exp\left(-\frac{|\tau|}{\tau_c} - i\omega\tau\right)$$
$$= \frac{1}{2\pi}\left(\frac{\tau_c}{1+i\tau_c\omega} + \frac{\tau_c}{\tau_c - i\tau_c\omega}\right) = \frac{\tau_c^2}{2\pi(1+\tau_c^2\omega^2)}. \tag{4.10}$$

  The spectrum behaves for $\omega\tau_c \ll 1$ as a constant and for $\omega\tau_c \gg 1$ decays as $\omega^{-2}$.

- Wiener process: although is Wiener process not a stationary process we can estimate the spectrum from knowledge of white noise. We know that in some sense holds $\frac{\mathrm{d}W(t)}{\mathrm{d}t} = \eta(t)$, so in Fourier image we have

$$-i\omega\mathcal{F}[W](\omega) = \mathcal{F}[\eta](\omega),$$

  so the power spectrum behaves as

$$S_W(\omega) \propto \omega^{-2}. \tag{4.11}$$

### 4.2.2 1/f Noise

Up to now we have seen only processes with constant power spectrum, or power spectrum decaying as $\omega^{-2}$. We now assume a process with power spectrum

$$S(\omega) = \frac{c}{\omega^\eta}, \tag{4.12}$$

where $0 < \eta < 2$. These processes are often called $1/f$ noises, or generally $1/f^\eta$ noises. Here $f$ denotes a frequency variable and the name wants us to say, how fast decays the power spectrum. It can be showed that these processes cannot be stationary. We use a result from [20], for short observation time we get that corresponding correlation function scales as

- $R(\tau) \propto |\tau|^{\eta-1}$ for $0 < \eta < 1$, $\tau > 0$

- $R(t_2, \tau) \propto \ln 4t_2 - \ln|\tau|$ for $0 < \tau \ll t_2$, $\eta = 1$
- $R(t_2, \tau) \propto t_2^{\eta-1} - C|\tau|^{\eta-1}$ for $1 < \eta < 2$.

In all three cases we see, that there does not exist any typical time scale, as for exponentially decaying correlation functions. That means that all of these processes have long memory and that gives us useful tool to recognize, whether the process has long memory or not.

## 4.3 Fractional Stochastic Processes

In this section we would like to show another approach leading to memory effects. This will be the fractional calculus. We firstly introduce the necessary mathematical formalism and then we show, how this calculus can be applied in order to create these kinds of processes. We note here that fractional calculus will be important also in the next section in connection to the scaling properties and super-diffusion behavior.

### 4.3.1 Fractional Integration and Differentiation

In this section we would like to extend the classical differentiation and integration operators from classical calculus to operators with non-natural numbers. We demand that for natural numbers we get classical derivatives and integrals. We begin from the classical formula for repeated integral

$$\int_{x_0}^{x}\int_{x_0}^{x_1}\dots\int_{x_0}^{x_{n-1}} f(x_n)\mathrm{d}x_n\dots\mathrm{d}x_1 \tag{4.13}$$

and rewrite the expression with help of Cauchy integral formula

$$\frac{1}{(n-1)!}\int_{x_0}^{x}(x-y)^{n-1}f(y)\mathrm{d}y. \tag{4.14}$$

This formula can be easily generalized for non-integer values, so for $\alpha > 0$ we define a fractional integral as

$${}_{x_0}\mathcal{I}^{\alpha}[f](x) := \frac{1}{\Gamma(\alpha)}\int_{x_0}^{x}(x-y)^{\alpha-1}f(y)\mathrm{d}y. \tag{4.15}$$

We show that on space $L^1(a,b)$ is the operator bounded. For arbitrary $a < x_0 < b$ we get

$$\|\mathcal{I}^{\alpha}_{x_0}f\|_1 = \frac{1}{|\Gamma(\alpha)|}\int_a^b\left|\int_{x_0}^{x}(x-y)^{\alpha-1}f(y)\mathrm{d}y\mathrm{d}x\right| \le \frac{|b-a|^{\alpha}}{\alpha|\Gamma(\alpha)|}\|f\|_1. \tag{4.16}$$

We also show that a set of operators $\{\mathcal{I}^{\alpha}_{x_0}\}_{\alpha>0}$ is a semigroup. For some $\alpha$, $\beta$

$$\mathcal{I}^{\alpha}_{x_0}\circ\mathcal{I}^{\beta}_{x_0}[f](x) = \frac{1}{\Gamma(\alpha)}\int_{x_0}^{x}\mathrm{d}y(x-y)^{\alpha-1}\frac{1}{\Gamma(\beta)}\int_{x_0}^{y}\mathrm{d}z(y-z)^{\beta-1}f(z). \tag{4.17}$$

In the next we use a special version of Dirichlet formula

$$\int_a^b\mathrm{d}y\int_a^y G(y,z)\mathrm{d}z = \int_a^b\mathrm{d}z\int_z^b G(y,z)\mathrm{d}y \tag{4.18}$$

where $G(y,z) = (x-y)^{\alpha-1}(y-z)^{\beta-1}f(z)$. So we get

$$\frac{1}{\Gamma(\alpha)\Gamma(\beta)}\int_{x_0}^{x} f(z)\mathrm{d}(z)\int_{z}^{x}(x-y)^{\alpha}(y-z)^{\beta-1}\mathrm{d}y. \tag{4.19}$$

We introduce a substitution $t = \frac{y-z}{x-z}$ and after this substitution we get

$$\begin{aligned}\frac{1}{\Gamma(\alpha)\Gamma(\beta)}\int_{x_0}^{x} f(z)\mathrm{d}z(x-z)^{\alpha+\beta-1}\int_{0}^{1}(1-t)^{\alpha-1}t^{\beta-1}\mathrm{d}t = \\ = \frac{1}{\Gamma(\alpha)\Gamma(\beta)}\int_{x_0}^{x} f(z)\mathrm{d}(z)(x-z)^{\alpha+\beta-1}B(\alpha,\beta),\end{aligned} \tag{4.20}$$

so together we actually get $\mathcal{I}_{x_0}^{\alpha+\beta}[f](x)$. We can now define also a fractional derivative operator as

$$\mathcal{D}_{x_0}^{\beta}[f](x) := \frac{\mathrm{d}^{\lfloor\beta\rfloor+1}}{\mathrm{d}^{\lfloor\beta\rfloor+1}x}\left(\mathcal{I}_{x_0}^{\lfloor\beta\rfloor-\beta+1}[f]\right)(x) \tag{4.21}$$

which is well defined for sufficiently smooth function.

We look at another interesting representation of fractional integrals and derivatives and concretely at the representation through Fourier transform. Unfortunately the class of functions, for which can be fractional integrals represented through Fourier transform, covers only a fraction of functions, for which a fractional integral or derivative can be made. But on the other hand, Fourier representation is very easy and elegant. We can formulate it as a lemma.

**Lemma 4.3** *We consider* $0 < \alpha < 1$ *and* $f \in L^1(a,b)$. *Then*

$$\mathcal{F}\left[\mathcal{I}_{-\infty}^{\alpha}[f]\right](k) = (ik)^{-\alpha}\mathcal{F}\left[f\right](k). \tag{4.22}$$

In order to show the validity of the lemma we look at integral

$$\begin{aligned}\int_{-t}^{t}\mathrm{d}x e^{-ikx}\,\mathcal{I}_{-\infty}^{\alpha}[f](x) = \int_{-t}^{t}\mathrm{d}x e^{-ikx}\frac{1}{\Gamma(\alpha)}\int_{-\infty}^{x}(x-y)^{\alpha-1}f(y)\mathrm{d}y = \\ = \left|\begin{array}{c} x-y=l \\ -\mathrm{d}y=\mathrm{d}l \end{array}\right| = \int_{-t}^{t}\mathrm{d}x\int_{0}^{\infty}\mathrm{d}l e^{-ikx}l^{\alpha-1}f(x-l) = \\ = \left|\begin{array}{c} x-l=z \\ \mathrm{d}z=\mathrm{d}l \end{array}\right| = \int_{0}^{\infty}\mathrm{d}\,l\;l^{\alpha-1}e^{-ikl}\int_{-t-l}^{t-l}\mathrm{d}z f(z)e^{-ikz}\end{aligned}$$

Now we make a limit transition in the principal value. We assume the first integral as a limit of a finite integral $\int_0^a$ and change the limit order. With that we get that in the principal value meaning is the last expression equal to

$$\stackrel{p.v.}{=} \mathcal{F}[f](k)\int_{0}^{\infty}\mathrm{d}\,l\;l^{\alpha-1}e^{-ikl} = (ik)^{-\alpha}\mathcal{F}[f](k). \tag{4.23}$$

We can easily extend the representation for some class of functions for which it does make sense, extend also to exponents $\alpha > 1$ and for fractional derivatives, as well. More details about this are in [6], but generally is possible to write for suitable functions

$$\mathcal{F}\left[\mathcal{I}_{-\infty}^{\alpha}[f]\right](k) = (ik)^{-\alpha}\mathcal{F}\left[f\right](k) \tag{4.24}$$

$$\mathcal{F}\left[\mathcal{D}_{-\infty}^{\beta}[f]\right](k) = (ik)^{\beta}\mathcal{F}\left[f\right](k). \tag{4.25}$$

### 4.3.2 Mittag-Leffler Functions

Let us look at special class of functions that are closely related to fractional derivatives. For exponential function we can find a series expansion in the following form:

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{x^n}{\Gamma(n+1)}. \tag{4.26}$$

We introduce a simple generalization of this series expansion. The function is called *Mittag-Leffler function* and is given by this expansion

$$E_\mu(x) = \sum_{n=0}^{\infty} \frac{x^n}{\Gamma(\mu n+1)}, \tag{4.27}$$

for $\mu > 0$. The function is not accidentally called generalized exponential function, the name comes from property that it is an eigenfunction of fractional differential equation

$$\mathcal{D}_0^\mu[E_\mu(x)] = E_\mu(x). \tag{4.28}$$

In order to show this relation, we calculate the fractional derivative of function $x^n$. For simplicity we show it for $\mu \in (0,1)$, for other values is the calculation analogous.

$$\mathcal{D}_0^\mu[x^n] = \frac{\mathrm{d}}{\mathrm{d}x}\left(\mathcal{I}_0^{1-\mu}[x^n]\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{\Gamma(1-\mu)}\int_0^x (x-y)^{1-\mu} y^n \mathrm{d}y\right).$$

Here we use the same substitution as in (4.20) and get

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{\Gamma(1-\mu)} x^{n+1-\mu} B(1-\mu, n+1)\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\Gamma(n+1)}{\Gamma(n+2-\mu)} x^{n+1-\mu}\right) = \frac{\Gamma(n+1)}{\Gamma(n+1-\mu)} x^{n-\mu}.$$

With that can be easily checked that Mittag-Leffler function really fulfills the equation (4.28). There are a few examples of Mittag-Leffler function expressible as elementary function:

- $E_1(x) = \exp(x)$,
- $E_2(x^2) = \cosh(x)$,
- $E_2(-x^2) = \cos(x)$.

There are some more cases for those we get elementary functions, more examples and properties are in [32]. From now on, we are interested in the behavior of $E_\mu(-\alpha_\mu t^\mu)$, where $\alpha_\mu$ is some constant. This expression will be helpful in the next section. Firstly, we look at the behavior for small $t$:

$$E_\mu(-\alpha_\mu t^\mu) \simeq \frac{1}{\Gamma(1)} - \frac{\alpha_\mu t^\mu}{\Gamma(\mu+1)} \simeq \frac{1}{\Gamma(1)} - \frac{c\alpha_\mu t^\mu}{\Gamma(2)} \simeq \exp(-c\alpha_\mu t^\mu). \tag{4.29}$$

Thus, for very small $t$ decays the function $E_\mu$ exponentially. On the other hand for large values of $t$ the function behaves as

$$E_\mu(-\alpha_\mu t^\mu) \simeq \frac{t^{-\mu}}{\alpha_\mu \Gamma(1-\mu)}. \tag{4.30}$$

We used the result from [32], so for large values it decays with power law. For us one more property will be important, namely Laplace transform (again from [32])

$$\mathcal{L}[E_\mu(-\alpha_\mu t^\mu)](s) = \frac{s^{\mu-1}}{s^\mu + \alpha_\mu} = \frac{1}{s + \alpha_\mu s^{1-\mu}}. \tag{4.31}$$

In the next section we see the application of fractional calculus and Mittag-Leffler function on the example of Langevin equation with memory.

### 4.3.3 Langevin Equation with Memory

Up to now we were interested only in processes with random noises that depended only on the present configuration. Now we take into account also the situation, when the fluctuating force and friction have non-zero correlations. We assume general Langevin equation with fluctuation force $F(t)$ in the form

$$m\ddot{x}(t) + m\int_0^t \mathrm{d}t'\gamma(t-t')\dot{x}(t') + U'(x) = F(t) \tag{4.32}$$

with conditions on the force $F(t)$

$$\mathrm{E}(F(t)) = 0, \tag{4.33}$$

$$\mathrm{E}(F(t)F(t')) = k_B T\gamma(t-t'). \tag{4.34}$$

The last equation is the generalization of fluctuation-dissipation theorem for systems with the memory. We can observe that the whole system, i.e. both fluctuations a and friction have characteristic memory $\gamma(t)$, also called *memory kernel*.

Now let us consider a Langevin equation for a free particle ($U'(x) = 0$) with the memory kernel

$$\mathrm{E}(F(t)F(t')) = \frac{mk_BT}{\Gamma(1-\beta)} \cdot (t-t')^\beta = \gamma_\beta \cdot (t-t')^\beta. \tag{4.35}$$

In order to find a solution we transform the equation (4.32) into its Laplace image and we exploit the well known rules for the Laplace transform of the convolution and derivative (we follow the natural assumption $\gamma(t') = 0$ for $t' < 0$):

$$\mathcal{L}[\gamma * v](s)] = \mathcal{L}\left[\int_{-\infty}^{\infty} \mathrm{d}t'\gamma(t-t')v(t')\right](s) = \mathcal{L}[\gamma](s)\mathcal{L}[v](s), \tag{4.36}$$

$$\mathcal{L}[\dot{x}](s) = s\mathcal{L}[x](s) - x_0. \tag{4.37}$$

The Laplace image of the Langevin equation has following form:

$$m\left(s^2\mathcal{L}[x](s) - sx_0 - v_0\right) + ms\mathcal{L}[\gamma](s)\left(s\mathcal{L}[x](s) - x_0\right) = \mathcal{L}[F](s). \tag{4.38}$$

After some manipulations we get the solution in the Fourier image:

$$\mathcal{L}[x](s) = \frac{x_0}{s} + \frac{v_0}{s(s + \mathcal{L}[\gamma](s))} + \frac{\mathcal{L}[F](s)}{ms(s + \mathcal{L}[\gamma](s))} \tag{4.39}$$

We define a function $\mathcal{B}(t)$, such as

$$\mathcal{L}[\mathcal{B}](s) := \frac{1}{s(s + \mathcal{L}[\gamma](s))}. \tag{4.40}$$

We can convert the solution back to variable $t$ and write

$$x(t) = x_0 + v_0\mathcal{B}(t) + \frac{1}{m}\int_0^t \mathrm{d}t'\mathcal{B}(t - t')F(t'). \tag{4.41}$$

We find a solution for $v(t)$ by differentiating of the solution for $x(t)$

$$v(t) = v_0\mathcal{C}(t) + \frac{1}{m}\int_0^t \mathrm{d}t'\mathcal{C}(t - t')F(t'), \tag{4.42}$$

where $\mathcal{C}(t) = \dot{\mathcal{B}}(t)$. The function $\mathcal{C}(t)$ is the normalized correlation function, because

$$E(v(t)v(0)) = v_0^2\mathcal{C}(t) \tag{4.43}$$

and its Laplace transform is

$$\mathcal{L}[\mathcal{C}](s) = \frac{1}{s + \mathcal{L}[\gamma](s)}. \tag{4.44}$$

Now, we consider two extreme cases of memory:

- no memory: $\mathcal{C}_0(t) = \gamma\delta(t) \Rightarrow \mathcal{L}[\mathcal{C}_0](s) = \gamma$
- constant memory: $\mathcal{C}_1(t) = \gamma = const. \Rightarrow \mathcal{L}[\mathcal{C}_1](s) = \gamma s^{-1}$

The first memory kernel belongs to a system with no memory and the second to a system, where every past configuration influences the future behavior with the same magnitude. In between are processes that depend also on the past values, but with descending strength. This property can be well characterized again in the Laplace image, so we define

$$\mathcal{L}[\mathcal{C}_\alpha(t)](s) := \frac{\gamma}{s^\alpha} \Rightarrow \mathcal{C}_\alpha(t) = \frac{\gamma\alpha t^{\alpha-1}}{\Gamma(\alpha)} \tag{4.45}$$

for $\alpha \in [0, 1]$. In our case we can recognize that the correlation function is a the Mittag-Leffler function because

$$\mathcal{L}[C](s) = \frac{1}{s + \gamma_\beta s^{\beta-1}} \tag{4.46}$$

hence

$$\mathcal{C}(t) = E_{2-\beta}(-\gamma_\beta t^{2-\beta}) \simeq \frac{t^{\beta-2}}{\gamma_\beta\Gamma(\beta - 1)} \qquad \text{for large } t. \tag{4.47}$$

The mean square displacement is from $x(t) - x(0) = \int \mathrm{d}t'v(t')$ equal to

$$\begin{aligned}\mathrm{Var}(x(t)) = \int_0^t \mathrm{d}t_1 \int_0^t \mathrm{d}t_2 \mathrm{E}(v(t_1)v(t_2)) &= 2\int_0^t \mathrm{d}t_1 \int_0^{t_1} \mathrm{d}t_2 \mathrm{E}(v(t_2)v(0)) \\ &= 2v_0^2 \int_0^t \mathrm{d}t_1 \int_0^{t_1} \mathrm{d}t_2 \mathcal{C}(t_2) \simeq \frac{2v_0^2 t^\beta}{\gamma_\beta\Gamma(\beta + 1)}.\end{aligned} \tag{4.48}$$

We can distinguish three cases of diffusion:

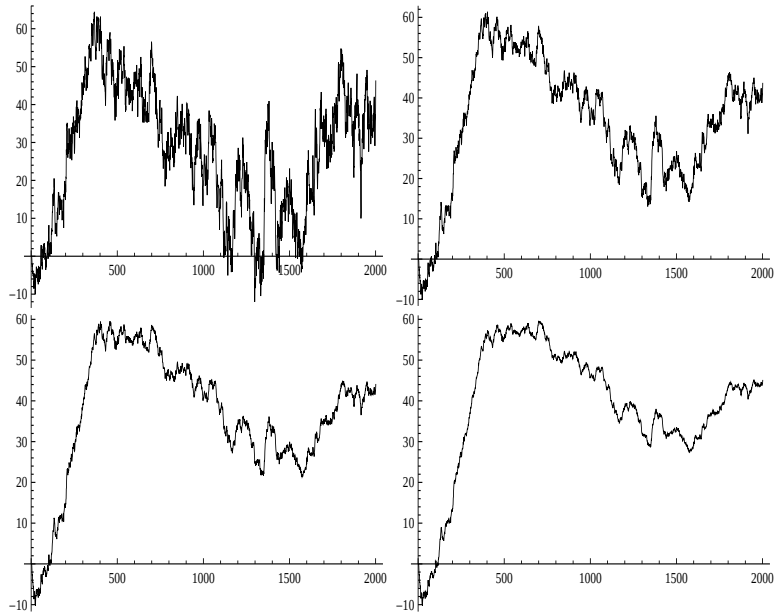



Figure 4.2: Sample functions of fBM for $H$=0.3, 0.5, 0.6 and 0.7. For 0.5 we get Brownian motion. Compared to that we see that for $H > 0.5$ are the sample functions smoother and the behavior is persistent. For $H < 0.5$ we get rough sample functions and anti-persistent behavior.

- $\beta = 1$: the mean square displacement is $\mathrm{Var}(x(t)) \sim t$, which is the same as for Langevin equation without memory, so $\mathcal{C}(t) \sim \exp(-ct)$. Velocities are therefore for large times uncorrelated.
- $\beta > 1$: $\mathcal{C}(t) > 0$, we get superdiffusive behavior, velocities are positively correlated.
- $\beta < 1$: $\mathcal{C}(t) < 0$, we get subdiffusive behavior, velocities are negatively correlated.

We have seen that the presence of memory can cause different behavior. The behavior of the process can be either persistent, which means that movement in some direction supports future movement in the same direction, or anti-persistent, which means that the movement in one direction causes the opposite movement. In the first case the trajectory is more straightforward, which means that in the same time the particle gets further than in case of an uncorrelated process. In the second case contributions cancel out and the particle remains probably nearer to the original position. In the next section we discuss similar processes that exhibit superdiffusive and subdiffusive behavior and in the next chapter we show that there exists a connection between memory effects and scaling and how can be this behavior revealed.

## 4.4 Fractional Brownian Motion

Another approach to processes with memory, that brings similar results as Langevin equation with memory, is generalization of Wiener process. With this

generalization came B.B.Mandelbrot [26]. It comes from an idea of fractional stochastic integration of Wiener process.

**Definition 4.4** *A Fractional Brownian motion (fBM) is a stochastic process defined as a stochastic integral*

$$W_H(t) := \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-s)^{H-\frac{1}{2}}\mathrm{d}W(t) \tag{4.49}$$

*where* $H \in [0,1]$.

We calculate the variance of Fractional Brownian motion

$$\mathrm{E}(W^H(t)W^H(t)) = \mathrm{E}\left(\frac{1}{\Gamma(H+\frac{1}{2})^2}\int_0^t\int_0^t (t-s)^{H-\frac{1}{2}}(t-s')^{H-\frac{1}{2}}\mathrm{d}W(s)\mathrm{d}W(s')\right)$$

$$= \frac{1}{\Gamma(H+\frac{1}{2})^2}\int_0^t\int_0^t (t-s)^{H-\frac{1}{2}}(t-s')^{H-\frac{1}{2}}\mathrm{E}(\mathrm{d}W(s)\mathrm{d}W(s'))$$

$$= \frac{1}{\Gamma(H+\frac{1}{2})^2}\int_0^t\int_0^t (t-s)^{H-\frac{1}{2}}(t-s')^{H-\frac{1}{2}}\delta(s-s')\mathrm{d}s\mathrm{d}s'$$

$$= \frac{1}{\Gamma(H+\frac{1}{2})^2}\int_0^t (t-s)^{2H-1}\mathrm{d}s = \frac{t^{2H}}{2H\Gamma(H+\frac{1}{2})^2}.$$

Sometimes is the fBM defined in a slightly different way, directly through definition of variance. One can say that fBM is a stochastic process with following properties:

- $W_H(0) \overset{a.s.}{=} 0$,
- $W_H(t)$ has stationary independent increments,
- $\mathrm{Var}(W_H(t)) = \frac{t^{2H}}{H\Gamma(H+\frac{1}{2})^2}$.

Similarly to variance can be autocorrelation function calculated as [26]

$$E(W^H(t)W^H(s)) = \frac{1}{2H\Gamma(H+\frac{1}{2})^2}(s^{2H}+t^{2H}-|s-t|^{2H}). \tag{4.50}$$

We see that it corresponds to the correlation functions calculated in the previous chapter, so the coefficient $2H$ corresponds to the exponent by $t$. Because correlations are non-zero and have power-law character, we get a class of non-Markovian processes with long memory. Similarly to the last section, for $H > \frac{1}{2}$ we get positive correlations and for $H < \frac{1}{2}$ we get negative correlations, which represent the superdiffusive behavior, resp. subdiffusive behavior.

Processes that we have introduced in this chapter are only an elementary approach to more realistic processes observed in many experiments from physics, biology or economics. In the next chapter we show another concept that will help us to construct better models and we will see that it is also partially related to the processes with memory.

# Chapter 5

# Scaling Properties of Random Processes

In this chapter we want to show, how the scaling is important in stochastic processes and what properties can be determined by scaling. We will see that scaling is a powerful tool, that causes nontrivial behavior of systems, because it discovers their inner structure. We will see, that processes with scaling properties exhibit both the memory effects and extreme behavior in comparison to other processes. The concept of scaling will lead us to self-similarity of processes and this can be well described by fractal geometry, which will be a springboard to further chapters.

## 5.1 Scaling and Memory

In the previous chapter we were talking about difference between different types of memories and corresponding correlation functions. Let us look at the problematic once more from a different point of view that will stand as a motivation for the next chapter. We have seen two types of correlation function. Firstly correlation functions that decay exponentially. These systems exhibit short-range memory effects, there is a typical scale, for which the values are independent from each other. This can be nicely understood if we rescale the variable $t$ to $t \mapsto \kappa t$, then the function $\gamma$, that determines the behavior of the system is

$$\gamma_c(t) = e^{-\frac{t}{\tau_c}} \Rightarrow \gamma_c(\kappa t) = e^{-\frac{\kappa t}{\tau_c}} \simeq 0 \tag{5.1}$$

for $\kappa \gg \tau_c$, so we see, that there exist a scale and values further than this scale are practically independent. But when we assume a polynomial behavior, then

$$\gamma_p(t) = \frac{\mu t^{\mu-1}}{\Gamma(\mu)} \Rightarrow \gamma_p(\kappa t) = \frac{\mu(\kappa t)^{\mu-1}}{\Gamma(\mu)} = \kappa^{\mu-1}\gamma_p(t). \tag{5.2}$$

We see that for large times remains the correlation function the same, just rescaled. That is the main difference between exponential decay and polynomial behavior. Generally we can distinguish between exponential and polynomial behavior in certain systems. We can observe by systems with power-law a typical long-range behavior and scale-invariance, which is also called self-similar. The proper definition of self-similarity comes in one of the next sections.

## 5.2 A Way beyond Central Limit Theorem

Central limit theorem (theorem A.1 in appendix) states that for sum of independent variables with finite variance is Gaussian distribution the limiting distribution. That means that for long times we can assume that for every system composed of many little random steps added together the sum of all that steps behaves as it was Gaussian. Nevertheless, the assumption of no correlations and finite variance can be limiting for many applications, where studied systems perform complex behavior with a row of nontrivial phenomena. In the last section we overcame the absence of memory, in this section we will be interested in processes, for which the variance is not finite. We will see that some properties of both processes are common.

In the latter case there is shown in appendix that there is a special class of limiting distributions called Lévy distributions that can be expressed in the form (A.5) with asymptotic behavior

$$L_\alpha(x) \simeq \frac{l_\alpha}{|x|^{1+\alpha}} \tag{5.3}$$

where

$$l_\alpha = \frac{\gamma\Gamma(\alpha+1)}{\pi}\sin\left(\frac{\pi\alpha}{2}\right) \tag{5.4}$$

and $\gamma$ is one of parameters of Lévy distribution. The expression comes from an asymptotic expansion of Lévy distribution (A.8). These distributions have interesting properties

- **sharper central peak**: Compared to Gaussian distribution have Lévy distributions sharper maximum in the central part.
- **fat tails**: Because Lévy distributions decay polynomially for large values, they decay therefore much slower than Gaussian distribution.

The second property is connected to the infiniteness of the second moment of the distribution, that means that large deviations from the center are much more probable than in case of normal distribution. Both properties are pictured in figure 5.1. In the next section we focus on processes that are generated by Lévy distributions and show some interesting properties of those processes.

### 5.2.1 Lévy Flights

In the third chapter we defined stable processes, but we were dealing only with processes with Gaussian distribution. The reason is that many models from other branches deal with Gaussian distribution, because dealing with other types of distributions is usually very difficult and as we saw, Gaussian distribution describes many systems without extreme behavior quite well because of central limit theorem. Our approach is to aim to processes that are driven by Lévy distributions. We adapt the definition of stable processes to case, when the stable distribution is Lévy distribution. Let us start with an auxiliary definition

**Definition 5.1** *Let $X_1$,$X_2$ be i.i.d. copies of random variable $X$ with the probability distribution $p(x)$. Then is this variable called strictly $\alpha$-stable if for any*

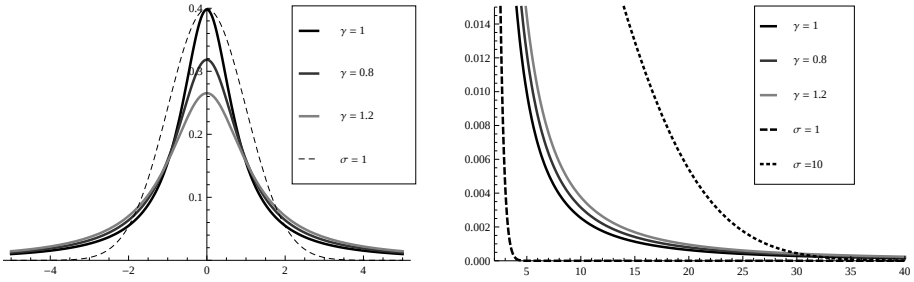



Figure 5.1: Probability distribution of Lévy stable distribution (concretely Cauchy distribution) and Gaussian distribution. On the left figure we can see the central part of distribution with sharper peak of Lévy distribution and on the right figure we see fat tails of Lévy distribution driven by power law and comparison with Gaussian distributions that decay faster to zero even for large $\sigma$'s.

$a, b$

$$a^{\frac{1}{\alpha}} X_1 + b^{\frac{1}{\alpha}} X_2 \stackrel{d}{=} (a+b)^{\frac{1}{\alpha}} X. \tag{5.5}$$

This property is direct generalization of the property of Wiener process, because from basic property of Gaussian distribution

$$\lambda \mathcal{N}(0, \sigma^2) = \mathcal{N}(0, \lambda^2 \sigma^2) \tag{5.6}$$

$$\mathcal{N}(0, \mu\sigma^2) + \mathcal{N}(0, \nu\sigma^2) = \mathcal{N}(0, (\mu+\nu)\sigma^2) \tag{5.7}$$

is easy to show that for normally distributed variable $w$ is

$$a^{1/2} w + b^{1/2} w \stackrel{d}{=} (a+b)^{1/2} w. \tag{5.8}$$

That perfectly corresponds to Gaussian limit, because for $\alpha = 2$ becomes the Lévy distribution into Gaussian distribution. It is also obvious from the form of characteristic function of Lévy distributions that strictly $\alpha$-stable variables are variables with symmetric Lévy distributions $L_\alpha$ with mean value 0 (i.e. $c = 0$), because when we add two random variables, we must do a convolution of probability distributions and for Fourier images it means to multiply them, so for functions

$$\mathcal{F}[L_\alpha(x)](k) = \exp(-\gamma |k|^\alpha) \tag{5.9}$$

this causes the right multiplication. The reason of redefining Wiener process is that we cannot define a Lévy process with variance, because it is infinite, so we help us with scaling properties. We can therefore define a process, that will be a generalization of Wiener process, but with Lévy increments.

**Definition 5.2** *A stochastic process $L_\alpha(t)$ is called Lévy motion, or alternatively Lévy flight if:*

- $L_\alpha(0) \stackrel{a.s.}{=} 0$,
- *Increments of $L_\alpha(t)$ are not correlated for different values,*
- *$L_\alpha(t)$ is strictly $\alpha$-stable.*

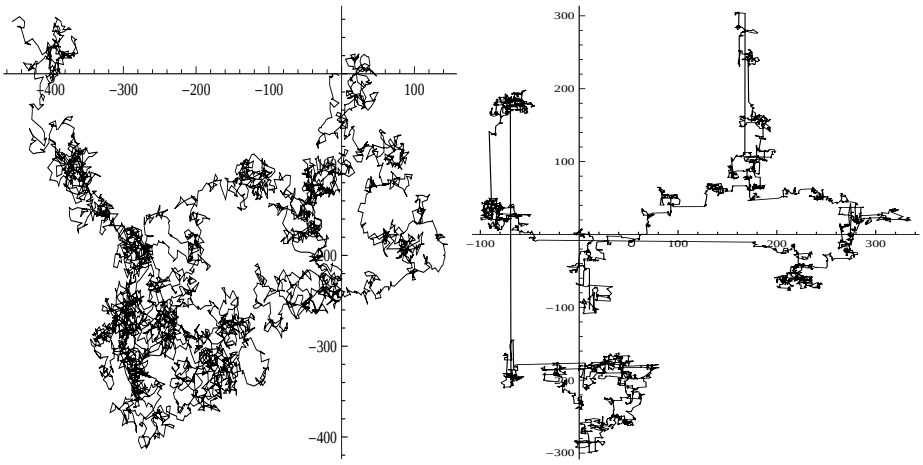



Figure 5.2: Simulation of sample path of Wiener process (left) and Lévy process (right) in 2D space. It was used 5000 steps to show the diffusion of Wiener process and so called supperdiffusion of Lévy processes which produces large jumps in comparison to Wiener process. For the simulation of Lévy distribution was used Student distribution with $\nu = 1.8$, which has Lévy distribution as its attractive distribution.

Alternatively, it is possible to define this property with help of fractional moment of the distribution, from asymptotic behavior of Lévy distribution is easy to see that for random variable $X$ with $\alpha$-stable Lévy distribution is for $l \leq \alpha$

$$\mathrm{E}(|X|^l) = \int \mathrm{d}x p(x)|x|^l < \infty. \tag{5.10}$$

So the third property can be captured in the following meaning: $L_\alpha$ is strictly $\alpha$-stable if

$$\mathrm{E}(|L_\alpha(t_1) - L_\alpha(t_2)|^\alpha) \sim |t_1 - t_2|. \tag{5.11}$$

Similarly to Wiener process, for $\alpha > 1$ can be shown that sample paths are continuous, but not differentiable [11]. We can see a difference between both distributions on figure 5.2, where is a simulation is shown in two dimensions. In every direction an independent random process was used. Here we can clearly see that processes driven by Lévy distributions produce extreme behavior, where periods of small moves are alternated by sudden large jumps.

### 5.2.2 Fractional Diffusion Equation

From the property of strictly $\alpha$-stability of Lévy flight we get the distribution function of Lévy motion as

$$p_\alpha(x,t)\mathrm{d}x = t^{\frac{1}{\alpha}} p_\alpha(x,0)\mathrm{d}x = t^{\frac{1}{\alpha}} p_\alpha(x)\mathrm{d}x \tag{5.12}$$

where $p_\alpha$ is an $\alpha$-stable Lévy distribution. That means that the Fourier transform of $p(x,t)$ for $\beta, c = 0$ is

$$\mathcal{F}[p](k,t) = \exp(-\gamma t|k|^\alpha) \tag{5.13}$$

and therefore

$$\frac{\partial \mathcal{F}[p](k,t)}{\partial t} = -\gamma|k|^\alpha \mathcal{F}[p](k,t). \tag{5.14}$$

The expression on the right side is very similar to the expression that we derived for fractional derivative, with only difference at absolute value of $k$. Actually it is a Riesz fractional derivative (defined in [6]). We can therefore rewrite this expression similarly to [11] to the form of Fractional diffusion equation

$$\frac{\partial p(x,t)}{\partial t} = \gamma \frac{\partial p(x,t)}{\partial |x|^{\alpha}} \tag{5.15}$$

for $\alpha \in (0, 2]$. This gives us another way how to describe processes with super-diffusive behavior and presence of large jumps.

### 5.2.3 Truncated Lévy Distributions

Lévy flights are undoubtedly a nice concept to describe large jumps and abnormal diffusion phenomena. Nevertheless, in real systems seems the assumption of infinite variance unnatural. Usually we are able to measure variance and we do not observe infinitely large jumps. It can be true that there are present larger jumps than can be predicted from Gaussian distribution, on the other hand very extreme jumps are usually not observed. For those systems can be a better description found, although the process will not be stable. We assume some characteristic scale $\kappa$ and for that we consider that for smaller jumps than $\kappa$ will be the system driven by Lévy distribution, for larger jumps will be driven by some other distribution, so together we get finite variance of the distribution. The easiest way is to define the distribution

$$p_{cut}(l) = \begin{cases} NL_{\alpha}(l) & \text{for } |l| \leq \kappa, \\ 0 & \text{for } |l| > \kappa. \end{cases} \tag{5.16}$$

The main disadvantage of this distribution is the fast convergence to Gaussian distribution. Another possibility that was introduced by R. Mantegna and E. Stanley [27] is to add exponentially decaying tails for large $l$'s

$$p_{trunc}(l) \sim \begin{cases} c_{+} e^{-\lambda |l|} |l|^{-(1+\alpha)} & \text{for } l < 0, \\ c_{-} e^{-\lambda l} l^{-(1+\alpha)} & \text{for } l > 0. \end{cases} \tag{5.17}$$

For this distribution can be found a characteristic function in the form taken from [5]:

$$\ln L^{t}_{\alpha\beta}(k, N) = c_0 - N\lambda^{\alpha} c \frac{(1 + (k/\lambda)^2)^{\alpha/2}}{\cos(\pi\alpha/2)} \cos\left(\alpha \arctan \frac{|k|}{\lambda}\right) \times \left(1 + i\beta \frac{|k|}{k} \tan\left(\alpha \arctan \frac{|k|}{\lambda}\right)\right) \tag{5.18}$$

where exponent $N$ denotes that this distribution holds for convolution of $N$ random variables with distribution with $N = 1$. It can be shown that for $N\lambda^{\alpha} \gg 1$ behaves this distribution as Gaussian and for $N\lambda^{\alpha} \ll 1$ as Lévy distribution. This distribution connects important properties of both Lévy and Gaussian distributions. These processes can be a successful model for simulating processes with large jumps, but finite variance.

## 5.3 Fractals and Self-similarity

In previous chapters we have seen an interesting property of many processes: they are not-differentiable and therefore extremely irregular. These properties cannot be described within classical calculus, because it deals with objects that are usually differentiable. We have seen this fact, when we defined an Itō integral. There arose terms that cannot be observed in classical Riemann integration. But these properties are for stochastic processes important and some of them depend on irregularity of sample paths. There exists a discipline that describes these irregular objects - fractal geometry. It deals with extremely irregular objects called fractals. The proper definition will be given later, but loose definition can be formulated in this manner:

> *Fractal is an object defined by simple rules and exhibiting nontrivial self-similar structure.*

This definition is not the most general, but nicely reflects main features of fractals. In order to study these objects more precisely, we introduce the main concept of fractal geometry, i.e. fractal dimension. It is a generalization of topological dimension to objects that are not smooth manifolds. However, these dimensions should have some properties that are naturally expected. Among these properties belong

- *monotonicity*, i.e. for $A \subseteq B$: $\dim(A) \leq \dim(B)$,
- *stability*, i.e. $\dim(A \cup B) = \max\{\dim(A), \dim(B)\}$,
- *invariance under regular transformations as rotation, affinity,...*
- for *countable set* $E$ is $\dim(E) = 0$,
- for *n-dimensional manifold* $M^n$ is $\dim(M^n) = n$.

All definitions will obey these properties, indeed. There exist more possible definitions of fractal dimensions, but we mention two most important ones. We begin with easier one, namely with box counting dimension. If we want to describe the dimension of the object, we cover it by balls with some diameter $\delta$. Then we cover it with balls with diameter for example $\frac{\delta}{2}$ and look, how changes the number of these balls, which are necessary to cover the object. It is clear that for a one-dimensional objects, as for example a line segment, we need 2 times more balls, for two-dimensional object we need 4 times more balls etc. The number of these balls behaves for sufficiently small $\delta$ as

$$N_\delta(M^n) \simeq c\delta^{-n}. \tag{5.19}$$

We can generalize this scaling property also for other objects and the exponent can measure the dimensionality of the object.

**Definition 5.3** *For a nonempty bounded set $F$ is $N_\delta(F)$ the smallest number of balls with diameter $\delta$ that cover $F$. Then the box-counting dimension of $F$ is*

$$\dim_B(F) = \lim_{\delta \to 0} \frac{\ln N_\delta(F)}{-\ln \delta} \tag{5.20}$$

*if the limit exists.*

The definition comes from the relation (5.19) when we make a logarithm of the relation and send *delta* to zero. The name comes from the fact that alternatively this definition can be given with number of $n$-dimensional cubes, or boxes that cover the set. The equivalence of both definition can be easily shown, because every ball of diameter $\delta$ can be covered by $2^n$ cubes with diagonal $\delta$. Since it is suitable for numerical calculations, we can also define the dimension trough $\delta$-mesh, which is a collection of cubes with side $\delta$ that create a lattice. That means that we divide the space into small cubic regions and look in how many of them the object is contained.

Another definition can be given via a special kind of measure. This measure will also deal with the property of scaling, but in more precise way.

**Definition 5.4** *For a bounded set $F$ let $\{U_i\}$ be a countable cover and if for every $U_i$*

$$|U_i| := \sup\{|x-y|; x, y \in U_i\} \leq \delta, \tag{5.21}$$

*then $U_i$ is a $\delta$-cover of $F$. We define a number*

$$\mathcal{H}^s_\delta(F) := \inf\left\{\sum_i |U_i|^s \mid U_i \text{ is } \delta-\text{cover}\right\}. \tag{5.22}$$

*Then is the Hausdorff measure defined as*

$$\mathcal{H}^s(F) := \lim_{\delta\to 0} \mathcal{H}^s_\delta(F). \tag{5.23}$$

It can be shown that $\mathcal{H}^s(F)$ is a measure, that it is nonincreasing function of $s$ and that for some $s_0$ is for $s > s_0$ equal to zero, for $s < s_0$ infinity. The point $s_0$ is very important point and defines the *Hausdorff dimension.* All of these properties are discussed in [13].

**Definition 5.5** *For a nonempty set $F$ is Hausdorff dimension*

$$\dim_H(F) := \inf\{s|\mathcal{H}^s(F) = 0\} \tag{5.24}$$

Equivalently can be used $\sup\{s|\mathcal{H}^s(F) = \infty\}$. With that definition we can define, what is fractal.

**Definition 5.6** *A fractal is an object, whose fractal dimension is greater than topological.*

As a fractal dimension any of the above defined dimensions can be taken, because usually they are both the same if they exist. Topological dimension is usually meant the dimension of objects, from which is the object compound as points, line segments, etc. If the fractal dimension is nonnatural, then the object is a fractal, as well.

An important class among fractals are (previously mentioned) self-similar fractals. We should specify this term and we show that for these objects the fractal dimension can be easily calculated. We begin with the definition of similarity as a transformation

**Definition 5.7** *Let $M$ be a set a let $S$ be a transformation on this set. $S$ is called a similarity if exist a constant $c$, such that for every $x, y \in M$*

$$|S(x) - S(y)| = c|x - y|. \tag{5.25}$$

We can say that an object is self-similar, if it is composed of several parts, that are similar to the original, as it is said in the next definition.

**Definition 5.8** *A set $F$ is self-similar, if exist a finite set of similarities $S_i$ with constants $c_i$ such that*

$$F = \bigcup_i S_i(F). \tag{5.26}$$

It is clear that $c_i < 1$. But another question is, if these constant somehow determine the fractal dimension of the fractal. The answer is yes. It is given by the next theorem, it comes from [13, theorem 9.3], where we can find also its proof.

**Proposition 5.9** *Let $F$ be a fractal with similarities $S_i$ that satisfy open set condition, i.e. for every open set $O$ is $\bigcup_i S_i(O) \subset O$, then is box-counting dimension equal to Hausdorff dimension* $\dim F = s$ *and it is solution of equation*

$$\sum_i c_i^s = 1. \tag{5.27}$$

With this we can calculate the dimension of some fractals, for example Koch curve (figure 5.3). It can be divided into four same parts that are similar to the whole and have size of $\frac{1}{3}$ in comparison to original, and therefore from

$$4\left(\frac{1}{3}\right)^s = 1 \tag{5.28}$$

we get that $\dim F = s = \frac{\ln 4}{\ln 3} \doteq 1.26$. We also note that these self-similar fractals are in practical computations created by iterative process, described by similarities.

We mention also one proposition that helps us in the next.

**Proposition 5.10** *Let $f$ be a function $F \to \mathbf{R}^n$, such that for all $x, y \in F$*

$$|f(x) - f(y)| \leq c|x - y|^\alpha, \tag{5.29}$$

*then*

$$\mathcal{H}^{s/\alpha}(F) \leq c^{s/\alpha}\mathcal{H}^s(f(F)). \tag{5.30}$$

We get this relation from definition, because if $\{U_i\}$ is $\delta$-cover of F, then $\{f(U_i \cap F)\}$ is $c\delta^\alpha$-cover of $F$. From the relation we have that $|f(U_i \cap F)| \leq |U_i|^\alpha$, so after limiting $\delta \to 0$ we get the relation above.

As a corollary of the last proposition we get similar relation for dimensions:

**Corollary 5.11** *For a function satisfying equation* (5.29) *holds*

$$\dim_H f(F) \leq \frac{1}{\alpha}\dim_H(F). \tag{5.31}$$

Because we would also like to describe some fractal random processes of random objects and therefore we also define a statistical self-similarity.

**Definition 5.12** *A object $F$ is statistically self-similar, if it is self-similar in distribution.*

We can see that if we slightly change the Koch curve, so the middle part is broken either up or down, then it significantly changes its shape, but a fractal dimension remains, as observed in figure 5.3.

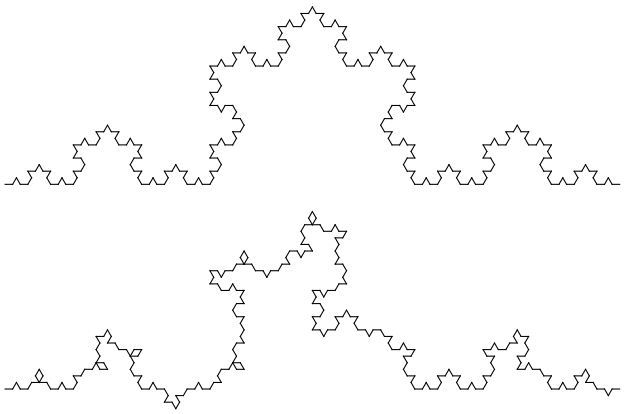


Figure 5.3: The fourth iteration of Koch curve and its random version. Both fractals have fractal dimension $s = \frac{\ln 4}{\ln 3}$, the first one is left-similar while the latter one is statistically self-similar.

## 5.4 Fractal Properties of Stochastic Processes

Now we would like to apply the knowledge from the last chapter to random processes. Here we have to make a little distinction between two possible representations of random processes.

**Definition 5.13** *A sample path of the n-dimensional process process $X(t)$ is a set of points $x \in \mathbf{R}^n$ that arose from one realization of random process.*

**Definition 5.14** *A sample function of the random process is mapping to one realization of random process $t \mapsto X(t)$.*

The distinction is natural, because a sample function cannot "go back in time". This has an influence on roughness of the sample path, and consequently on its fractal dimension. Another problem can be in measuring the dimension for the graph of sample function, because it is not defined in two-dimensional space with norm, but there are independent values of $t$ and values of the process, so we cannot construct open sets. This problem can be solved by introducing a relation between these two coordinates, called affinity. Roughly speaking we choose the ratio between these two variables when we want to draw a graph of the function (which we do always implicitly when drawing a graph) and then we can look at the graph as a set of points in $\mathbf{R}^2$.

The important property that determines fractal dimension is of course scaling of the system. We know that for Wiener process holds that the increments are given by normal distribution

$$P[W(t+T) - W(t) \in [x, x + \mathrm{d}x]] = \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{x^2}{2T}\right) \mathrm{d}x, \tag{5.32}$$

so if we rescale variables to $T \mapsto \alpha T$ and $x \mapsto \sqrt{\alpha} x$ then we get the same distribution. We can illustrate this property very easily, if we notice that the space increment is similar to the square root of time, i.e.

$$|\Delta x| \simeq \sqrt{\Delta t}. \tag{5.33}$$

This argument can be used to estimate the upper bound of the dimension of fractal dimension of Wiener process, because when we apply 5.11 to $|W(1) - W(0)|$ we get that the dimension is at least two. This argument can be formulated correctly, which is partially shown in [22] and then in [13]. Similarly, we can calculate a fractal dimension for Lévy flights and fractal Brownian motion. So together we have:

- a sample path of Wiener process in $\mathbb{R}^n$ $(n \geq 2)$ is the dimension 2,
- a graph of Wiener sample function $t \mapsto W(t)$ has dimension $\frac{3}{2}$,
- a sample path of Lévy process $L_\alpha(t)$ has dimension $\max\{1, \alpha\}$,
- a graph of Lévy sample function has dimension $\max\{1, 2 - \frac{1}{\alpha}\}$,
- a graph of fractional Brownian motion $t \mapsto W_H(t)$ has dimension $2 - H$.

We see that processes that are at first sight different, have some similar properties in the meaning of roughness, which coheres with the presence of antipersistent behavior and large jumps. We can use this knowledge to estimate, how the process scales with time and from that we can reveal if the series contains memory effects and/or large jumps.

## 5.5 Hurst Exponent

In the last section we saw that an important quantity of the series is an exponent $H$, which is given as

$$|\Delta x| \propto |\Delta t|^H. \tag{5.34}$$

We saw this exponent a few times, firstly by fractional Brownian motion, where for $H = \frac{1}{2}$ we got Wiener process, secondly by Lévy flight, where we could not use variance, so we expressed it through other moments. Therefore the exponent $H \in (0, 1)$, given by the relation

$$H \simeq \frac{\ln |\Delta x|}{\ln |\Delta t|}, \tag{5.35}$$

is an important quantity that helps us recognize these nontrivial effects.

### 5.5.1 Rescaled Range Analysis

In order to estimate the exponent for the real time series, we introduce a statistical method called Rescaled range analysis (R/S analysis) that is based on ideas from the last section. The estimation of the series is following: suppose that we have a wide-sense stationary series $x = \{x_k\}$ for $k = \{1, \dots, n, \dots\}$. At first we look at a part of series of length $n$. We define such an integrated-increments series $w$, that

$$w_k = \sum_{i=1}^{k} (x_i - \bar{x}_n) \tag{5.36}$$

where $\bar{x}_n$ is a mean of the partial series

$$\bar{x}_n = \frac{1}{n} \sum_{j=1}^{n} x_j. \tag{5.37}$$

Then we introduce a rescaled range function $R/S(n)$, so

$$R/S(n) := \frac{R(n)}{S(n)} = \frac{\max(w_1, \ldots, w_n) - \min(w_1, \ldots, w_n)}{\left(\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x}_n)^2\right)^{\frac{1}{2}}} \tag{5.38}$$

so the series $R(n)$ is a range and $S(n)$ is a standard deviation. The Hurst exponent can be estimated from a limit expression

$$\lim_{\tau\to\infty} R/S(\tau) = C\tau^H. \tag{5.39}$$

For finite series we divide the length of the series by factor of two and try to get the exponent from maximal likelihood estimation. The mechanism is described in [33]. Important thing is that we have one mechanism, how to reveal memory effects and non-gaussian behavior of time series.

## 5.6 Detrended Fluctuation Analysis

An alternative to R/S analysis is a detrended fluctuation analysis (DFA). As the name prompts, we get rid of trends and investigate only fluctuations. It was firstly described in [31]. The method helps to find the scaling properties of the process. We assume, that the process is self-similar and try to find relations between scaling of axes, which corresponds to Hurst exponent. The procedure is following: We divide the series of length $N$ into $l$ non-overlapping intervals. In every interval we find the linear trend by least-square method and then average all fluctuations from trend. When we formulate it mathematically, similarly to R/S analysis we have to do these steps: Suppose an increment process. Similarly to the last section, we introduce a integral process with

$$y_k = \sum_{i=1}^{k}(x_i - \bar{x}_n). \tag{5.40}$$

We define for $l_j \in \left(\frac{N(j-1)}{l} + 1, \frac{Nj}{l}\right)$ a series $y_k^{l_j}$ which arose as series given by linear fit of series $y_k$ on this interval. We define a fluctuation from the trend in one interval as

$$f_{l_j}^2 = \frac{l}{N}\sum_{l_j} |y_k - y_k^{l_j}|^2 \tag{5.41}$$

and the fluctuation function for the whole series as a mean of these fluctuations, therefore

$$F(l) = \left(\frac{1}{l}\sum_{j=1}^{l} f_{l_j}^2\right)^{\frac{1}{2}}. \tag{5.42}$$

The properties of fluctuation function have been widely studied, the important thing for us is the fact, that for non-correlated random walks the function should scale as $l^{\frac{1}{2}}$ and for larger exponents, we get power-law properties with scaling invariance, as discussed before.

# Chapter 6

# Multifractal Stochastic Processes

This chapter is the crucial part of the thesis. We would like to answer a question, how a various behavior of assets traded on different markets can arise and what kind of behavior, that seems to be extreme, can be included into some suitable model. There have been studied many different approaches, as double-stochastic models [18], superstatistics [2], theory of entropy [19, 36] and many others. For this thesis a multifractal approach has been chosen for several reasons. Firstly, multifractal formalism is a natural generalization of fractals, which gives a multi-scaling, that means, we get processes with scaling properties changing in time. Secondly, as we will see in the section about trading time and clock time, the multifractal model uses time deformation, which arises from the difference between some "inner" time on the market and the outer clock time. At the first sight this seems quite unnatural, but it can elegantly explain many phenomena on financial markets. We begin with the formalism and then we move to applications to time series.

## 6.1 Multifractal Formalism

In this section we would like to generalize fractal formalism to objects that have locally different scaling properties. The reason is that the demand of self-similarity seems to be too strict for objects observed in real experiments, but on the other hand it is locally possible to observe these scaling properties. Multifractals have a wide range of applications, for example in measuring length of coasts, chaotic dynamics, and others [17]. The definition given in this section is not only possible definition of a multifractal, but seems to be the most suitable for our needs.

We begin with the measure $\mu$ with support $F$, where $F$ is a bounded subset of $\mathbf{R}^n$. The subset can be fractal, or does not have to be. Let $\{U_i\}$ be the smallest $\delta$-cover of $F$ (similarly to box-counting dimension) and we define a number $N_\delta(\alpha)$ which equals to

$$N_\delta(\alpha) = \#\{U_i | \mu(U_i) \geq \delta^\alpha\} \tag{6.1}$$

for $\alpha \in \mathbf{R}$, so that are balls with measure larger than $\delta^\alpha$. We also introduce a sum over all balls

$$S_\delta(q) = \sum \mu(U_i)^q. \tag{6.2}$$

Because $S_\delta(0) = N_\delta(\mathrm{supp}(\mu)) = N_\delta(F)$, then the expression $-\frac{\ln S_\delta(q)}{\ln \delta}$ is a generalization of the fractal dimension, but we have to be care about existence of this limit for every $q$. We show the existence of the limit and also the relation between $N_\delta(\alpha)$ and $S_\delta(q)$. We now consider only balls for those is

$$\delta^{\alpha+\epsilon} \leq \mu(U_i) < \delta^\alpha \tag{6.3}$$

for $\delta < 1$ and small $\epsilon$. Let us assume that for those balls exist a function $f(\alpha)$ such that

$$\delta^{-f(\alpha)} \sim \#\{U_i | \delta^{\alpha+\epsilon} \leq \mu(U_i) < \delta^\alpha\}. \tag{6.4}$$

The connection between the sum $S_\delta$ and $N_\delta$ can be approximately formulated as

$$S_\delta(q) \sim \int_0^\infty \mathrm{d}\alpha \delta^{-f(\alpha)} (\delta^\alpha)^q = \int_0^\infty \mathrm{d}\alpha \delta^{\alpha q - f(\alpha)}. \tag{6.5}$$

In the integral we added measures of all ball that scale as $\delta^\alpha$ and multiplied it with the number of these balls. On the other hand the integrand can be estimated in other way, concretely as $N_\delta(\alpha+\epsilon) - N_\delta(\alpha-\epsilon)$ so we get

$$\delta^{f(\alpha)-\eta} \lesssim N_\delta(\alpha+\epsilon) - N_\delta(\alpha-\epsilon) \lesssim \delta^{f(\alpha)+\eta} \tag{6.6}$$

which gives us the relation for $f(\alpha)$. We formulate it as a theorem and omit the proof, it can be found in [13].

**Proposition 6.1** *For every $\alpha > 0$ the limit*

$$\lim_{\delta\to 0} \lim_{\epsilon\to 0} \frac{\log(N_\delta(\alpha+\epsilon) - N_\delta(\alpha-\epsilon))}{-\log \delta} =: f(\alpha) \tag{6.7}$$

*always exists and is called multifractal spectrum.*

Now we look closer to the relation to $S_\delta(q)$. The main contribution to the integral (6.5) for small $\delta$ gives $\delta^{q\alpha - f(\alpha)}$ when the exponent is minimal. We define a *scaling function* $\tau(q)$ as

$$\tau(q) = \inf_\alpha (q\alpha - f(\alpha)) \tag{6.8}$$

and we consider that the integral has a sharp maximum around $\tau(q)$, so $S_\delta(q) \sim \delta^{\tau(q)}$. Again, we cite a result from [13] and give a precise expression.

**Proposition 6.2** *For a $\tau(q)$ defined in* (6.8) *holds the following expression:*

$$\tau(q) = \lim_{\delta\to 0} \frac{\ln S_\delta(q)}{\ln \delta} \tag{6.9}$$

*and the limit on the right side always exist.*

Even more interesting situation is, when we assume differentiability and convexity of $f(\alpha)$. We denote $\alpha(q)$ as a value, for which is the expression $q\alpha - f(\alpha)$ for given $q$ minimal. Therefore, we get

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}(q\alpha - f(\alpha))|_{\alpha(q)} = 0, \tag{6.10}$$

so

$$f'(\alpha(q)) = q. \tag{6.11}$$

If we put $\alpha(q)$ back into the definition of $\tau(q)$ and calculate a derivative of $\tau(q)$ we get an interesting relation between $\alpha$ and $\tau$:

$$\frac{\mathrm{d}}{\mathrm{d}q}\tau(q) = \alpha(q) + q\alpha'(q) - f'(\alpha(q))\alpha'(q) = \alpha(q). \tag{6.12}$$

This is nothing else than a Legendre transformation from variables $(f(\alpha), \alpha)$ to $(\tau(q), q)$. The transformation is given by equations

$$\tau(q) = q(\alpha(q)) - f(\alpha(q)), \tag{6.13}$$

$$\frac{\mathrm{d}\tau(q)}{\mathrm{d}q} = \alpha(q), \tag{6.14}$$

$$q = \frac{\mathrm{d}f(\alpha(q))}{\mathrm{d}q}. \tag{6.15}$$

The inverse Legendre transformation has the same form, but we formulate these terms as functions of $\alpha$. It seems to be advantageous to use both the descriptions and we will be doing it from time to time in further calculations. We also note that we get special values for some appropriate choices of $q$. For example $\tau(0) = -\dim_B(F)$. This can be seen directly from definition. For other choices we can get other fractal dimensions, that were not defined in this work, but they are used as well (for example in [17]).

## 6.2 Trading Time and Clock Time

For modeling financial market simply models based on Brownian motion are mostly used. We already saw and will also see on real data that these models work in some times quite well, but in some situations they fail, or do not give the right answer. The reason is that these models work with very idealized conditions and oversimplified presumptions. One important restriction is the assumption of constant volatility. In classical models based on Brownian motion we assume a geometric Brownian motion, so for logarithm of the price we assume

$$\ln S(t) - \ln S(0) = W(t). \tag{6.16}$$

The problem is that the volatility is not constant over time, but varies quite noticeably. One way to grasp it is to look at the markets and see, if there is the same volume of traded assets, if the information stream is constant in time etc. We see that during time these quantities are not constant and for example it happens that the most rush times during the day are the first minutes after opening the market. That means that in some time interval more things happen and we can compare this interval to some other less rush interval during the day,

that brings the same amount of information, but it is longer. We can model this unbalanced behavior with the deformation of the time. We assume that there exists one more "virtual" time on the stock market called *trading time*, for which the price can be modeled with Brownian motion, but it differs from our real time and this deformation causes, that if the time interval at trading time is longer than in clock time, we get probably larger jump than we expected and vice versa. The price can be therefore modeled as a compound process

$$\ln S(t) - \ln S(0) = B[\theta(t)] \tag{6.17}$$

where $B$ is usually Brownian motion (but can be easily generalized to fBM, Lévy flight, etc.) and $\theta(t)$ is the time deformation - a increasing function of the clock time. We consider that $B$ and $\theta$ are independent.

## 6.3 Modeling of Time Deformations

We will model these time deformations from knowledge of scaling properties through fractal patterns. This approach affects all scales and we will therefore observe time deformations in all orders of magnitude, which is a natural expectation. This approach will be formalized into the multifractal formalism in the next part, but it is really helpful to see the motivation, why we will use these measures. The approach of creating time deformations from fractal patterns is widely discussed for example in [23].

As an example we create one deformation given by difference between so called Wiener pattern and Multifractal pattern. Wiener pattern is a fractal that represents a time flow in the trading time, where the time flows equally and we model processes by Brownian motion. This pattern is the one that scales as Brownian motion, i.e. $\Delta x \simeq \Delta t^{\frac{1}{2}}$. We create it, as many self-similar objects, by iteration. We begin with a straight line from $[0,0]$ to $[1,1]$ and we break it into a few parts, all scaling as Brownian motion. We use the simplest way, i.e. we divide it into three parts in the symmetric way, so the first line leads from $[0,0]$ to $[4/9, 2/3]$ then to $[5/9, 1/3]$ and then to $[1,1]$, so every line scales as Brownian motion. We repeat this iteration (theoretically up to infinity) and the resulting object is Wiener pattern that represents classical random walk behavior. The iterating process is drawn in figure 6.1.

In order to create a time deformation, we introduce the second pattern that will represent the non-gaussian behavior. As shown in figure 6.2, we introduce a few generators, by shifting original points, and therefore disrupt Wiener scaling. It is possible to move the first, the second point or both, or when we use more complicated pattern, we can produce a complex multiscaling behavior. We can observe the time deformation, which is generated as a difference between appropriate points of patterns and is therefore a multifractal itself (figure 6.3. Then we generate a multifractal process as a compound process and we get regions, where large deviations can be observed (figure 6.5). Generally, a multifractal process is a process with fluctuating volatility, and the great benefit is that we can observe these fluctuations in all scales.

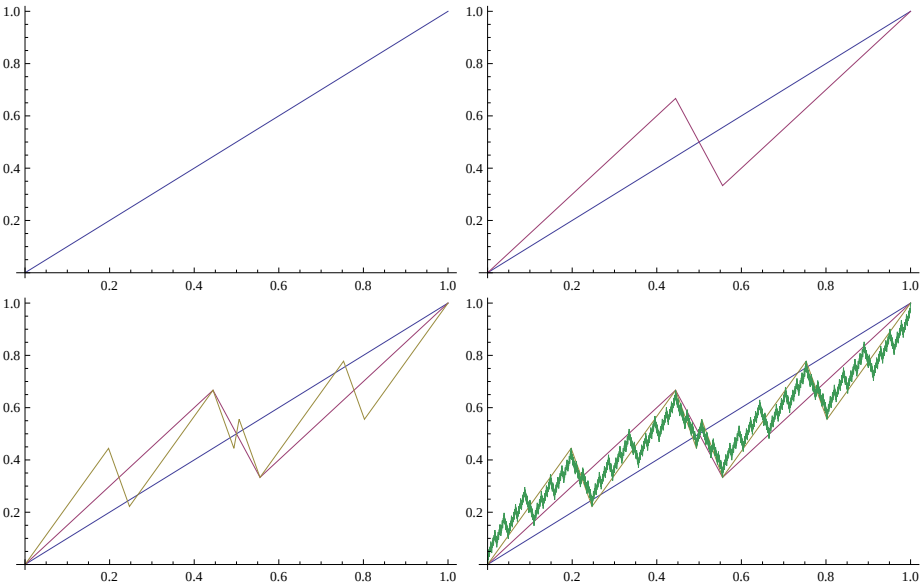



Figure 6.1: First few iterations of Wiener pattern. We begin with an initiator, a straight line and brake into three parts for that $\Delta x = (\Delta t)^{\frac{1}{2}}$.

## 6.4 Multifractal Measures

In this section we do a mathematical formalization of multifractal measures used in the next chapter to simulate price evolution on financial markets (but can be also applied to many other problems.) We showed in the last section the idea of time deformations and the concept of multifractal measures will serve us as a tool to generate these transformations. Further, we will generate the time deformation function $\theta(t)$ as a c.d.f. of some given multifractal measure $\mu(t)$, so

$$\theta(t) = \mu[0, t]. \tag{6.18}$$

We begin with some simple examples of multifractal measures and generalize them into measures that are used in economical models.

### 6.4.1 Binomial Measure

Apart from Koch curve (or Koch snowflake), other types of fractal are known, we mention a Cantor dust, which will be an important example for this section. The dust is generated from the line, by dividing the line into three thirds and leaving out the middle third of the line. In the next iteration we do the same with remaining parts of the line. It is easy to see that the fractal has the dimension $\frac{\ln 2}{\ln 3}$. We will use the same iterative approach for generating binomial measure, but we will not leave out some interval, we will redistribute "weights" of intervals. Without loss of generality we show the iteration for a measure on the interval $[0, 1]$. We simply begin with an uniform measure, for which $\mu_0[0, 1] = 1$. Then we bisect the interval into two subintervals and define the measure, which is uniform on every subinterval and

$$\mu_1[0, 1/2] = m_0, \qquad \mu_1[1/2, 1] = m_1. \tag{6.19}$$

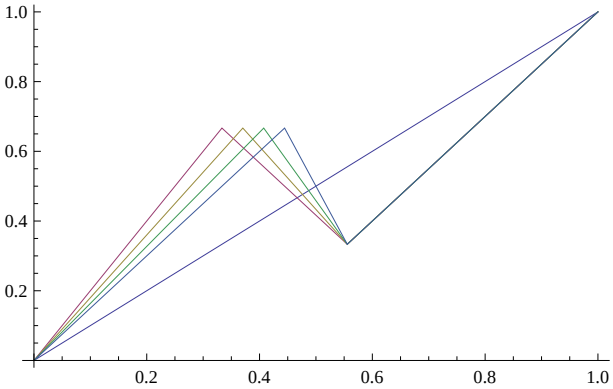



Figure 6.2: Four possible generators of a multifractal pattern. The original point $\{\frac{4}{9}, \frac{2}{3}\}$ from Wiener pattern is replaced by one of the points $\{\frac{1}{3}, \frac{2}{3}\}$, $\{\frac{10}{27}, \frac{2}{3}\}$ or $\{\frac{11}{27}, \frac{2}{3}\}$, by every iteration is the point chosen randomly. The pattern represents the nonlinear time transformation.

where holds $m_0 + m_1 = 1$, so the measure of the whole interval $[0, 1]$ remains unchanged. In the second step we use the procedure to every part, so we get

$$\begin{aligned} \mu_2[0, 1/4] &= m_0\mu_1[0, 1/2] = m_0^2, \\ \mu_2[1/4, 1/2] &= m_1\mu_1[0, 1/2] = m_1 m_0, \\ \mu_2[1/2, 3/4] &= m_0\mu_1[1/2, 1] = m_0 m_1, \\ \mu_2[1/2, 1] &= m_1\mu_1[1/2, 1] = m_1^2. \end{aligned}$$

This iterative process continues up to infinity, so we get the binomial multifractal measure $\mu$ as the limit $\mu_k \to \mu$ for $k \to \infty$.

We should note that there is one nice definition of this measure. Let us assume an arbitrary interval of $k$-th iteration $\Delta t = [t, t + 2^{-k}]$ where $t = n2^{-k}$. In this interval are numbers that in binary system begin with the same $k$ first digits, so for every $a \in \Delta t$ we have $a = 0.\beta_1\beta_2 \ldots \beta_k \ldots$ where $\beta_i = \{0, 1\}$ are digits characterizing the interval $\Delta t$. We define $\phi_0, \phi_1$ as frequencies of zeros and ones in the first $k$ digits of the interval. We can then write the $k$-th approximation of the measure as

$$\mu_k(\Delta t) = m_0^{k\phi_0} m_1^{k\phi_1}. \tag{6.20}$$

One possible generalization of this binomial measure is easy: by every iteration we toss a coin and decide if measure $m_0$ should go to the left or to the right. $m_1$ then goes to the other one.

### 6.4.2 Multiplicative Measure

Both two previous examples can be generalized into a bigger class of multifractal measures - multiplicative measures. We assume the same iterative process as in the last section, but the weight division will be driven by identical copies of a random variable $M$. We divide the original interval into $n$ subintervals, and to every interval we assign the measure $M_\gamma$, where $M_\gamma$ is one realization of $M$. For the last two examples we had that $m_0 + m_1 = 1$. Generally, we can write

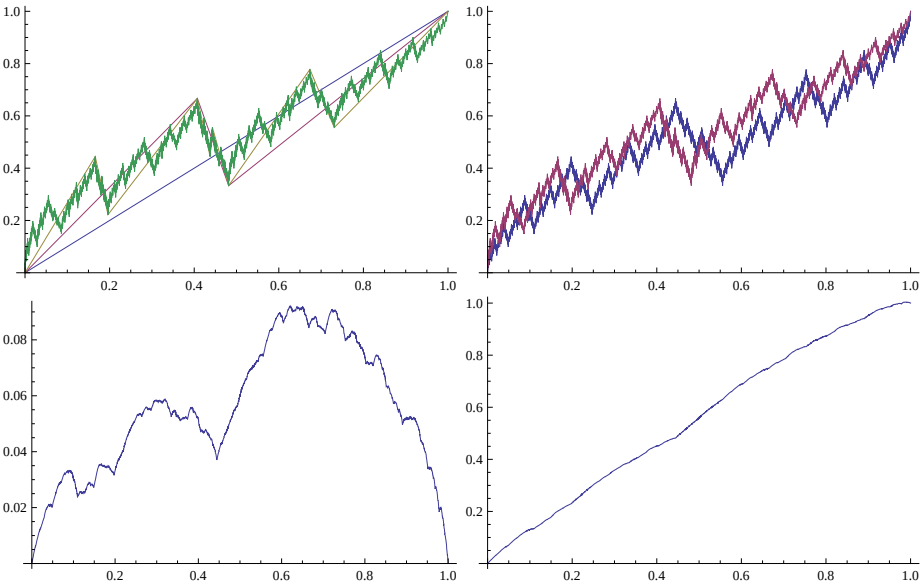



Figure 6.3: On top figures we can see the comparison of Wiener and multifractal pattern. On bottom figures we can see difference of both times and dependence of trading time on clock time.

that $\sum M_{\gamma=1}^{\beta} = 1$ These measures are called conservative, or microcanonical (similarly to the microcanonical ensemble statistical physics). The disadvantage is that these random variables cannot be independent. But on the other hand we can weaken the condition of conservation and only request the conservation on average, so $\sum_{\gamma=1}^{\beta} \mathrm{E}(M_\gamma) = 1$ which means $\mathrm{E}(M) = \frac{1}{\beta}$. The advantage is that we can use independent variables $M_\gamma$.

We can calculate the scaling function $\tau(q)$. The expression $S_{2^{-k}}(q)$ is equal to

$$S_{\beta^{-k}}(q) = \mathrm{E}\left(\sum_{i=1}^{\beta^k} \mu_k(t_i)^q\right) = \beta^k \mathrm{E}([M]^q)^k, \tag{6.21}$$

therefore is

$$\tau(q) = \lim_{k\to\infty} \frac{\log_\beta(\beta^k \mathrm{E}([M]^q)^k)}{\log_\beta \beta^{-k}} = -\log_\beta \mathrm{E}([M]^q) - 1. \tag{6.22}$$

Finally, we get a scaling relation between the measure and $\Delta t$. Firstly for conservative measure:

$$\mathrm{E}(\mu(\Delta t)^q) = (\Delta t)^{\tau(q)+1}. \tag{6.23}$$

We get the similar expression for canonical measures. Now is $\mu[0,1]$ a random variable, we denote it $\Omega$. With this notation we get

$$\mathrm{E}(\mu(\Delta t)^q) = \mathrm{E}(\Omega^q)(\Delta t)^{\tau(q)+1}. \tag{6.24}$$

For a finite interval $\Delta t$ we can show these relations only on intervals $\Delta t = [t, t + 2^{-k}]$ where $t = n2^k$. Alternatively, we can define the measure with this

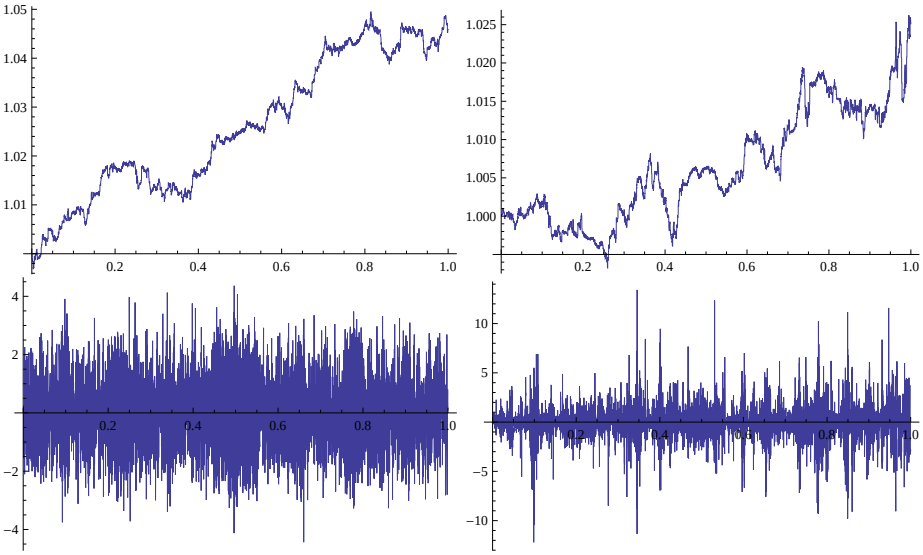



Figure 6.4: Processes that produce periods with large jumps can be created with multifractal patterns. On the left side there is an example of Geometric Brownian motion and its logarithmic returns, on the right side there is a multifractal process with its returns.

condition for infinitesimally small interval. The measure is then given by an equation

$$\mathrm{E}(\mu(\mathrm{d}t)^q) = c(q)(\mathrm{d}t)^{\tau(q)+1} \tag{6.25}$$

when $\mathrm{d}t \to 0$.

In both cases is possible to show that $\tau(q)$ is concave function. We assume that $q = \alpha_1 q_1 + \alpha_2 q_2$. From Hölder inequality we get

$$E(\mu(\Delta t)^q) \leq [E(\mu(\Delta t)^{q_1})]^{\alpha_1} \, [E(\mu(\Delta t)^{q_2})]^{\alpha_2} \tag{6.26}$$

and because the mean scales as $(\Delta t)^{\tau(q)+1}$ and $\Delta t$ is small, we get that

$$\tau(q) \geq \alpha_1 \tau(q_1) + \alpha_2 \tau(q_2). \tag{6.27}$$

We can therefore calculate a multifractal spectrum of the measure as a Legendre transform. The spectrum can be calculated from

$$f(\alpha) = \inf_q [q\alpha - \log_\beta \mathrm{E}([M]^q) - 1]. \tag{6.28}$$

As an example we show the form of multifractal spectrum for binomial measure and dividing into two intervals, as defined before. The scaling function has an easy form

$$\tau_b(q) = -\frac{\ln(m_0^q + m_1^q)}{\ln 2}. \tag{6.29}$$

From that is $\alpha(q)$ equal to

$$\alpha(q) = \frac{\mathrm{d}\tau(q)}{\mathrm{d}q} = -\frac{m_0^q \ln m_0 + m_1^q \ln m_1}{(m_0^q + m_1^q)\ln 2} \tag{6.30}$$

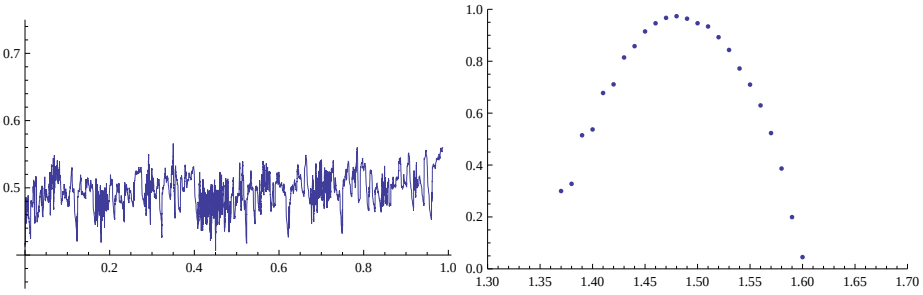



Figure 6.5: Left: Local Hurst exponent of the multifractal pattern. The Hurst exponent is calculated from the ratio $\frac{\ln(\Delta|x|)}{\ln(\Delta t)}$ for every part of the fractal. We can see that it rapidly fluctuates around 0.5. Right: Multifractal spectrum $f(\alpha)$ for the mulfractal pattern.

So the multifractal spectrum is implicitely given as

$$f_b(\alpha(q)) = q\alpha(q) - \tau(q) = \frac{\ln(m_0^q + m_1^q)}{\ln 2} - \frac{q(m_0^q \ln m_0 + m_1^q \ln m_1)}{(m_0^q + m_1^q)\ln 2}. \quad (6.31)$$

It is not difficult to express $q$ as a function of alpha

$$\begin{aligned} \alpha = \frac{\mathrm{d}\tau(q)}{\mathrm{d}q} &= -\frac{m_0^q \ln m_0 + m_1^q \ln m_1}{(m_0^q + m_1^q)\ln 2} \\ -\alpha \ln 2 &= \frac{\ln m_0 + \left(\frac{m_1}{m_0}\right)^q \ln m_1}{1 + \left(\frac{m_1}{m_0}\right)^q} \\ (-\alpha \ln 2 - \ln m_1)\left(\frac{m_1}{m_0}\right)^q &= (\alpha \ln 2 + \ln m_0) \end{aligned}$$

so finally we get

$$q(\alpha) = \frac{\ln\left(-\frac{\ln m_0 + \alpha \ln 2}{\ln m_1 + \alpha \ln 2}\right)}{\ln m_1 - \ln m_0}. \quad (6.32)$$

It is also possible to calculate the spectrum in other cases, but long and heavy calculations are needed. Nevertheless, the knowledge of multifractal spectrum helps to characterize properties of multifractal processes.

We should also mention that it has been shown [7, 25] that if we consider a random measure, for that $E(M^q) < \infty$ for every $q > 0$ and if random multipliers are greater than one with positive probability, then a critical exponent $q_{crit}$ exists and the random variable $\Omega$ has polynomial right tail, which means

$$P[\Omega > \omega] \sim C\omega^{-q_{crit}} \quad (6.33)$$

where $1 < q_{crit} < \infty$, i.e. we can observe a heavy-tail behavior of $\Omega$.

# Chapter 7

# Models of Price Evolution

The last theoretical chapter is aimed at financial models of asset prices. We begin with classical models, frequently used by both the theoretics and practitioners, as ARMA or GARCH, and then introduce models that are based on multifractal approaches. The main advantage of these approaches compared to other models, is a natural way how to create series with multi-frequency volatility, where cycles of many frequencies can be observed. We will see that these models create series with complex behavior naturally, which is favorable for us, with regard to the practical applications.

## 7.1 Classical Models

We firstly provide a brief overview of the most used models for modeling time series. These models will be a good basis for us, because we will generalize these processes in the second part of this chapter, or at least we will highlight some concrete properties of these processes.

### 7.1.1 ARMA

Autoregressive moving average model is an illustrative and popular model applied to autocorrelated data. The model consists of two parts. Let $r_t$ be a time series in time $t$ and $\eta_t$ a series of independent random variables with normal distribution $\mathcal{N}(0, \sigma^2)$ (white noise) that will represent new information supplied into the market. There exist also processes with $\eta_t$ that has other distributions, as student distribution or others. We will stay only by processes that are normally distributed.

The first part of the model is an autoregressive process

$$r_t = c + \sum_{i=1}^{p} \phi_i r_{t-i} + \eta_t. \tag{7.1}$$

Constants $\phi_i$ are parameters of the model. This model considers that the price return is correlated with previous values. It is necessary to say that if the series has to be stationary, then all roots of the polynomial

$$x^n - \sum_{i=1}^{n} \phi_i x^{n-i} \tag{7.2}$$

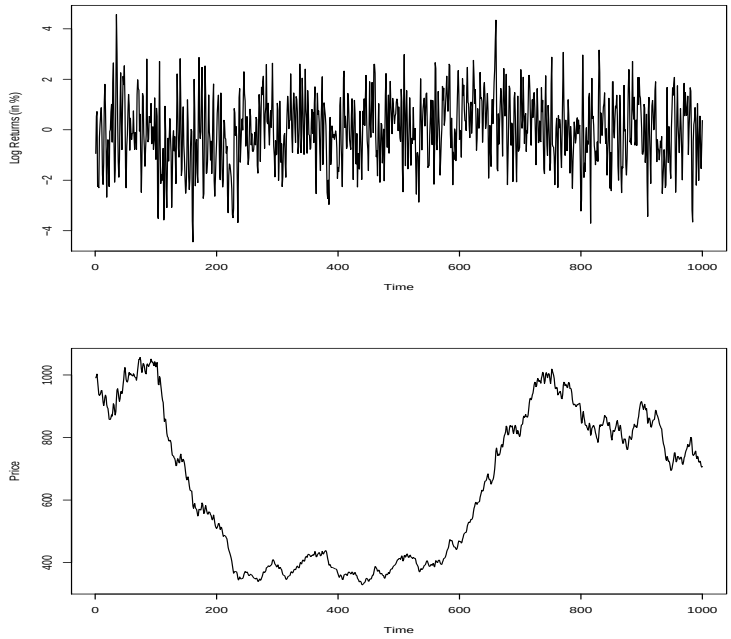



Figure 7.1: Simulation of ARMA(2,1) returns and corresponding prices.

have to lie in the unit circle [4]. For example, for $AR(1)$ is the condition $|\phi_1| < 1$. If we look closer to the process $AR(1)$, and we neglect the constant $c$, we get that

$$r_t = \phi_1 r_{t-1} = \phi_1^2 r_{t-2} + \phi_1 \eta_{t-1} + \eta_t = \cdots = \sum_{i=1}^{n} \phi_1^i \eta_{i-1}. \tag{7.3}$$

The autocorrelation function is therefore

$$E(r_t r_{t-i}) = \phi_1^i, \tag{7.4}$$

so the autoregressive model assumes decaying correlations.

The second part of the model is moving average process

$$r_t = c + \eta_t + \sum_{j=1}^{q} \theta_i \eta_{t-i}. \tag{7.5}$$

Together, the model ARMA(p,q) can be formulated as

$$r_t = c + \eta_t + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{j=1}^{q} \theta_i \eta_{t-i} \tag{7.6}$$

The interpretation of the model is the following: the actual price is influenced by actual information represented by $\eta_t$, by past prices $r_{t-i}$ and past information $\eta_{t-i}$. The model has $p + q + 2$ unknown parameters, $c$, $\sigma^2$, $\phi_i$ and $\theta_i$. Often is enough to use models with small $p$ and $q$.

### 7.1.2 ARCH, GARCH

Other models that are often used are models based on volatility autoregressive models. That means that returns are specified as

$$r_t = \sigma_t \eta_t, \tag{7.7}$$

where $\eta_t$ is again the white noise unit variance and volatility $\sigma_t^2$ obeys an autoregressive process defined in the last section. If the process fulfills ARMA(p,q) process,

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \eta_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{7.8}$$

we talk about GARCH(p,q) (generalized autoregressive conditional heteroscedasticity) process. In case, when $p = 0$, we talk about ARCH(q) process. There also exist large number of different extensions of GARCH models. We shall look a bit closer to some important properties of the most popular models ARCH(1), respectively GARCH(1,1).

The ARCH(1) is specified by equation

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 \tag{7.9}$$

or equivalently

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 \eta_{t-1}^2. \tag{7.10}$$

If we are interested in the average value of volatility and assume that $\mathrm{E}(\sigma_t^2) \approx \mathrm{E}(\sigma_{t-1}^2) = \sigma^2$ we get that the average volatility is

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1}. \tag{7.11}$$

If we look at GARCH(1,1)

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{7.12}$$

the similar formula for volatility could be derived, we get that

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}. \tag{7.13}$$

By iterating of the process back we get

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1(\alpha_0 + \alpha_1 r_{t-2}^2 + \sigma_{t-2}^2) = \dots \\ &= \frac{\alpha_0}{1 - \beta_1} + \alpha_1 \sum_{i=1}^{\infty} \beta_1^{i-1} \sigma_{t-i}^2. \end{aligned} \tag{7.14}$$

That shows us an exponential decay in the volatility which means short-range correlations. This argument can be shown directly, it can be shown that autocovariance function $\mathrm{Cov}(x_t^2, x_{t+n}^2)$ fulfills relation [27]

$$\mathrm{Cov}(x_t^2, x_{t+n+1}^2) = (\alpha_1 + \beta_1)\mathrm{Cov}(x_t^2, x_{t+n}^2). \tag{7.15}$$

This enables us to consider the covariance function in the form $Ae^{-\frac{t}{\tau_c}}$ and it confirms our hypothesis and for modeling long-range correlated processes we have to use other processes.

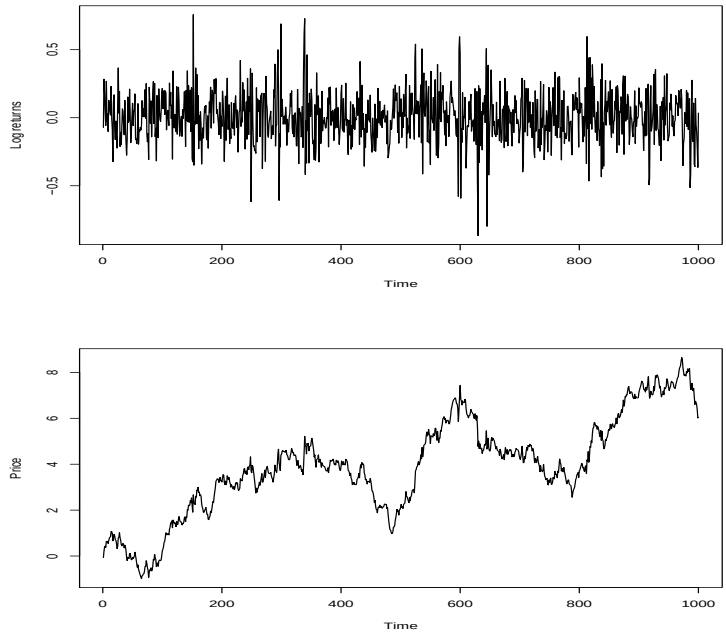



Figure 7.2: Simulation of GARCH(1,1) returns and corresponding prices.

### 7.1.3 Regime Switching Models

Regime-switching models are a completely different approach. These models are based on some unobservable latent variables that can suddenly change and influence observable quantities. The important model is Markov-Switching model for which we have an equation

$$r_t = \mu(M_t) + \sigma(M_t)\eta(t) \tag{7.16}$$

where $M_t$ is Markov chain, so it is fully characterized by transition matrix $\mathbb{P}_{ij} = (M_{t+1} = m_i | M_t = m_j)$. This model brings us the possibility to simulate sudden changes and economical cycles.

A good combination of both methods seems to be the MS-GARCH process. We consider a state Markovian variable $s_t$ which can have two values $s_1, s_2$. The returns are described by relation $r_t = \sigma_t \eta_t$ and conditional volatility $\sigma_t^2(i) = \mathrm{Var}(r_t | s_t = i)$ is given by an GARCH(1,1) equation

$$\sigma_t^2(i) = \alpha_0(i) + \alpha_1(i) r_{t-1}^2 + \beta_1(i) \mathrm{E}\left[\sigma_{t-1}^2(s_{t-1}) | s_t = i\right]. \tag{7.17}$$

This model contains both the autoregressive model and regime switching model, so it is good for simulation of complex behavior of the time series and allows also analytical forecasting [21].

## 7.2 Multifractal Models

We now apply our knowledge of multifractal measures in order to generate processes based on them. These models seem to be a good alternative to classical models and have wide field of applications in many branches of economics. The first model is exactly based on assumptions that we specified in the previous chapter, the next models are also based on multifractal formalism, but they combine multifractal scaling and Markov regime-switching.

### 7.2.1 Multifractal Model of Asset Returns

As we said in the previous chapter, we want to model the complex behavior of financial time series by time deformations, so we assume that the process is given by

$$\ln P(t) - \ln P(0) = W[\theta(t)] \tag{7.18}$$

where $\theta(t)$ is a time deformation that arise as c.d.f. of some multifractal measure, so $\theta(t) = \mu[0,t]$. We assume that $B(t)$ and $\theta(t)$ are independent processes. It is easy to generalize the process, when we assume the time deformation of fractional Brownian motion. From the last chapter we know that there exists a critical exponent of the measure $\mu$, $q(\theta)_{crit} > 1$, such that

$$P[\theta(t) > \gamma] \sim C\gamma^{q(\theta)_{crit}}. \tag{7.19}$$

Thus, for the process $r(t) = \ln P(t) - \ln P(0)$ the q-th moment is from conditions above and from scaling of Wiener process given as

$$\mathrm{E}(|r(t)|^q) = \mathrm{E}(\theta(t)^{q/2})\mathrm{E}(|W(1)|^q), \tag{7.20}$$

so therefore $q(r)_{crit} = 2q(\theta)_{crit} > 2$ and $\tau_r(q) = \tau_\theta(q/2)$, which follows from (6.24) or (6.25). The process $r(t)$ is a multifractal, because it depends on multifractal measure and moreover we know its scaling function $\tau_r$. It can be also shown that the process $P(t)$ is semimartingale and process $r(t)$ is martingale [24], so we can use a stochastic integration of the process.

Now, we define the multifractal model of asset returns according to [24], by specifying the measure $\mu$.

**Definition 7.1** *Multifractal model of asset returns (MMAR) is a time deformation process with random multiplicative measure.*

Now we focus on the distribution of returns and especially to its tail behavior. We know that there exists every moment for conservative measures, so $q_{crit} = +\infty$. That can be shown from the fact that that the measure $\mu[0,t]$ is constant from the deterministic construction of the measure, so it is bounded. On the other hand, for random multiplicative measure we have shown that there is a finite critical exponent $q_{crit} > 2$. This implies that the process has finite variance and therefore has Gaussian distribution as its limiting distribution, but on the other hand it exhibits quite wild behavior, which we can compare with truncated Lévy flight, that was created in quite artificial fashion.

### 7.2.2 Markov Switching Multifractal

Inspired from previous models, we introduce the final model that combines properties of all models that we have already mentioned. We begin with construction of the model in discrete time. Firstly, we consider a state vector $M(t) = (M_1(t), M_2(t), \ldots, M_k(t))$. Every component of $M(t)$ can be in every step updated or can remain unchanged with given probability. When it has to be updated, we take one realization of random variable $M$ with some given distribution so together

$$\text{in every time } t \quad \begin{cases} M_j(t) \text{ is updated from } M & \text{with probability } \gamma_j \\ M_j(t) = M_j(t-1) & \text{with probability } 1-\gamma_j. \end{cases}$$

The distribution $p_M$ can be arbitrary, we only demand that it has positive support and $\mathrm{E}(M) = 1$. From that it implies that $M_j(t)$ are independent and the mean of every component is one.

The volatility is then given as a product of all components, so

$$\sigma^2(M(t)) = \sigma^2 \prod_{j=1}^{k} M_j(t) \tag{7.21}$$

and returns are given by

$$r(t) = \sigma(M(t))\eta(t) \tag{7.22}$$

as usual. This approach has a great advantage, because every component of the state vector represents actual volatility of some economical cycle. We specify the switching of regimes of these cycles by defining transition probabilities $\gamma_j$.

$$\gamma_j = 1 - (1-\gamma_1)^{b^{(k-1)}} \tag{7.23}$$

where constant $\gamma_1 \in (0,1)$ and $b \in (1,\infty)$. If we consider a small parameter $\gamma_1$, then transition rates for small $j$ are approximately $\gamma_1 b^{j-1}$, so they grow geometrically, but for larger $j$ it stops to increase geometrically and the cycles are smaller, and it is so defined, that every $\gamma_j < 1$. The model seems to be very plausible, because it connects the properties of unpredicted regime switches with presence of longer periods of similar behavior and is quite easily defined.

We can assign the random variable $M$ in order to get more precise results. In the simplest case, when $M$ is binomial with values $m_0$ and $m_1 = 2 - m_0$ (from the mean condition) is therefore for given $k$ described by only four variables $\sigma^2$, $m_0$, $\gamma_1$ and $b$. Compared to GARCH(p,q) is the number of need variables smaller, because by GARCH(p,q) we needed $p + q + 2$ parameters. In the case of Markov-switching models we need the second power of Markov states, which even for small number of states is quite large number. We will compare these models a bit more at the end of this chapter.

### 7.2.3 Continuous Time MSM

As the final result of this chapter we generalize MSM into continuous time. Similarly to discrete version we introduce a state vector $M(t) = (M_1(t), \ldots, M_k(t))$ for $t > 0$. The update of $M(t)$ can be formulated also similarly:

$M_j(t+\mathrm{d}t)$ is updated from $M$ with probability $\gamma_j \mathrm{d}t$
$M_j(t+\mathrm{d}t) = M_j(t)$ with probability $(1-\gamma_j)\mathrm{d}t$

where $\gamma_j = \gamma_1 b^{j-1}$. Here are transition probabilities exactly geometrical, the reason is that it can be shown that the discrete MSM weakly converges to continuous time MSM if we choose transition probabilities in this way [7]. The volatility is again

$$\sigma^2(M(t)) = \sigma^2 \prod_{j=1}^{k} M_j(t) \tag{7.24}$$

and the log-price $r(t) = \ln P(t) - \ln P(0)$ is defined through a stochastic differential equation

$$\mathrm{d}r(t) = \mu \mathrm{d}t + \sigma(M(t))\mathrm{d}W(t) \tag{7.25}$$

where $\mu \mathrm{d}t$ is (constant) drift term.

We want to show the connection between MMAR and MSM, as well. We therefore assume the continuous-time MSM and consider that number of frequencies $k$ rises to infinity. The problem in the differential representation is that the volatility is given by an infinite product. Another possibility is to use an integral approach, so we introduce a time deformation

$$\theta_k(t) = \int_0^t \mathrm{d}t' \sigma^2(M_k(t')) \tag{7.26}$$

It can be shown that $\theta_k$ weakly converges to a deformation $\theta_\infty$ with finite expectation [7]. That means that differential representation (7.25) can be formulated in alternative integral equation

$$\ln P(t) - \ln P(0) = \mu t + W[\theta_\infty(t)]. \tag{7.27}$$

Thus, we see that a continuous MSM with countably many frequencies is the multifractal processes with time deformation $\theta_\infty$. The advantage of this model in comparison with MMAR is that it can be easily generalized for the infinite interval and it is generated independently of the interval division.

**Comparison of models:** At the end of the chapter, we would like to compare models of asset returns and show advantages of multifractal models, especially MSM. We divide our arguments into two groups and begin with theoretical arguments. Main advantages of MSM are persistent behavior in all scales followed by change of behavior and existence of large deviations and low number of free parameters at the same time. These properties are very important for practical calculations. In comparison to other models, MSM has very complex behavior that can be fitted to many series very well. That lead us to the second part of arguments, empirical observations, in-sample and out-sample simulations and comparison of likelihood functions. We will not do this kind of analysis, but a very deep analysis is in [7] and often shows that MSM performs good results for many applications.

# Chapter 8

# Real Financial Markets: Memory, Jumps

In the last chapter we would like to show some practical results from real financial markets and compare them with theory. In the first chapter we introduced a stochastic calculus based on Wiener process. For asset returns is often in many applications assumed that it follows a geometric Brownian motion, as in case of Black-Scholes equation for option pricing, portfolio theory and many others. We want to show that Brownian motion seems to be an elegant model for financial market simulations, nevertheless, in some cases it oversimplifies the complexity of studied system and neglects some effects that are caused by various situations on the market. We try to apply the theory from the first part of the work to some concrete financial data in order to show the presence of such effects like memory, or large jumps in financial time series.

| stock index | $\alpha$ | $\beta$ | $\gamma$ | $c$ |
|---|---|---|---|---|
| S&P 500<br>3. 1. 1950 | 1.623 | -0.081 | 0.00493 | 0.00052 |
| DJIA<br>10. 1. 1928 | 1.451 | -0.044 | 0.00520 | 0.00049 |
| NASDAQ<br>5.2.1971 | 1.366 | -0.178 | 0.00539 | 0.00127 |
| DAX<br>26. 1. 1990 | 1.647 | -0.137 | 0.00786 | 0.00090 |
| FTSE 100<br>2. 4. 1984 | 1.711 | -0.156 | 0.00619 | 0.00067 |
| EUR/USD<br>12. 10. 2003 | 1.804 | -0.084 | 0.00416 | 0.00010 |
| Gold ($/oz.)<br>3.1.1973 | 1.449 | 0.046 | 0.00599 | 0.00010 |

Table 8.1: In the table are calculated coefficients of stable distribution that was fitted to the density of returns of given indices by MLE estimation, and for comparison is also added one exchange rate and one commodity. The begin of period used for the fit is listed under the index and the end of the period is for all the end of 2010. We can see that $\alpha$ parameter is often smaller than two which indicated heavy-tail behavior. The estimation was done by R package *fBasics* [37].


Figure 8.1: Standard and Poor's 500 index between years 1985 - 2010. In this period were financial markets affected by many unexpected movements, crises and crashes.

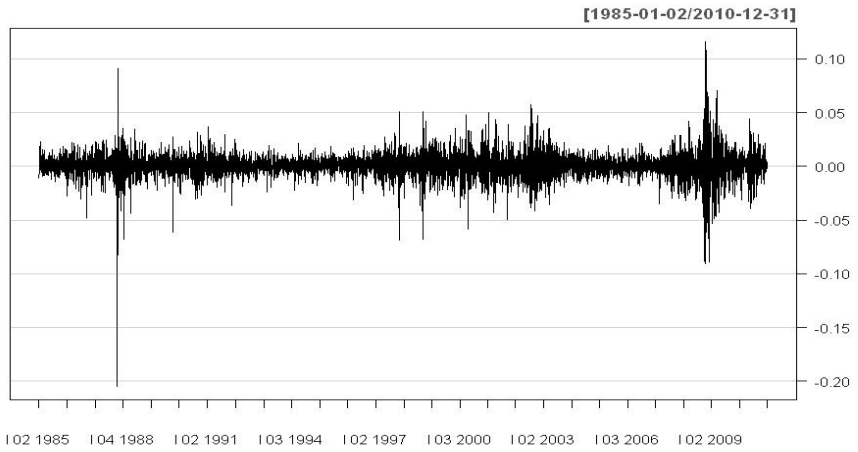



Figure 8.2: Daily returns of S&P 500. We can distinctly recognize large shocks and periods with large volatility that are often connected to some period of crisis.

For analysis an index Standard and Poor's 500 was chosen, which is composed of 500 largest shares traded on New York Stock Exchange. The studied period is 1985 - 2010. We chose this period, because it contains many interesting situations, financial crashes and unexpected evolution. The first big crash was in 1987 and it began on Black Monday, October 19. Then around year 2000 a speculative bubble called "dot-com bubble" arose, but affected more technological indexes, like NASDAQ. In the year 2001 the index dropped because of the terrorist attack. In 2007 began an American mortgage crisis, followed by a American and Global financial crisis.

Table 8.1 show us a comparison of parameters of stable distribution that fits best to the return distribution for S&P 500 and a few other indices, exchange rate and commodity price. We see that all of them have the value of the $\alpha$ parameter much smaller than two, which signalizes heavy tails. We can see that

| year | $\alpha$ | $\beta$ | $\gamma$ | $c$ | R/S | DFA |
|---|---|---|---|---|---|---|
| 1985 | 1.876 | 1.000 | 0.00427 | 0.00029 | 0.666 | 0.596 |
| 1986 | 1.851 | -1.000 | 0.00577 | 0.00182 | 0.660 | 0.534 |
| 1987 | 1.543 | -0.223 | 0.00773 | 0.00201 | 0.541 | 0.458 |
| 1988 | 1.662 | 0.158 | 0.00579 | 0.00032 | 0.457 | 0.438 |
| 1989 | 1.776 | -0.185 | 0.00462 | 0.00141 | 0.754 | 0.455 |
| 1990 | 1.905 | -0.375 | 0.00678 | 0.00007 | 0.783 | 0.554 |
| 1991 | 1.804 | 0.464 | 0.00559 | 0.00034 | 0.745 | 0.548 |
| 1992 | 2.000 | 0.157 | 0.00430 | 0.00019 | 0.374 | 0.381 |
| 1993 | 1.758 | 0.211 | 0.00326 | 0.00013 | 0.465 | 0.408 |
| 1994 | 1.757 | -0.344 | 0.00382 | 0.00037 | 0.528 | 0.460 |
| 1995 | 1.835 | -0.043 | 0.00317 | 0.00123 | 0.429 | 0.611 |
| 1996 | 1.883 | -1.000 | 0.00481 | 0.00155 | 0.609 | 0.504 |
| 1997 | 1.864 | 0.035 | 0.00695 | 0.00117 | 0.828 | 0.487 |
| 1998 | 1.647 | -0.260 | 0.00695 | 0.00184 | 0.478 | 0.528 |
| 1999 | 2.000 | 0.263 | 0.00803 | 0.00077 | 0.305 | 0.380 |
| 2000 | 1.842 | 0.425 | 0.00900 | -0.00101 | 0.461 | 0.382 |
| 2001 | 1.822 | 0.038 | 0.00856 | -0.00062 | 0.717 | 0.564 |
| 2002 | 1.876 | 1.000 | 0.01080 | -0.00267 | 0.421 | 0.496 |
| 2003 | 1.887 | 0.244 | 0.00714 | 0.00007 | 0.670 | 0.511 |
| 2004 | 2.000 | -0.207 | 0.00493 | 0.00037 | 0.574 | 0.429 |
| 2005 | 2.000 | -0.007 | 0.00457 | 0.00014 | 0.559 | 0.377 |
| 2006 | 1.748 | 0.067 | 0.00389 | 0.00042 | 0.563 | 0.460 |
| 2007 | 1.524 | -0.374 | 0.00537 | 0.00150 | 0.537 | 0.404 |
| 2008 | 1.418 | -0.089 | 0.01178 | 0.00097 | 0.559 | 0.304 |
| 2009 | 1.620 | -0.202 | 0.00969 | 0.00181 | 0.608 | 0.495 |
| 2010 | 1.499 | -0.132 | 0.00590 | 0.00184 | 0.736 | 0.511 |

Table 8.2: The table shows us fitted parameters of daily return distribution of Standard and Poor's 500 index between years 1985 - 2010. Parameters differ between years. In the time of no financial shocks or disasters moves $\alpha$ around 2 and in the time of crisis it falls down to values around 1.5. Values are often far from average value in table 8.1. Two last columns show us the estimation of Hurst exponent with R/S method and DFA. Hurst exponent also varies widely around 0.5. Different values are caused by the sensitivity of both methods to the depth of the sample, but we can see some trends that are for parameters estimated by both methods common. For estimation of Hurst parameter was used R package *fractal* [9].

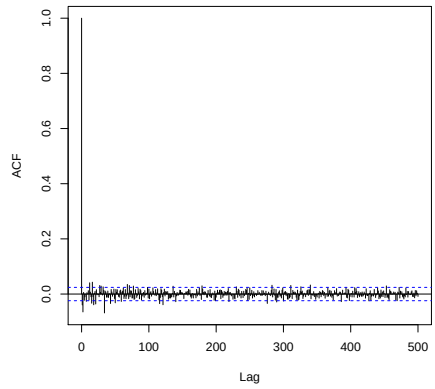
it is similar for all assets, regardless of the kind. In the table 8.2 we can see the estimation of stable parameters for S&P 500 for every year. We see that values are different year to year - $\alpha$ is usually smaller in the period of crisis, but also $\beta$ fluctuates from year to year. We have also estimated the Hurst exponent with two methods described earlier for every year. Here we also see a rapid variation around the value 0.5, which is the value for Brownian motion. This table shows us that we have to take into account both the heavy tail property and memory effects.

The presence of memory is also illustrated on collection of figures 8.3, where we see the graph of autocorrelation function of returns and the graph of autocorrelation function of absolute value of daily returns. We see the difference in typical correlation time. In the first case the correlation time is much shorter than one day (many authors are talking about the order of minutes as in [27]) and in the second case the typical correlation time is in the order of days, or

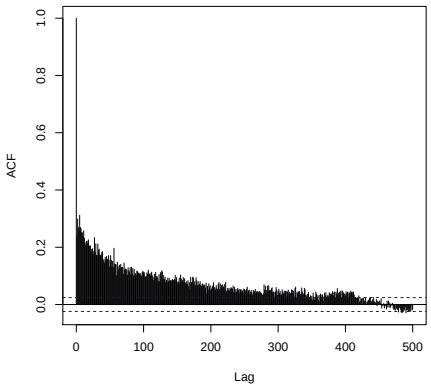
even months. That supports the idea that there are economic cycles, i.e. periods with large fluctuations and higher volatility, that means with more difficult prediction of future evolution, and periods of relative calm, where are observed large changes only rarely.

On the third figure is the distribution of daily returns and for comparison Gaussian distribution with fitted mean and variance from the data. We can observe the presence of heavy tails, which means higher probability of large returns than in case of Gaussian distribution. We have seen that correlation and large jumps in time series have very often common reasons and usually if we observe one of them, the other is present as well.

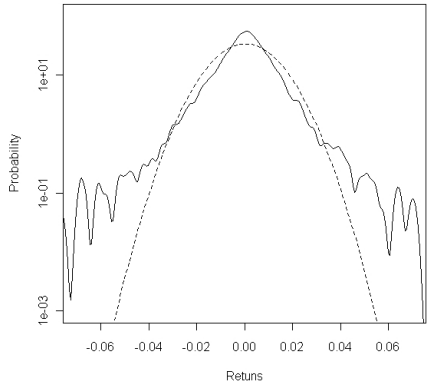
Altogether, we have shown on the example of one concrete index that it is necessary to take into account some nontrivial phenomena in order to simulate and predict future evolution on financial markets. We have not talked about other things, like correlation between shares, markets, about other models that also imply some external information to the model and many other things. There are many possibilities, when constructing a model. But in every case it is necessary to consider if the model is sufficient. Unfortunately in some theoretical applications, like option pricing theory, or portfolio theory every generalization leads to difficult relations and equations that are tough to solve.



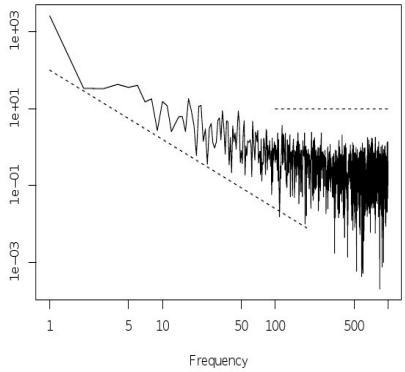

(a) autocorrelation of returns




(b) autocorrelation of absolute returns




(c) return distribution




(d) power spectrum of absolute returns

Figure 8.3: In the figures are analyzed daily returns of S&P500 index in the period 1985-2010. On figure **(a)** is the autocorrelation function of daily returns. We see that there are practically no correlations, i.e. correlation time is shorter than one day. On figure **(b)** is the autocorrelation function of absolute value of returns, here we can observe exponential decay, which means that the size of returns is correlated, which points to presence of periods with large fluctuations and with small fluctuations. Figure **(c)** shows probability distribution of returns in logarithmic plot and shows estimated Gaussian distribution (dashed line). We can see that the tails of the distribution are much broader than in the case of Gaussian distribution. Finally, on figure **(d)** is drawn a log-log plot of power spectrum of absolute returns. We can see that for small lag it exhibits a decay, but for larger lag we see that records are practically uncorrelated. For comparison there are two lines to recognize the trend of power spectrum. The first dashed line shows the decay $\sim \omega^{-1.8}$, the second line is constant.

# Chapter 9

# Conclusion

The main task of this work was to extend models that are commonly used in financial markets and to show theoretical background leading to these models. The thesis goes beyond Central limit theorem and discusses important properties that are typically observed on financial markets. We have shown the connection between physical systems and financial markets and introduced the calculus for stochastic processes. The stochastic calculus provides us a good way to deal with systems, such as non-equilibrium physical systems or financial time series.

Important property of real processes, no matter if economical, physical or any other, is presence of correlations and inner structure. Correlations in time are typical for systems that exhibit some memory behavior. Although correlations in returns are negligible for periods longer than a few minutes, for absolute values and volatility the correlations are observed for time lag in the order of days or weeks. On the other hand there are also observed unexpected jumps. The theoretical background of that behavior can be found in scaling and self-similarity, which shows some inner structure and enables to classify processes and reveal superdiffusive or subdiffusive behavior and presence of huge fluctuations by geometrical instruments. The concept of scaling is not important only for stochastic processes, but also in critical phenomena in statistical physics and other problems.

There are many ways how to introduce models with these properties: Lévy flights [11], Lévy stochastic processes [10], superstatistics [2], Tsallis entropy [36] and many others. We have chosen processes with multi-scaling properties generated by multifractal time deformations. The concept of multifractals has wide range of applications [17], and also in this case provides an elegant way to define the transformation between clock time and trading time. We saw that these processes provide a very good diversity and are able to produce many realistic forecasts. The concept of multifractal deformations seems to be applicable to other branches, for example to biology or physics, because it is built on very general ideas of multi-scaling. The class of processes with multifractal behavior is wide enough to produce interesting behavior, but the properties of these processes are given by an idea of fluctuating Hurst exponent and presence of some typical cycles.

For the future work I see the possibility to focuse more on a few different topics. Firstly, to make a detailed analysis of the MSM model with all of its

generalizations, to do precise simulations and comparison with other models, secondly to study applications of these models to portfolio theory, option pricing theory, risk-management and other fields of economics and/or econophysics. There is also possible to apply the concept of multi-scaling processes to some other areas, where can play also a handy role and can classify some important properties of these phenomena. It is also important to connect this approach to other concepts like theory of entropy, Bayesian statistics and other fast growing scientific areas. I guess that there is still enough place to study these systems as financial markets, because they still offer the possibility to explore a lot of interesting features and to broaden the knowledge about this interesting branch.

# Appendix A

# Stable distributions

We ask about the limiting distribution of the infinite sum of independent random variables $X_n$

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

If $X_n$ have finite variance, then the limiting distribution is known by Central limit theorem (here is for simplicity presented the version with identically distributed random variables).

**Theorem A.1 (Central limit theorem)** *Let $X_i$ be i.i.d. variables with finite mean $\mu$ and finite variance $\sigma^2$. We define $Y_n = \sum_{i=1}^{n} X_i$. Then*

$$\frac{Y_n - \mu n}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1). \tag{A.2}$$

The proof can be found in many books, for example in [14].

The question is,if there are any limiting distributions for distributions with infinite, or not defined variance. In order to show answer the question we need to define some new terms.

**Definition A.2** *Probability distribution $p(x)$ is stable, if it is invariant under convolution:*

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

Equivalently, we can reformulate the criterion with Fourier transform

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

so if two Fourier images of stable distribution are multiplied, then the functional form remains unchanged.

The full answer to our question give us next two theorems:

**Theorem A.3** *A probability density $L(x)$ can be limiting distribution of sum* (A.1) *of independent, randomly distributed variables, only if $L(x)$ is stable distribution.*

**Theorem A.4** *A probability density $L_{\alpha\beta}(x)$ is stable, if and only if logarithm of its characteristic function has this form:*

$$\ln L_{\alpha\beta}(k) = ick - \gamma|k|^{\alpha}\left(1 + i\beta sgn(k)\omega(k,\alpha)\right) \tag{A.5}$$

*where:* $0 < \alpha \leq 2, \quad -1 \leq \beta \leq 1, \quad \gamma \geq 0, \quad c \in \mathbb{R},$

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

Proofs are to find in [16].

Parameters, that influence the behavior of the distribution are $\alpha$ and $\beta$, $\gamma$ and $c$ are scaling, resp. shifting parameters. Let us consider $\beta = 0$, so we have a symmetric distribution.

For large values the distribution decays polynomially:

$$L_{\alpha}(x) \simeq \frac{l_{\alpha}}{|x|^{1+\alpha}} \quad \text{for} \quad |x| \to \infty. \tag{A.7}$$

This property is a direct corollary of the asymptotic expansion

$$L_{\alpha}(x) = -\frac{1}{\pi}\sum_{n=1}^{\infty}\frac{(-\gamma)^n}{n!}\frac{\Gamma(\alpha n+1)}{|x|^{\alpha n+1}}\sin\left(\frac{\pi\alpha n}{2}\right). \tag{A.8}$$

We show it for $c = 0$. The probability distribution is a inverse Fourier transform of its characteristic function:

$$\begin{aligned} L_{\alpha}(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}k e^{-\gamma|k|^{\alpha}}e^{ikx} = \frac{1}{2\pi}\int_{0}^{\infty}\mathrm{d}k\ e^{-\gamma k^{\alpha}}\left(e^{ikx} + e^{-ikx}\right) \\ &= \frac{1}{2\pi}\int_{0}^{\infty}\mathrm{d}k\ e^{-\gamma k^{\alpha}} 2\Re\left(e^{ikx}\right) = \frac{1}{\pi}\Re\left[\int_{0}^{\infty}\mathrm{d}k\ e^{-\gamma k^{\alpha}}e^{ikx}\right] \end{aligned}$$

we expand the exponential in the integral to a power series, integrate and with an expression for $\Gamma$ function we get

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

We calculate the real part and with identity

$$\Re\left((\pm i)^{\alpha n+1}\right) = -\sin\left(\frac{\pi\alpha n}{2}\right) \tag{A.9}$$

we get the expression (A.8).

# Bibliography


[1] *Stochastic Processes, Formalism and Applications*, Lecture Notes in Physics. Springer, 1983.

[2] Christian Beck and E. D. G. Cohen. Superstatistics. *Physica A*, (322), 2003.

[3] Jean-Phillipe Bouchard and Marc Potters. *Theory of Financial Risks: From Statistical Physics to Risk Management*. CUP, Cambridge, 2001.

[4] G. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice-Hall, third edition, 1994.

[5] Zdzisław Brzeźniak and Tomasz Zastawniak. *Basic Stochastic Processes*. Springer, London, 1999.

[6] P.L. Butzer and U. Wespthal. An introduction to fractional calculus. In L. Hilfer, editor, *Applications of Fractional Calculus in Physics*, pages 1–85. World Scientific Publ. Co., 2000.

[7] L. E. Calvet and A. J. Fisher. *Multifractal Volatility: Theory, Forecasting and Pricing*. Academic Press, 2008.

[8] W. T. Coffey, Yu. P. Kalmykov, and J.T.Waldron. *Langevin Equation: With applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*. World Scientific, 1996.

[9] William Constantine, Donald Percival (Applied Physics Laboratory, and University of Washington). *fractal: Insightful Fractal Time Series Modeling and Analysis*, 2010. R package version 1.0-3.

[10] Rama Cont and Peter Tankov. *Financial Modelling with Jump processes*. CRC Press, 2004.

[11] A. A. Dubkov, B. Spagnolo, and V. V. Uchaikin. Lévy flight superdiffusion: An introduction. *arXiv:0810.1492 [cond-mat.stat-mech]*, 2008.

[12] A. Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der Physik*, 1905.

[13] Kenneth Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, Inc., 1990.

[14] Willian Feller. *An Introduction to Probability theory and its Applications*, volume 1. Wiley, third edition, 1967.

[15] C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and Natural Sciences*. Springer, second edition, 1985.

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

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

[18] Steven 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), 1993.

[19] Petr Jizba and Toshihico Arimitsu. The world according to rényi: thermodynamics of multifractal systems. *Annals of Physics*, 312(1):17 –59, 2004.

[20] M.S. Keshner. 1/f noise. *Proceedings of the IEEE*, 70(3):212–218, march 1982.

[21] Franc Klaassen. Improving garch volatility forecasts with regime-switching garch. *Empirical Economics*, 27:363–394, 2002.

[22] Jan Korbel. *Application of Multifractals in Financial Markets*. Research work. Faculty of Nuclear Sciences and physical Engeneering, CTU in Prague, 2011.

[23] B. B. Mandelbrot. *Multifractals and 1/f Noise: Wild Self-Affinity in Physics (1963-1976)*. Springer, 1999.

[24] B. B. Mandelbrot, L. E. Calvet, and A. J. Fisher. A multifractal model of asset returns. *Cowles Foundation Discussion Paper 1164*, 1997.

[25] B.B. Mandelbrot. Heavy tails in finance for independent or multifractal price increments. In S. T. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, chapter 1. Elsevier, 2003.

[26] Beniot B. Mandelbrot and John W. van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968.

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

[28] H. Nyquist. Thermal agitation of electric charge in conductors. *Phys. Rev.*, 32:110–113, Jul 1928.

[29] Leonard Ornstein and George Eugene Uhlenbeck. On the theory of the brownian motion. *Phys. Rev.*, 36, 1930.

[30] Wolfgang Paul and Jörg Baschangel. *Stochastic Processes: From Physics to Finance*. Springer, Berlin, 1999.

[31] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, and A.L. Goldbergerz. Mosiac organization of dna nucleotides. *Phys. Rev. E*, 49(2):1685–1689, 1994.

[32] Igor 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.

[33] Bo Qian and Khaled Rasheed. Hurst exponent and financial market predictability. In *IASTED conference on "Financial Engineering and Applications"(FEA 2004)*, pages 203–209, 2004.

[34] Hannes Risken. *The Fokker-Planck Equation: Methods of Solution and Application*. Springer, second edition, 1989.

[35] Jeffrey A. Ryan. *quantmod: Quantitative Financial Modelling Framework*, 2010. R package version 0.3-15.

[36] Constantino Tsallis, Silvio Lévy, André Souza, and Roger Maynard. Statistical-mechanical foundation of the ubiquity of lévy distributions in nature. *Phys. Rev.*, 75(20), 1995.

[37] Diethelm Wuertz and Rmetrics core team members. *fBasics: Rmetrics - Markets and Basic Statistics*, 2010. R package version 2110.79.