CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Nuclear Sciences and Physical Engineering

Department of Physics

Bachelor thesis

Applications of econophysics in quantitative finance

Václav Svoboda

Supervisor: Petr Jizba

Prague, 2014

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ
V PRAZE
Fakulta Jaderná a Fyzikálně Inženýrská
Katedra Fyziky

Bakalářská práce

Aplikace ekonofyziky v kvantitativním finančnictví

Václav Svoboda

Supervisor: Petr Jizba

Praha, 2014

Prohlášení:

Prohlašuji, že jsem svou bakalářskou práci vypracoval samostatně a použil jsem pouze podklady (literaturu, software, atd.) uvedené v přiloženém seznamu.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu 60 Zákona .121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze dne

Title:
Applications of econophysics in quantitative finance

Author: Václav Svoboda

Specialization: Mathematical physics

Sort of project: Bachelor thesis

Supervisor: Petr Jizba
Consultant:

Abstract: Econophysics tries to describe and predict behaviour of financial markets. In the first chapters we introduce the most useful mathematical tools for describing financial markets or dynamical systems in general - stochastic calculus, fractal geometry and statistics generalized beyond central limit theorem. In last two chapters we will focus on concrete applications in finance, namely we will derive the famous Black-Scholes option pricing formula and use a parallel between extensive and non-extensive statistical mechanics for generalization of this formula.

Key words: econophysics, stochastic processes, generalized statistics, fractal geometry, option pricing, non-extensive statistical mechanics

Název práce:
Aplikace ekonofyziky v kvantitativním finančnictví

Autor: Václav Svoboda

Abstrakt: Cílem ekonofyziky je popsat a předvídat chování finančních trhů. V prvních kapitolách představíme nejpotřebnější matematické nástroje k popisu finačních trhů respektive dynamických systemů obecně - stochastický počet, fraktální geometrii a statistiku zobecněnou za centrální limitní teorém. V posledních dvou kapitolách se budeme více soustředit na konkrétní aplikace ve finančnictví, především odvodíme slavnou Black-Scholesovu rovnici pro cenu opcí a dále využijeme paralelu mezi mezi extensivní a neextensivní statistickou mechanikou k zobecnění této rovnice.

Klíčová slova: ekonofyzika, stochastické procesy, zobecněná statistika, fraktální geometrie, oceňování opcí, neextensivní statistická mechanika

Acknowledgement

I would like to thank my supervisor Ing. Petr Jizba, PhD., for many useful consultations, suggestions and advices that have helped me to discover for me a new and exciting field of study - econophysics. His lectures helped me greatly with my understanding of many problems and also made me realize how complicated, interesting and interdisciplinary the econophysics is. I would also like to thank Zdeněk Svoboda for careful proofreading of my thesis.

Contents

1 Probability theory **14**
- 1.1 Preliminaries . . . 14
- 1.2 Central limit theorem . . . 16
- 1.3 Beyond central limit theorem: Levy distributions . . . 17
  - 1.3.1 Stable distributions . . . 17
  - 1.3.2 Generalized central limit theorem . . . 21
  - 1.3.3 Infinitely divisible distributions . . . 22

2 Stochastic calculus **25**
- 2.1 Stochastic processes . . . 25
  - 2.1.1 Brownian motion . . . 27
  - 2.1.2 Conditional expectation . . . 29
  - 2.1.3 Martingales . . . 34
- 2.2 Stochastic integral . . . 36
  - 2.2.1 Ito integral . . . 37
  - 2.2.2 Ito lemma . . . 39
  - 2.2.3 Stochastic differential equation . . . 42
- 2.3 Levy processes . . . 45
  - 2.3.1 Stable processes . . . 48
- 2.4 Different approach to stochastic processes: Fokker-Planck equation 49
  - 2.4.1 Markov processes . . . 49
  - 2.4.2 Fokker-Planck equation . . . 52

3 Fractal geometry **54**
- 3.1 Fractals . . . 54
  - 3.1.1 Haussdorf dimension and measure . . . 56
  - 3.1.2 Box counting dimension . . . 59
  - 3.1.3 Self similar sets . . . 60
- 3.2 Random fractals . . . 61
  - 3.2.1 Brownian motion . . . 61
  - 3.2.2 Fractional Brownian motion . . . 62
  - 3.2.3 Levy stable processes . . . 63
- 3.3 Multifractals . . . 64
  - 3.3.1 Multiftactal spectrum . . . 64

4 Black-Scholes formula **68**
4.1 Financial markets 68
4.1.1 Options 68
4.2 Black-Scholes option pricing 69
4.2.1 Building portfolio 70
4.2.2 Option pricing problem 70
4.2.3 Change of measure 72
4.3 Beyond Black-Scholes formula 74
4.3.1 Implied volatility 74
4.3.2 Beyond geometric Brownian motion 75

5 From thermodynamics to finance **77**
5.1 Boltzmann-Gibbs statistics 77
5.1.1 MaxEnt principle 77
5.1.2 Limitations of Boltzmann-Gibbs statistics 79
5.2 Non extensive statistical mechanics 79
5.2.1 Tsallis entropy 79
5.2.2 Quantum calculus 81
5.3 Non-Gaussian option pricing 84

A **90**

B **91**

C **92**

List of Figures

3.1 Koch curve 55

[chapter]

Introduction

Econophysics is a field of study, which uses methods of physics to analyse and predict behaviour of financial markets, but econophysics is only branch of the much more established field of mathematical finance. L.Bachelier is considered as a founder of this field with his doctoral thesis from 1900, where he used Brownian motion to describe the a development of prices on financial markets. It was the first attempt to use advanced mathematics to describe a complex system of financial markets. Since then many improvements and new ideas were made in this area, we will name just few of them.

In 1973 F.Black and M.Scholes published the first rigorous option pricing formula. It was a huge breakthrough correct prices of options were just guessed till then. Their formula was widely celebrated and both of them were rewarded Nobel prize for it twenty years later. However in 1994 Black and Merton were in charge of a big hedge fund, they were very successful the first few years but they lost more than a half of the value of their fund and failed during financial crisis in 1997 and 1998 .

As the last but certainly not least we will mention Benoit Mandelbrot. He is best known for his work in physics and as a father of fractal geometry, but he also published number of articles about mathematical finance. His biggest contribution to mathematical finance beside a usage of fractals was idea that prices on financial markets do not follow normal Gaussian distributions but so called heavy-tailed distributions, i.e. distributions with a polynomial decay. In fact, he explained why Black-Scholes approach based on a normal distribution had to fail during the crisis even before their formula was published, but no one paid much attention to his ideas at that time.

In this thesis we will first describe a mathematical apparat that mathematical finance and econophysics use and we show some concrete applications to finance then. We hope that this thesis will explain main ideas behind econophysics and why there is a need to use physical and not just mathematical methods in finance.

In the first chapter, after stating basic results of probability theory including central limit theorem, we will explain a need for more general limit theorems and discuss an important class of distributions - stable distributions in details. In the second and third chapters we will thoroughly discuss stochastic calculus and fractal geometry with a special focus to Brownian motion. Throughout these chapters, we try to outline a connection between stochastic processes and

fractals and their applicability to finance.

In the last two chapters we will focus more on the applications. In the fourth chapter we will discuss Black-Scholes formula in details and we will explain its limitations and possible generalizations then. In the last chapter we will focus on relation between thermodynamics and statistical physics with finance.

Chapter 1

Probability theory

1.1 Preliminaries

We will start with stating some basic definitions and results of the probability theory. We will not go into details because these results are well know.

Definition 1.1. *(Probability space) Let Ω be set of elementary events, $\mathcal{F}$ be σ-algebra on Ω and μ probability measure ($\mu(\Omega) = 1$). Then we will call $(\Omega, \mathcal{F}, \mu)$ probability space.*

Definition 1.2. *(Random variable) We will call any Borel measurable function $X : \Omega \to \mathbb{R}(\mathbb{R}^n)$ a random variable.*

Definition 1.3. *(Cumulative distribution) Let $X : \Omega \to \mathbb{R}$ be a random variable on probability space $(\Omega, \mathcal{F}, P)$. Then $F(x) = P[\omega \in \Omega, X(\omega) \leq x] \equiv P[X \leq x]$ is a cumulative distribution (CD) of a random variable X.*

Definition 1.4. *(Probability density function) Let $F(x)$ be a cumulative distribution function. Then if exists a function $p(x)$ such as $F(x) = \int_{-\infty}^{x} p(y)\mathrm{d}y$ then we we will call $p(x)$ a probability density function (PDF).*

Remark. *We can also define PDF as $d(P \circ X^{-1}) = p(x)dx$. An existence and a uniqueness is guaranteed by Radon - Nikodym theorem (see Appendix A) for all equivalent measures with Lebesque measure. We can easily see from this definition that $P[a < X \leq b] = \int_{a}^{b} p(x)\mathrm{d}x$.*

If PDF exists, we will call given CD absolutely continuous. We will consider almost only absolutely continuous CD throughout this thesis.

Definition 1.5. *We define expected (mean) value for every random variable X by $EX = \int_{\Omega} X\mathrm{d}P = \int_{-\infty}^{\infty} xp(x)\mathrm{d}x$.*

Remark. *Analogically, we can define $Ef(X)$ for every Borel measurable function f. Specially for $f(x) = x^k$ we call EX^k kth moment and $E(X - EX)^k$ kth central moment. The second central moment is usually called variance and a following relation holds*

$$VarX = E(X - EX)^2 = EX^2 - (EX)^2 \tag{1.1}$$

For random variables X, Y on same probability space we define covariance by following relation

$$Cov(X, Y) = E((X - EX)(Y - EY)) = E(XY) - EXEY \tag{1.2}$$

If $Cov(X, Y) = 0$, we say that X and Y are uncorrelated.

Definition 1.6. *(Random vector) Let X_i be random variables to $\mathbb{R}$ than we call $X = (X_1, X_2, ..., X_n)$ random vector. Cumulative distribution and PDF of X is defined in following way*

$$F(x_1, ..., x_n) = P[X_1 \leq x_1, ..., X_n \leq x_n] = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p(x_1 ... x_n) \mathrm{d}x_1 ... \mathrm{d}x_n$$

Expected value $EX = (EX_1, ..., EX_n)$

Definition 1.7. *Let X_1, X_2 be random variables then we say that X_1 and X_2 are statistically independent $\Leftrightarrow p_{X_1,X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2.)$.*
Independence of random variables $(X_1, X_2, ..., X_n)$ is defined analogically.

Definition 1.8. *(Conditional CD) Let X, Y be random variables then we a define conditional CD*

$$F_{X|Y}(x \mid y) = \lim_{\varepsilon \to 0} P[X \leq x \mid y - \varepsilon < Y \leq y + \varepsilon]$$

Remark. *Following relation for PDF of $F_{X|Y}$ can be easily proven .*

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \tag{1.3}$$

Clearly $p_{X|Y}(x \mid y) = p_X(x)$ if X and Y are independent. Further, we can define a conditional expected value in following sense

$$E_{X|Y}(\quad \mid y) = \int_{-\infty}^{\infty} x p_{X|Y}(x \mid y) \mathrm{d}x$$

We will simplify marking for a conditional PDF with more arguments

$$p_{n|m}(x_1, ..., x_n \mid y_1, ..., y_m) = \frac{p_{n+m}(x_1, ..., y_m)}{p_m(y_1, ..., y_m)}$$

We will end this section with an important theorem about transformation of PDF.

Theorem 1.1. *Let $X : \Omega \to \mathbb{R}^n$ be a random vector with absolutely continuous CD, $g : \mathbb{R}^n \to \mathbb{R}^n$ differentiable, invertible and regular function. Then a random vector $Y = g(X)$ has an absolutely continuous CD and*

$$p_Y(y) = p_X(g^{-1}(y))|\mathbb{J}_{g^{-1}}(y)|$$

1.2 Central limit theorem

In physics, we examine effects that are caused by many small, random influences very often. Typical example is a pressure of gas on walls of a containment vessel.

It is not surprising that a similar situation arises when we are trying to describe behaviour of financial markets. As we already mentioned, early attempts to describe financial markets were often somehow based on Gaussian distribution, a reason for that is in following well known theorem.

Remark. *Random variable X obeys normal (Gaussian) distribution*
$X \sim N(\mu, \sigma^2) \Leftrightarrow p_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x-\mu)^2}{2\sigma^2}$

Theorem 1.2. *(CLT) Let X_i be identically and independently distributed (iid) random variables with finite variance, $EX_i = \mu$ and $Var(X_i) = \sigma^2$ then*

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathcal{D}} X \sim N(0, 1)$$

Remark. $\xrightarrow{\mathcal{D}}$ *means convergence in distribution respectively a weak convergence, for a definition see appendix B.*

In other words, a sum of iid random variables with a finite variance converge to a normal distribution. An assumption about identic distribution of random variables may be relaxed, for example, if all X_i have the same expected value but different bounded variance convergence to a normal distribution remains.

Assumptions of CLT are not very restrictive, indeed many systems do follow normal distribution, but many do not. For example, it turns out on financial markets that CLT works quite well when market is in a equilibrium, but when market becomes unstable, CLT stops working at all. This reminds a situation in equilibrium and non - equilibrium thermodynamics, this connection will be discussed in the last chapter.

What happens when a sequence of random variables has not a finite variance? Or when variables are not independent? The first question will be in detail discussed in the next section, the second one is much more complicated and will be briefly mentioned in the last chapter.

1.3 Beyond central limit theorem: Levy distributions

In this section, we examine distributions that can be a limiting distribution of a sum of iid random variables with an infinite variance. A typical example are variables with PDF in the following form:

$$p(x) \sim \frac{1}{|x|^{1+\alpha}} \quad for \quad |x| \to \infty \quad and \quad \alpha \in (0,2)$$

Convergence of sequence of random variables is as in the case of CLT presumed in distribution and after proper normalization .

$$S_n = \frac{1}{a_n}(\sum_{i=1}^{n} X_i - b_n)$$

So $a_n = \sqrt{n}\sigma$ and $b_n = n\mu$ is a proper normalization in the case of random variables with finite variance .

1.3.1 Stable distributions

Proves of theorems in following sections are usually very lengthy and can be found [4].

Definition 1.9. *Probability density is stable if it does not change functional form under convolution i.e. for all $a_i > 0$ and all $b_i \in \mathbb{R}$:*

$$p(a_1 y + b_1) * p(a_2 y + b_2) \equiv \int_{-\infty}^{\infty} p(a_1(x-y)+b_1)p(a_2 y + b_2)\mathrm{d}y = p(ax+b)$$

Where $a > 0$ and b is any real constant.

Remark. *An idea behind this definition becomes clear, when we realize that when we have two independent variables with PDFs p_1, p_2 then the sum of them has PDF $p = p_1 * p_2$, this follows from theorem (1.1). So if we sum stable, independent random variables, they preserve their distribution and a change will appear only in scale.*

Normal distribution is an example of stable distribution , this fact is obvious because it is well known that when we sum independent random variables $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

But for the sake of example, we will verify this relation and stability of normal distribution directly from the definition. A helpful tool for verification of stability is Fourier transform. Fourier transformation of PDF is usually called a characteristic function and is defined by:

$$\varphi(k) = \mathcal{F}[p(x)](k) = \int_{-\infty}^{\infty} p(x) exp(ikx)\mathrm{d}x \tag{1.4}$$

We will use the fact that Fourier transform of convolution is a product of Fourier transforms.

$$\mathcal{F}[f * g] = \mathcal{F}[g]\mathcal{F}[f]$$

Characteristic function of random variable $X \sim N(\mu_1, \sigma_1^2)$ *has form* $\varphi_X(k) = exp(ik\mu_1 - \frac{1}{2}k^2\sigma_1^2)$. *So we can see that characteristic function of sum* $X + Y$ *equals :*

$$\varphi_{X+Y}(k) = exp(ik(\mu_1 + \mu_2) - \tfrac{1}{2}k^2(\sigma_1^2 + \sigma_2^2))$$

and this is characteristic function of random variable with distribution $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. *So we have verified stability of normal distribution.*

An importance of stable distributions follows from Levy-Khintchin theorem.

Theorem 1.3. *(Levy, Khintchin) A probability density can only be a limiting distribution of a sum of iid random variables, if it is stable.*

Levy and Khintchin have also found the most general class of a stable distributions. This class is often called Levy distributions or α - stable distributions.

Theorem 1.4. *A probability density* $p_{\alpha,\beta}(x)$ *is stable* $\Leftrightarrow$ *logarithm of its characteristic function has form :*

$$\ln \varphi_{\alpha,\beta}(k) = i\gamma k - c|k|^\alpha(1 + i\beta\tfrac{|k|}{k}\omega(k,\alpha))$$

where $\gamma \in \mathbb{R}$, $c \geq 0$, $\alpha \in (0, 2\rangle$ *and* $\beta \in \langle -1, 1\rangle$ *and function* ω *has form:*

$$\omega(k,\alpha) = \begin{cases} -tan(\pi\alpha/2) & for\ \alpha \neq 1 \\ (2/\pi)ln|k| & for\ \alpha = 1 \end{cases}$$

Remark. *Meaning of parameters* α, β, γ *and* c *can be easily understood. For* $\alpha = 2$ *we have normal distribution with the expected value* $\mu = \gamma$ *and variance* $\sigma^2 = 2c$ *regardless of parameter* β.
For $\alpha \in (0, 2)$ *is* α *a tail exponent or an index of stability.*

$$p_{\alpha,\beta}(x) \sim \tfrac{1}{|x|^{1+\alpha}} \quad for \quad |x| \to \infty$$

This can be easily shown. For simplicity we will presume $\gamma = \beta = 0$ *. Then*

$$p_\alpha(x) = \tfrac{1}{2\pi}\int_{-\infty}^{\infty} exp(-c|y|^\alpha)exp(-iyx)\mathrm{d}y = \tfrac{1}{2\pi}\int_{0}^{\infty} exp(-c|y|^\alpha)(exp(-iyx)+exp(iyx))\mathrm{d}y =$$

$\frac{1}{\pi}\Re\int_0^\infty exp(-c|y|^\alpha)exp(iyx)\mathrm{d}y = \frac{1}{\pi}\Re\int_0^\infty \sum_{n=o}^\infty \frac{(-c|y|^\alpha)^n}{n!}exp(iyx)\mathrm{d}y =$

We can use Lebesque's dominated convergence theorem now and change an order of summation and integration.

$= \frac{1}{\pi}\Re\sum_{n=o}^\infty \frac{(-c)^n}{n!}\int_0^\infty |y|^{\alpha n}exp(iyx)\mathrm{d}y = \frac{1}{\pi}\Re\sum_{n=o}^\infty \frac{(-c)^n}{n!}\frac{\Gamma(\alpha n+1)}{(-ix)^{\alpha n+1}}$

Finally, with a use of identity:

$$\Re((\pm i)^{\alpha n+1}) = -sin(\pi\alpha n/2)$$

we will get

$$p_\alpha(x) = -\frac{1}{\pi}\sum_{n=1}^\infty \frac{(-c)^n}{n!}\frac{\Gamma(\alpha n+1)}{x^{\alpha n+1}}sin(\pi\alpha n/2)$$

So only the first term in sum matters for $|x| \to \infty$ and we will get a correct asymptotic behaviour. Clearly variance diverges for $\forall\alpha \in (0,2)$ and even an expected value is infinite for $\alpha \in (0,1\rangle$. Values $\alpha > 2$ have no mathematical sense because inverse Fourier transform will give functions, which are not generally positive.

The parameter γ gives a peak position or an expected value, if it exists. β determines how asymmetric probability density is. Specially for $\gamma = \beta = 0$ is probability density even a function. On the other hand for $\beta = \pm 1$ is distribution very asymmetric and if at the same time $\alpha < 1$, one tail will vanish completely. The parameter c is scale factor that determines a width of a distribution.

We will show that parameters c and γ are really only responsible for scaling and a peak position, therefore a shape of distribution is completely given by the parameters α, β. We will show that by a simple replacement of $x-\gamma$ with $c^{1/\alpha}x$ in a formula for PDF we can eliminate γ from the formula and c will remain only as a multiplication constant.

$p_{\alpha,\beta}(x) = \mathcal{F}^{-1}[\varphi_{\alpha,\beta}](x) = \frac{1}{2\pi}\int_{-\infty}^\infty exp(-i(x-\gamma)y - c|y|^\alpha(1+i\beta sgn(y)\omega(y,\alpha))\mathrm{d}y =$

$\frac{1}{2\pi c^{1/\alpha}}\int_{-\infty}^\infty exp(-ixz - |z|^\alpha(1+i\beta sgn(z)\omega(c^{1/\alpha}z,\alpha))\mathrm{d}z$

Where we used substitution $z = c^{1/\alpha}y$ in integral and replaced $x-\gamma$ with $c^{1/\alpha}x$.

Now we will give concrete examples of stable distributions:

Normal distribution: *For $\alpha = 2$, $c = \sigma^2/2$, $\gamma = \mu$ and any $\beta \in \langle -1,1\rangle$ we have*

$$p_{2,0}(x) = N(\mu,\sigma^2)$$

Levy - Smirnov distribution: *If we set $\alpha = 1/2$, $\beta = 1$, we will obtain by inverse Fourier transform of characteristic function following PDF:*

$$p_{1/2,1}(x) = \sqrt{\frac{c}{2\pi}} \frac{exp(-\frac{c}{2}(x-\gamma)^{-1})}{(x-\gamma)^{3/2}}, \qquad x \geq \gamma$$

We will note that only stable distributions with $\beta \pm 1$ and $\alpha < 1$ are not defined on entire $\mathbb{R}$. Those distributions are defined on $\langle \gamma, \infty)$ respectively $(-\infty, \gamma\rangle$.

Cauchy distribution: *Fixing $\alpha = 1$ and $\beta = 0$ gives Cauchy distribution:*

$$p_{1,0}(x) = \frac{c}{\pi((x-\gamma)^2 + c^2)}$$

These are only stable distributions with PDFs that can be written in terms of elementary functions. However, other stable distributions with PDFs that can be written in a closed form using special functions exist .

We will summarize basic properties of stable distributions. We should mention first that a stability holds only for random variables with the same parameter α. So if we sum stable random variables with the same α parameter, we will get a stable distribution with the parameter α. However, if we sum stable distributions with different parameters α_i, we will not generally get a stable distribution.

We will consider independent stable random variables X, Y where $X \sim p_{\alpha,\beta_1,\gamma_1,c_1}(x)$ and $Y \sim p_{\alpha,\beta_2,\gamma_2,c_2}(x)$, then

$$X + Y \sim p_{\alpha,\beta,\gamma,c}(x)$$

where $c = c_1 + c_2$, $\gamma = \gamma_1 + \gamma_2$ and $\beta = \frac{c_1\beta_1 + c_2\beta_2}{c}$. Clearly these values are in right ranges, therefore $X+Y$ is a stable random variable. We used that characteristic function of sum of independent variables is a product of characteristic functions in calculation.

Stable distributions also fulfil something what is often called a scaling property. In case of normal distribution is well known that for independent random variables $X_i \sim N(0, \sigma_i^2)$ following relation holds $\sum_{i=1}^n a_i X_i \sim N(0, \sum_{i=1}^n a_i^2 \sigma_i^2)$. So for iid random variables $Y_i \sim N(\mu, \sigma^2)$

$$Y_n - \mu \overset{d}{=} n^{-1/2} \sum_{i=1}^n (Y_i - \mu) \sim N(0, \sigma^2)$$

where $\overset{d}{=}$ means an equality in distribution.

We will show that similar scaling relation holds for all stable distributions. Let us consider astable random variable $U' \sim p_{\alpha,\beta,\gamma,c}(x)$ first. We want distribution of random variable $U = aU'$ where $a > 0$. We know that $p_U(x) = \frac{1}{a} p_{U'}(\frac{x}{a})$ from theorem (1.1), we use property of Fourier transform

$$\mathcal{F}[f(ax)](z) = \frac{1}{|a|^n}\mathcal{F}[f(x)](z/a) \quad for \quad f : \mathbb{R}^n \to \mathbb{R}$$

By applying this to $p_U(x)$ we will get $\varphi_U(z) = \varphi_{U'}(az)$, so we can see that

$$U \sim p_{\alpha,\beta,a\gamma,a^\alpha c}(x)$$

As expected a stability of U holds and only scale parameters have changed. If we now consider iid stable random variables U_i' with parameter $\gamma = 0$ we get

$$\sum_{i=1}^n s_i U_i' \overset{d}{=} (\sum_{i=1}^n s_i^\alpha)^{1/\alpha} U_n'$$

If we consider more generally iid random variables $U_i \sim p_{\alpha,\beta,\gamma,c}(x)$, we will show that $V_i = U_i - \gamma \sim p_{\alpha,\beta,0,c}(x)$. Once again we can use theorem (1.1) and get $p_Y(x) = p_X(x + \gamma)$. We use a formula

$$\mathcal{F}[f(x + b)](z) = e^{-ibz}\mathcal{F}[f(x)](z)$$

and we get $\varphi_{V_i}(z) = e^{-i\gamma z}\varphi_{U_i}(z) = exp(-c|z|^\alpha(1 + i\beta\frac{z}{|z|}\omega(z,\alpha)))$. It means $V_i \sim p_{\alpha,\beta,0,c}(x)$ and we can write

$$U_n - \gamma \overset{d}{=} n^{-1/\alpha}\sum_{i=1}^n (U_i - \gamma) \tag{1.5}$$

This scaling relation for $\alpha = 2$ clearly coincides with scaling relation for normal distribution. We will get equality $p_{S_n}(x) = \frac{1}{n^{1/\alpha}}p_{V_n}(\frac{x}{n^{1/\alpha}})$ where $S_n = \sum_{i=1}^n V_i$ by reformulating (1.5) to equality of PDFs. This is typical sign of fractal behaviour, this connection will be further discussed in the chapter 3.

Now we will proceed to a generalized CLT where stable distributions play an important role.

1.3.2 Generalized central limit theorem

Following theorem shows an importance of stable distributions and gives them a prominent position between others heavy-tailed distributions.

Theorem 1.5. *(generalized CLT) Let X_i be a sequence of iid random variables, with asymptotic behaviour of PDF given by*

$$p_{X_i}(x) = C_\pm|x|^{-(1+\alpha)} \quad for \quad x \to \pm\infty \ , \ \alpha \in (0, 2)$$

Let us define the parameter $\beta = \frac{C_+ - C_-}{C_+ + C_-}$ then $\gamma \in \mathbb{R}$ exists such as

$$\frac{\sum_{i=1}^n (X_i - \gamma)}{n^{1/\alpha}} \xrightarrow{\mathcal{D}} X \sim p_{\alpha,\beta}(x)$$

Remark. *So in a combination with classical CLT, we can say that every properly normalized sequence of iid random variables converges to a stable distribution and the limiting distribution is normal distribution iff those variables have an finite variance.*

Assumptions of this theorem are mild but still restrictive, it is possible to weaken an assumption of identic distribution as in case of CLT but the same asymptotic behaviour of all variables in $\pm\infty$ is still required. To go beyond an independence of variables is even more complicated. It requires substantial generalization to so called q-stable distributions, which we will mention in the last chapter. Also this theorem does not work for variables with a different tail behaviour in $\pm\infty$, not even for iid variables. This assumption seems to be the key one.

We can from this theoretical discussion sense a problem that we have to face when describing complicating and complex systems like financial markets. Assumptions of any limit theorem will not be completely fulfilled, at least not always. Limit theorems give us classes of distributions we should expect to come across when describing a complicating systems effected by many small, random influences. However, many unpredictable events can occur in reality. They can for example create strong correlations between individual influences or one influence can overshadow a compound effect of the rest of them. Limit theorems that use to work well will fail then and give completely wrong predictions. Despite these limitations, limit theorems are still one of the most useful tools we have for describing such systems.

A concrete example of a situation I described above is well known from mathematical finance. Most of the traders believed that CLT is an omnipotent principle and they do not need to bother with any models based on something else than normal distribution for long time. Misguidedness of this assumption was theoretically shown by Mandelbrot. He showed that while during calm times financial markets are well described by a normal distribution, during crises normal distribution gives deeply flawed predictions. But it took more than 20 years and bankruptcy of hedge fund led by Black and Scholes before the main focus of mathematical finance shifted from a normal distribution to alternatives like stable distributions.

1.3.3 Infinitely divisible distributions

Definition 1.10. *Random variable X is an infinitely divisible iff for every $n \in \mathbb{N}$ there exists iid random variable $X_{n,1}, ..., X_{n,n}$ so that*

$$X \stackrel{d}{=} X_{n,1} + ... + X_{n,n}$$

Remark. *We can easily see necessary and sufficient condition for a characteristic function of an infinitely divisible random variable X from this definition*

$$\varphi_X(x) = (\varphi_n(x))^n \quad \forall n \in \mathbb{N}$$

where $\varphi_n(x)$ is a characteristic function of some random variable, that necessarily means φ_n is continuous and $\varphi_n(0) = 1$.

It is worth mentioning that sometimes a definition of infinitely divisible distributions is more general and we require random variables $X_{n,1}, ..., X_{n,n}$ to be only independent.

Every stable distribution is also infinitely divisible, it is easy to see because for any stable distribution with parameters α, β, γ, c we can consider n independent random stable variables with parameters $\alpha, \beta, \gamma/n, c/n$.

There are many infinitely divisible distributions, both with a finite and an infinite variance. For example, Gamma distribution, Student's t-distribution or Poisson distribution are infinitely divisible. But for example binomial or uniform distribution are not infinitely divisible. We can show the last statement easily because a characteristic function of uniformly distributed random variable Y

$$p_Y(x) = \begin{cases} l/2 & \text{for } |x| < l \\ 0 & \text{for } |x| > l \end{cases}$$

has form

$$\varphi_Y(z) = \frac{sin(zl)}{zl}$$

and therefore Y is not infinitely divisible because nth root does not generally exists.

An interesting property of infinitely divisible distributions is a completeness, i.e. the limit of weakly-convergent (convergent in distribution) sequence of infinitely-divisible distributions is again infinitely divisible.

The main reason why we mention infinitely divisible distributions is their role in another generalization of CLT. Let us consider a triangular array of independent random variables $X_{n,k}$ where $n > k$, fulfilling the condition

$$\lim_{n\to\infty} \max_{1\le k\le n} P(|X_{n,k}| > \varepsilon) = 0 \quad \forall \varepsilon > 0 \tag{1.6}$$

This condition is called a uniform asymptotic negligibility. It implies, that for big n, the influence of individual random variables is negligible compared to the compound influence of the rest of random variables in line. The uniform asymptotic negligibility can be easily arranged by a proper scaling like in cases of limit theorems.

We can state limit theorem for infinitely divisible distributions now.

Theorem 1.6. *Let us consider a triangular array of independent random variables fulfilling uniform asymptotic negligibility condition, then*

$$S_n = X_{n,1} + ... + X_{n,n} \xrightarrow{\mathcal{D}} Y$$

and Y is infinitely divisible random variable.

Other theorems similar to theorems about stable distributions can be proven for infinitely divisible distributions. Generally, these theorems are quite complicated and often their applicability is questionable. Though infinitely divisible distributions are used in mathematical finance rarely, they offer a way how generalize limit theorems for sequences of distributions with different asymptotic behaviour.

The assumption of identic distribution is not fulfilled in finance, that is actually quite logical. Big investors have bigger influence on the market than smaller ones and also they often have a different approach to trading. Anyway, it turns out that individual PDFs are similar enough so stable distributions give most of the time good predictions.

Chapter 2

Stochastic calculus

In this chapter, we will introduce probably the most important tool to describe systems such as financial markets - stochastic processes. We will focus on their properties. We will introduce stochastic integral and discuss some important classes of stochastic processes in details. The proves of most of the theorems stated here can be found in [6] or [9].

2.1 Stochastic processes

Definition 2.1. *Let $T \subset \mathbb{R}$ and $\forall t \in T$ X_t is random variable to $\mathbb{R}(\mathbb{R}^n)$ then we will call $(X_t, t \in T)$ a stochastic process.*

Definition 2.2. *Function $X_t(\omega)$ for some fixed $\omega \in \Omega$ is called a sample path of the stochastic process X.*

Remark. *For $T = \mathbb{N}$ (or some other countable subset of $\mathbb{R}$) we will call X_n discrete-time stochastic process, for $T = (a, b)$ we will call X_t continuous-time stochastic process.*

It is obvious that stochastic processes are a better option for describing financial markets then just random variables. The parameter t is interpreted as real time very often, so stochastic process is a random variable changing in time.

We can write that $\forall t \in T \quad X_t \sim p(x, t)$, where $p(x, t)$ is PDF of stochastic process X in time t or, in other words, a density of sample paths in time t in neighbourhood of x. In the last section of this chapter, we will describe an approach where we are only interested in time development of $p(x, t)$. Now, we will focus on an approach, which works with concrete sample paths as well.

Definition 2.3. *For every stochastic process X_t, we will define finite-dimension distribution (fidis) as distribution of random vector $(X_{t_1}, ..., X_{t_n})$ for any $n \in \mathbb{N}$ and $\forall (t_1, ..., t_n) \in T$.*

Remark. *We will call stochastic process X Gaussian iff all fidises are Gaussian. From an elementary probability theory we know that a distribution of n-dimensional*

Gaussian random vector is fully given by correlation matrix $\Sigma \in \mathbb{R}^{n,n}$ *and by expected value* $\mu \in \mathbb{R}^n$.

From now on we will use marking $C_X(t,s) = Cov(X_t, X_s)$ *and* $\mu_X(t) = EX_t$ *where* X *is any stochastic process.*

Definition 2.4. *Stochastic process* X *is said to be strictly stationary* $\iff$ $\forall (t_1, ..., t_n), \forall h > 0$

$$(X_{t_1}, ..., X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, ..., X_{t_n+h})$$

where $t_1, ..., t_n, t_1 + h, ..., t_n + h \in T$

Remark. *For example in case of Gaussian process we only need to check conditions* $\mu_X(t+h) = \mu_X(t)$ *and* $C_X(t,s) = C_X(t+h, s+h)$ *in order to verify stationarity.*

In reality this property is too strong and we encounter strictly stationary processes in applications rarely. Following properties are on the other hand very common.

Definition 2.5. *Stochastic process* X *has stationary increments if for every* $t, s \in T$

$$X_t - X_s \stackrel{d}{=} X_{t+h} - X_{s+h}$$

for such h *that* $t + h, s + h \in T$.

In other words, increments of such processes depend only on a time interval in between.

Definition 2.6. *Stochastic process* X *has independent increments if for* $\forall (t_1, ..., t_n)$ *where* $t_1 < t_2 < ... < t_n$ *random variables*

$$X_{t_2} - X_{t_1}, .., X_{t_n} - X_{t_{n-1}}$$

are independent.

Remark. *So for processes with independent and stationary increments if* $t_i = t_0 + i\Delta t$, *then random variables* $X_{t_2} - X_{t_1}, .., X_{t_n} - X_{t_{n-1}}$ *are iid.*
Now we will give some examples of stochastic processes.

Poisson process: *Process* $(X_t, t \geq 0)$ *is Poisson process with parameter* $\lambda > 0$ *if*
1) $X_0 = 0$ *a.s.*
2) X *has independent and stationary increments*
3) $X_t \sim Poi(\lambda t)$ *that means* $P(X_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$

A.s. means almost surely i.e. with probability 1. From this definition also immediately follows that

$$X_t - X_s \stackrel{d}{=} X_{t-s} - X_0 = X_{t-s} \sim Poi(\lambda(t-s))$$

We should mention a role of Poisson process in theorem known as law of rare events. Let X_t be a number of occurrences of some event up to time t, let X_t be a process with stationary increments. Then if assumption of rareness is fulfilled $X_t \sim Poi(\lambda t)$. Mathematically speaking rareness of event means following conditions.
a) $P(X_{t+h} - X_t = 1) \approx \lambda h + o(h)$
b) $P(X_{t+h} - X_t > 1) \approx o(h)$
for $h \to 0$
We will discuss the most important example of stochastic process - Brownian motion further.

2.1.1 Brownian motion

Definition 2.7. *Stochastic process $(B_t, t \geq 0)$ is called (standard) Brownian motion if*
1) $B_0 = 0$ *a.s.*
2) B_t has independent and stationary increments
3) $B_t \sim N(0, t)$
4) sample paths $X_t(\omega)$ are continuous

Remark. *Brownian motion (BM) is named after biologist Robert Brown, whose research dates back to 1820s but a firm mathematical background to Brownian motion was given 100 years later by mathematician Norbert Wiener. It is why Brownian motion is also often called Wiener process.*

We can also consider Brownian motion in $\mathbb{R}^n$ i.e. random vector $(B_t^{(1)}, ..., B_t^{(n)})$ where all $B_t^{(i)}$ are Brownian motions in $\mathbb{R}$ and for every t, $(B_t^{(1)}, ..., B_t^{(n)})$ has normal n-dimensional distribution i.e every linear combination $\sum_{i=1}^n \alpha_i B_t^{(i)}$ is normally distributed. This is clearly always true for $B_t^{(i)}$ independent.

Properties of Brownian motion:
Brownian motion is clearly Gaussian process. We will calculate expected value and covariance of Brownian motion first, $\forall t, s \geq 0$ following relations hold.
i) $\mu_B(t) = EB_t = 0$
ii) $Var_B(t) = EB_t^2 = t$
iii) $C_B(t, s) = min(t, s)$
Because $C_B(t,s) = E(B_t B_s) = E((B_t + B_s - B_s)B_s) = E((B_t - B_s)B_s) + EB_s^2 \stackrel{t>s}{=} E(B_t - B_s)EB_s + s = s$
Now we will focus on path properties of BM.

Definition 2.8. *Stochastic process $(X_t, t \geq 0)$ is H - self similar for some $H > 0$ if all its fidis satisfy condition*

$$(T^H X_{t_1}, ..., T^H X_{t_n}) \stackrel{d}{=} (X_{Tt_1}, ..., X_{Tt_n})$$

for every $T > 0$.

Remark. *Roughly speaking, self-similarity means that sample paths look similarly on every scale. Sample paths are only similar on every scale but not identical, because equality in definition is only in distribution. This is also a clear sign of fractal-like behaviour.*

A self similarity has also important consequences that can be formulated by following theorems.

Theorem 2.1. *Let $(X_t, t \in T)$ be H - self similar stochastic process with stationary increments for some $H \in (0,1)$ then $\forall t_0 \in T$*

$$\limsup_{t \to t_0} \frac{|X_t - X_{t_0}|}{|t - t_0|} = \infty$$

i.e. such processes have nowhere differentiable sample paths with probability 1.

Remark. *This theorem is directly applicable to Brownian motion because from properties of normal distribution we can easily see*

$$(T^{1/2} B_{t_1}, ..., T^{1/2} B_{t_n}) \stackrel{d}{=} (B_{Tt_1}, ..., B_{Tt_n})$$

for every $T > 0$, $\forall n \in \mathbb{N}$ and for any partition $(t_1, ..., t_n)$. Therefore we can state following properties of Brownian motion:

1) Brownian motion is $\frac{1}{2}$-self similar
2) Brownian motion has continuous but nowhere differentiable sample paths

Another important property of Brownian motion is that it has an unbounded variation.

Definition 2.9. *Function f on a finite interval $\langle 0, T \rangle$ has bounded p-variation, if*

$$\sup_\tau \sum_{i=1}^n |f(t_i) - f(t_{i-1})|^p < \infty$$

where supremum is over every partition of $\langle 0, T \rangle$, $0 = t_0 < ... < t_n = T$ for any $n \in \mathbb{N}$. In case of $p = 1$ we will just say that f has bounded variation.

Theorem 2.2. *Brownian motion has unbounded variation on any finite interval $\langle 0, T \rangle$ with probability 1, i.e.*

$$\sup_\tau \sum_{i=1}^n |B_{t_i}(\omega) - B_{t_{i-1}}(\omega)| = \infty \ \ a.s.$$

Remark. *This theorem can be further generalized: Brownian motion has unbounded variation for $p \leq 2$ and bounded variation for all $p > 2$.*

Unbounded variation and nowhere differentiability are main reasons why classical differential and integral calculus fails when applied to these paths or to stochastic processes in general. In other section of this chapter, we will introduce integral to work with when integrating along Brownian sample path.

We will introduce some examples of stochastic processes derived from Brownian motion further.

Brownian motion with linear drift

Stochastic process

$$X_t = \mu t + \sigma B_t \tag{2.1}$$

where $\sigma > 0, \mu \in R$ and B_t is standard Brownian motion is called Brownian motion with linear drift.
It is clearly Gaussian process with expected value $\mu_X(t) = \mu t$ and covariance $C_X(t,s) = \sigma^2 min(t,s)$.

This process was used in 1900 by L.Bachelier to model a development of financial prices but his approach had many flaws. For example, this process is not always positive, which is a natural requirement, when describing prices of assets. That is one of the reasons why geometric Brownian motion play much bigger role in finance then Brownian motion with a drift.

Geometric Brownian motion

Process

$$X_t = \exp(\mu t + \sigma B_t) \tag{2.2}$$

is called geometric Brownian motion. We will calculate its expected value and covariance.

$$\mu_X(t) = e^{\mu t} E e^{\sigma B_t} = e^{\mu t} E e^{\sigma t^{1/2} B_1} = e^{(\mu + \frac{1}{2}\sigma^2)t}$$

We used self similarity of B_t and relation $Ee^{aZ} = e^{a^2/2}$ for $Z \sim N(0,1)$ in calculations.
Similarly, we can calculate

$$C_X(t,s) = e^{(\mu+\frac{1}{2}\sigma^2)(t+s)}(e^{\sigma^2 min(t,s)} - 1)$$

An importance of geometric Brownian motion in finance will be discussed in the chapter 4, in fact, entire Black-Scholes option pricing formula stands on an assumption that prices of assets on financial markets follow geometric Brownian motion.

2.1.2 Conditional expectation

In this section, we will define a conditional expectation given by σ-algebra, which will be needed for a definition of stochastic integral. This approach is alternative to approach based on conditional probability densities as defined in the chapter 1. The other approach will be discussed in the section 4 of this chapter.

First, we remind a notion of σ-algebra.

Definition 2.10. *Let $\mathcal{F}$ be a set of subsets of Ω then $\mathcal{F}$ is σ-algebra if following conditions are fulfilled*

1) $\emptyset \in \mathcal{F}$

2) $(A \in \mathcal{F}) \Rightarrow (A^C \in \mathcal{F})$
3) $(\forall k \in \mathbb{N}, A_k \in \mathcal{F}) \Rightarrow (\bigcup_{k\in\mathbb{N}} A_k \in \mathcal{F})$

where A^C *denotes complement of* A *in* Ω.

Remark. *Properties of* σ*-algebras are well known, so we will not go into details and we will state just basic properties of* σ*-algebras following immediately from the definition.*
a) $\Omega \in \mathcal{F}$
b) $(\forall k \in \mathbb{N}, A_k \in \mathcal{F}) \Rightarrow (\bigcap_{k\in\mathbb{N}} A_k \in \mathcal{F})$

Definition 2.11. *Let* A *be any set of subsets of* Ω *then* $\sigma(A)$ *denotes smallest* σ*-algebra containing* A.

Definition 2.12. *Let* $Y = (Y_1, ..., Y_n)$ *be a random vector to* $\mathbb{R}^n$ *then we define* σ*-algebra generated by* Y *as smallest* σ*-algebra containing all sets of form*

$$\{\omega \in \Omega, \quad a_i < Y_i \leq b_i, \quad i = 1, ..., n\}$$
$$-\infty < a_i < b_i < \infty$$

we denote this σ*-algebra by* $\sigma(Y)$.

Remark. *In this definition, we can change generating set of subsets* $(a_i, b_i\rangle$ *for any other subset that generates Borel* σ*-algebra. For example, all open or all close sets in* $\mathbb{R}(\mathbb{R}^n)$. *Alternatively, we can take inverse images of all Borel sets in* $\mathbb{R}(\mathbb{R}^n)$ *and we will get entire* $\sigma(Y)$.

The importance of $\sigma(Y)$ *lies in fact that* $\sigma(Y)$ *contains essential information about variable* Y. *It gives us all information we need about a structure of* Y. *We say that* Y *contains or carries information* $\sigma(Y)$.

We will define σ*-algebra generated by stochastic process, it can be defined the most generally for stochastic process* $(Y_t, t \in T)$ *as the smallest* σ*-algebra containing sets in form* $\{\omega \in \Omega, (Y_t(\omega), t \in T) \in C\}$ *where* C *is "suitable" set of functions on* T. *To make this definition more precise we would need to go deep into functional analysis, which is beyond this thesis.*

Instead of that we will consider for example Brownian motion $B = (B_s, s \leq t)$, *then* σ*-algebra generated by* B *will be generated by all sets in form*

$$\{\omega \in \Omega, (B_{t_1}(\omega), ..., B_{t_n}(\omega)) \in \mathcal{B}_n\} \quad \text{for any } n \in \mathbb{N} \quad \text{and} \quad t_i \in \langle 0, t\rangle$$

where $\mathcal{B}^n$ *denotes Borel sets in* $\mathbb{R}^n$. *We will use a symbol* $\mathcal{F}_t = \sigma(B) = \sigma((B_s, s \leq t))$ *for information contained by Brownian motion .*

Let us consider a random variable Y *and* σ*-algebra* $\mathcal{F}$ *then if* $\sigma(Y) \subset \mathcal{F}$, *we will say that all information about* Y *is contained in* $\mathcal{F}$. *If* $\mathcal{F} = \sigma(X)$ *for some random variable* X *we will say that* X *contains more information than* Y.

We can see from the definition that for any Borel measurable function $f : \mathbb{R}^n \to \mathbb{R}^n$ *and any random vector* Y *to* $\mathbb{R}^n$

$$\sigma(f(Y)) \subset \sigma(Y)$$

This is actually quite logical. Let us consider function $f(x) = x^2$ then from knowledge of x we can easily construct $f(x)$ but from knowledge of $f(x)$ we cannot construct x because we cannot determine a right sign of x.

We can finally define conditional expectation given by σ-algebra.

Definition 2.13. *A random variable Z is called conditional expectation of X given by σ-algebra $\mathcal{F}$ (we write $Z = E(X|\mathcal{F})$) if*
1) $\sigma(Z) \subset \mathcal{F}$
2) $E(XI_A) = E(ZI_A) \quad \forall A \in \mathcal{F}$
Where I_A is an indicator of event A i.e. $I_A(\omega) = 1$ for $\omega \in A$ and $I_A(\omega) = 0$ else.
Similarly for a random variable Y we define $E(X|Y) = E(X|\sigma(Y))$.

Remark. *In other words, Z is a coarser version of X, which we can construct having only information given by $\mathcal{F}$. So if we have all information contained in X, we can reconstruct X completely. Mathematically we have*

$$\sigma(X) \subset \mathcal{F} \Rightarrow E(X|\mathcal{F}) = X$$

which is obvious from the definition.

Theorem 2.3. *Let X be a random variable and $E|X| < \infty$, then for any σ-algebra $\mathcal{F}$, $E(X|\mathcal{F})$ exists and is given uniquely.*

Remark. *Because of this theorem, we will usually consider only random variables with a finite first moment, when dealing with a conditional expectation.*

We will mention that a conditional expectation can be easily calculated, when σ-algebra is generated by a discrete random variable.
Let Y be a discrete random variable, let

$$A_i = \{\omega \in \Omega, Y(\omega) = y_i\} \quad i \in I \subset \mathbb{N}$$

and let us assume $P(A_i) > 0$ and $\bigcup_i A_i = \Omega$ then for any random variable X with a finite first moment

$$E(X|Y)(\omega) = E(X|A_i) \quad \forall \omega \in A_i$$

where $E(X|A_i)$ is defined as

$$E(X|A_i) = \frac{E(XI_A)}{P(A)}$$

Now we will show that these relations are compatible with the original definition of conditional expectation for an arbitrary random variable. So we want to prove

$$Z(\omega) = E(X|A_i) \stackrel{?}{=} E(X|\sigma(Y)) \quad \forall \omega \in Ai$$

Clearly $\sigma(Z) \subset \sigma(Y)$ *so we only need to prove*

$$E(XI_A) = E(ZI_A) \quad \forall A \in \sigma(Y)$$

Because Y *is a discrete variable* $\sigma(Y) = \sigma(\{A_i, i \in I\})$ *where* A_i *are disjoint so any* $A \in \sigma(Y)$ *can be written as* $A = \bigcup_{i \in I'} A_i$*, we can make calculations*

$E(XI_A) = E(X \sum_{i \in I'} I_{A_i}) = \sum_{i \in I'} E(XI_{A_i})$

$E(ZI_A) = \sum_{i \in I'} E(X|A_i)P(A_i) = \sum_{i \in I'} E(XI_{A_i})$

In the calculation of $E(XI_A)$*, we used the assumption* $E|X| < \infty$ *to change an order of summation and integration. So now we know, how to calculate a conditional expectation for* σ*-algebras generated by discrete random variables, but we also need some tools to work with a conditional expectation given by arbitrary* σ*-algebra.*

Properties of conditional expectation:

1. $E(\quad|\mathcal{F})$ is linear
2. $EX = E(E(X|\mathcal{F}))$
3. Let X and $\mathcal{F}$ be independent i.e. $\forall A \in \mathcal{F}$, I_A and X are independent then $EX = E(X|\mathcal{F})$
4. $\sigma(X) \subset \mathcal{F}$, G random variable then $E(XG|\mathcal{F}) = XE(G|\mathcal{F})$
5. Let $\mathcal{F} \subset F'$
   (a) $E(X|\mathcal{F}) = E(E(X|\mathcal{F}')|\mathcal{F})$
   (b) $E(X|\mathcal{F}) = E(E(X|\mathcal{F})|\mathcal{F}')$

Property 1. is obvious, property 2. follows immediately from the definition (2.13) if we select $A = \Omega$.
Before proving the property 3., we will mention that random variables X and Y are independent $\Leftrightarrow$ X and $\sigma(Y)$ are independent. A prove of property 3. follows from calculation

$$E(XI_A) = (EX)P(A) = E((EX)I_A)$$

We will not prove the property 4. rigorously because it is obvious that we can treat X as a constant when we have all information that X contains.
The case (b) of property 5. follows from

$$\sigma(E(X|\mathcal{F})) \subset \mathcal{F} \subset \mathcal{F}'$$

and from the property 4.
For prove of the case (a) we will use that $\forall A \in \mathcal{F}$

$$E(E(E(X|\mathcal{F}')|\mathcal{F})I_A) = E(E(E(XI_A|\mathcal{F}')|\mathcal{F})) = E(XI_A)$$

where we use that $\sigma(I_A) \subset \mathcal{F} \subset \mathcal{F}'$ first. We can treat I_A as a constant then and apply the property 2. twice. The second condition

$$\sigma(E(E(X|\mathcal{F}')|\mathcal{F})) \subset \mathcal{F}$$

is obvious so we have proved the case (a).

We will calculate conditional expectation for Brownian motion now. Let us consider $\mathcal{F}_s = \sigma(B_x, x \leq s)$ then

$$E(B_t|\mathcal{F}_s) \overset{s\leq t}{=} E(B_t - Bs + Bs|\mathcal{F}_s) = E(B_t - B_s|\mathcal{F}_s) + E(B_s|\mathcal{F}_s) = E(B_t - B_s) + B_s = B_s$$

and because evidently $E(B_t|\mathcal{F}_s) \overset{s\geq t}{=} B_t$ we get

$$E(B_t|\mathcal{F}_s) = B_{min(t,s)} \tag{2.3}$$

Similarly, we can derive

$$E(B_t^2 - t, \mathcal{F}_s) = B_{min(t,s)}^2 - min(t,s) \tag{2.4}$$

In next the section, we will examine a general class of processes fulfilling this condition.

Definition 2.14. *Let us the set of random variables Z, which fulfils conditions*

1. $EZ^2 < \infty$
2. $\sigma(Z) \subset \mathcal{F}$

denote by $L^2(\mathcal{F})$.

Remark. *Clearly $L^2(\mathcal{F})$ is a vector space, a closed subspace of space L^2 of all random variables with a finite variance . So for any random variable $X, EX^2 < \infty$ we can calculate a projection to $L^2(\mathcal{F})$. It can be shown that a projection of such X to $L^2(\mathcal{F})$ is $E(X|\mathcal{F})$. In other words*

$$E|X - E(X|\mathcal{F})|^2 = \min_{Z\in L^2(\mathcal{F})} E(X - Z)^2 \tag{2.5}$$

2.1.3 Martingales

Definition 2.15. *A collection of σ-algebras $(\mathcal{F}_t, t \geq 0)$ where for $s \leq t$ is $\mathcal{F}_s \subset \mathcal{F}_t$ is called a filtration. The filtration can be also discrete $(\mathcal{F}_n, n = 1, 2...)$ fulfilling $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.*

Remark. *Thus the filtration is an increasing stream of information. This definition is natural for applications in finance, where as time goes, we have more and more information. However, thermodynamic system on its own is heading towards equilibrium, where entropy reaches its maximum and information we have about system is minimal. Therefore for thermodynamics, it would be more logical to define the filtration as a decreasing stream of information.*

Stochastic processes and filtrations are closely connected.

Definition 2.16. *A stochastic process $(Y_t, t \geq 0)$ is adopted to the filtration $(\mathcal{F}_t, t \geq 0)$ if for $(\forall t)$ $(\sigma(Y_t) \subset \mathcal{F}_t)$. Analogically, a discrete time process $(Y_n, n = 1, 2...)$ is adopted to the filtration $(\mathcal{F}_n, n = 1, 2...)$ if for $(\forall n)$ $(\sigma(Y_n) \subset \mathcal{F}_n)$.*
Clearly, every stochastic process Y is adopted to a natural filtration given by $\mathcal{F}_t = \sigma(Y_s, s \leq t)$.

Remark. *We will often use a natural filtration of Brownian motion given by $\mathcal{F}_t = \sigma(B_s, s \leq t)$. We will say that processes adopted to natural filtration of Brownian motion are adopted to Brownian motion. An example of such processes are*

$$X_t = B_t^n + f(t) \quad or \quad Z_t = \max_{0 \leq s \leq t} B_s$$

where $f(t)$ is any deterministic function.

For example, a process $X_t = B_t^2$ has smaller natural filtration than Brownian motion but $Y_t = B_t^3$ has the same natural filtration. It is so because knowledge of value of Y_t gives us value of B_t but same is not true about X_t.

An example of process not adapted to Brownian motion is

$$Z_t = B_{t+1}$$

Definition 2.17. *Process $X = (X_t, t \geq 0)$ is called martingale with respect to filtration $(\mathcal{F}_t, t \geq 0)$ if*

1. $E|X_t| < \infty \quad \forall t \geq 0$
2. *X is adopted to $\mathcal{F}_t$*
3. $E(X_t|\mathcal{F}_s) = X_s \quad \forall s \in \langle 0, t \rangle$

For a discrete time process is a definition analogical, the third assumption has the form $E(X_{n+k}|\mathcal{F}_n) = X_n \quad \forall n, k \in \mathbb{N}$.

Remark. *The third assumption for discrete time martingale may be weaken to*

$$E(X_{n+1}|\mathcal{F}_n) = X_n \quad \forall n \in \mathbb{N} \tag{2.6}$$

because
$E(X_{n+1}|\mathcal{F}_n) = E(E(X_{n+2}|\mathcal{F}_{n+1})|\mathcal{F}_n) = E(E(X_{n+2}|\mathcal{F}_n)|\mathcal{F}_{n+1}) = E(X_{n+2}|\mathcal{F}_n) = E(X_{n+k}|\mathcal{F}_n)$
Also for a discrete time process X_n, the third condition can be rewritten to the form

$$E(Y_{n+1}|\mathcal{F}_n) = 0$$

where $Y_{n+1} = X_{n+1} - X_n$. Such Y_n is called a martingale difference sequence with respect to filtration $\mathcal{F}_n$.

Another important property of martingales follows from a calculation

$$EX_s = E(E(X_t|\mathcal{F}_s)) = EX_t \quad \forall s, t \geq 0 \tag{2.7}$$

So martingale processes have a constant expected value. We should mention that this condition is necessary but not sufficient. There exist processes with constant expected value which are not martingale (for example B_t^3).

From this condition, we can immediately say that the process B_t^2 is not martingale because $E(B_t^2) = t$. Also we already know from calculations made in the section (3.1.2) that $B_t, B_t^2 - t$ are martingale processes with respect to natural filtration of Brownian motion.

Let us consider a martingale difference sequence $Y = (Y_n, n = 1, 2...)$ with respect to filtration $(\mathcal{F}_n, n = 1, 2...)$ and a sequence $C = (C_n, n = 1, 2...)$ that fulfils the condition

$$\sigma(C_n) \subset \mathcal{F}_{n-1}$$

We will call this property predictability with respect to $\mathcal{F}_n$. Now we will define the process X by

$$X_n = \textstyle\sum_{i=1}^n C_i Y_i \quad \textit{and} \quad X_0 = 0$$

Such process X is called a martingale transform of Y by C. We will write $X = Y \cdot C$, further if the processes C_n, Y_n have finite variance X is a martingale with respect to $\mathcal{F}_n$.
We will prove that X fulfils all three defining properties of a martingale.

$E|X_n| \leq \sum_{i=1}^n |C_i Y_i| \leq \sum_{i=1}^n (EC_i^2 EY_i^2)^{1/2} < \infty$

Here we used Schwarz-Cauchy inequality.
Second condition $\sigma(X_n) \subset \mathcal{F}_n$ is obvious. Third condition follows from

$$E(X_n - X_{n-1}|\mathcal{F}_{n-1}) = E(C_nY_n|\mathcal{F}_{n-1}) = C_nE(Y_n|\mathcal{F}_{n-1}) = 0$$

here we used predictability of C.

Now let us consider Brownian motion $B = (B_s, s \leq t)$ *and partition* $0 = t_0 < ... < t_n = t$. *We will define martingale difference by*

$$\Delta_i B = B_{t_i} - B_{t_{i-1}} \quad and \quad \Delta_0 B = 0$$

with respect to filtration $\mathcal{F}_i = \sigma(B_{t_j}, j = 1, ..., i)$ *and* $\mathcal{F}_0 = \{\emptyset, \Omega\}$. *If we define* $\widetilde{B}_{t_i} = B_{t_{i-1}}$, *then martingale transform of* $\widetilde{B}$ *given by* $\Delta_i B$ *has form*

$$(\widetilde{B} \cdot \Delta_i B)_k = \sum_{i=1}^{k} B_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})$$

We know that this transformation is martingale from calculations we made in general case. Importance of this sum lies in fact that it is actually discrete form of Ito stochastic integral.

Often are martingales interpreted as fair games. That can be best seen from relation

$$E(X_t - X_s|\mathcal{F}_s) = 0 \quad \forall t \geq s$$

The best prediction of our profits in any time interval is 0 with no regard to previous history.

2.2 Stochastic integral

We will discuss first, what class of functions we want to integrate and why a classical approach fails. We want to be able to calculate integrals in a form

$$\int_0^T f(t, B_t)\mathrm{d}B_t(\omega).$$

where f is a "reasonable" function. It is well known from the theory of Riemann integral that for Riemann integrable functions

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(y_i)(t_i - t_{i-1}) = \int_0^T f(t)\mathrm{d}t$$

where $0 = t_0 < ... < t_n = T$ is partition of $\langle 0, T\rangle$, $y_i \in \langle t_{i-1}, t_i\rangle$ and $\lim_{n\to\infty} \max_i(t_i - t_{i-1}) = 0$. So the left side does not depend on a choice of partition and y_i.

We are facing a problem now, how does it work for integral in a form

$$\int_0^T f(t)\mathrm{d}g(t) \stackrel{?}{=} \lim_{n\to\infty} \sum_{i=1}^{n} f(y_i)(g(t_i) - g(t_{i-1}))$$

It is not easy to determine right conditions on functions f, g, so that a limit of Riemann sum would exists and would not depend on a choice of partition and y_i. Without going into details, we will state the following result.

Theorem 2.4. *Integral* $\int_0^T f(t)\mathrm{d}g(t)$ *exists in Riemann sense, if following conditions are fulfilled*

1. f *and* g *do not have discontinuity at the same point* $t \in \langle 0, T\rangle$
2. *There exist* $p, q > 0$ *that* f *has bounded* p*-variation and* g *has bounded* q*-variation and* $\frac{1}{p} + \frac{1}{q} > 1$

Remark. *This theorem gives sufficient conditions for existence of Riemann integral, but these conditions are also very close to necessity.*

So for Brownian motion, we do not have a problem with the first condition because Brownian motion is continuous. The second condition implies that we can only integrate deterministic functions with bounded q*-variation where* $q < 2$ *respectively random variables fulfilling the same condition almost surely. We can easily calculate that a differentiable function* f *with bounded derivation* $|f'(t)| < K\ \forall t \in \langle 0, T\rangle$ *has a bounded variation because*

$$\sup_{\tau} \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})| \leq K \sum_{i=1}^{n} |t_i - t_{i-1}| = KT$$

So we can integrate functions like $e^t, \sin t$ *in Riemann sense but we cannot even integrate all continuous functions. For example integral*

$$\int_0^T B_t(\omega)\mathrm{d}B_t(\omega)$$

does not exist in Riemann sense so need for a different stochastic integral is clear.

2.2.1 Ito integral

It is beyond this thesis to build stochastic integral mathematically rigorously. A construction of stochastic integral reminds construction of Daniell integral but is much more complicated. We will not go into details here and will focus mainly on properties of such integral. For more thorough description of this construction see Appendix C or the completely rigorous approach can be found in [9].

Let us start with the concrete example. We try to calculate $\int_0^T B_t(\omega)\mathrm{d}B_t(\omega)$, as always $B = (B_t, t \geq 0)$ is standard Brownian motion. Let us consider a sum

$$S_n = \sum_{i=1}^{n} B_{t_{i-1}} \Delta_i B \tag{2.8}$$

where $0 = t_0 < ... < t_n = T$ and $\Delta_i B = B_{t_i} - B_{t_{i-1}}$, this is actually Riemann sum for integral $\int_0^T B_t(\omega)\mathrm{d}B_t(\omega)$ but with the concrete choice of $y_i = t_{i-1}$. With use of calculations

$$S_n = \tfrac{1}{2}B_T^2 - \tfrac{1}{2}\sum_{i=1}^{n}(\Delta_i B)^2 := \tfrac{1}{2}B_T^2 - \tfrac{1}{2}Q_n(T)$$

and

$$E(\Delta_i B \Delta_j B) = \delta_{ij} Var(\Delta_i B) = \delta_{ij} \Delta_i$$

where we defined $\Delta_i = t_i - t_{i-1}$. Now we can calculate

$$EQ_n(T) = \sum_{i=1}^{n} E(\Delta_i B)^2 = T$$

and similarly

$$Var(Q_n(T)) = \sum_{i=1}^{n} Var(\Delta_i B)^2 = \sum_{i=1}^{n} (E(\Delta_i B)^4) - \Delta_i^2) = 2 \sum_{i=1}^{n} \Delta_i^2.$$

in calculations we used self similarity of B_t and fact that $EB_1^4 = 3$.
If we send $\underset{i}{max}(\Delta_i) \to 0$ we will get

$$Var(Q_n(T)) \leq 2 \ \underset{i}{max}(\Delta_i) T \to 0$$

and because from the definition $Var(Q_n(T)) = E(Q_n(T) - T)^2$ we have

$$Q_n(T) \xrightarrow{\mathcal{L}_2} T$$

For a definition of $\xrightarrow{\mathcal{L}_2}$ and other types of convergence see Appendix B.
If we presume $\mathcal{L}_2$ convergence of S_n, we will get

$$\int_0^T B_t(\omega) \mathrm{d}B_t(\omega) = \frac{1}{2}(B_T^2 - T) \tag{2.9}$$

So we can see for a stochastic integral, that we have to

1. choose concrete y_i
2. take convergence of Riemann sum in mean square sense

The choice

$y_i = t_{i-1}$ is characteristic for Ito stochastic integral
$y_i = \frac{1}{2}(t_i - t_{i-1})$ characterize Stratonovich integral.

So we calculated Ito stochastic $\int_0^T B_t(\omega) \mathrm{d}B_t(\omega)$. Stratonovich integral will be mentioned later.

The relation $E(\Delta_i B)^2 = \Delta_i$ and the fact that we assume convergence in mean square suggest

$$(\mathrm{d}B_t)^2 = \mathrm{d}t \tag{2.10}$$

respectively

$$\int_0^T (\mathrm{d}B_t)^2 = \int_0^T \mathrm{d}t = T \tag{2.11}$$

we got this rule only by a heuristic argument but it can also be proven rigorously.

This is still not enough to correctly define Ito stochastic integral but we believe that we showed the main idea behind stochastic integral. We will move to properties of Ito integral. For a more correct approach, a reader can see Appendix C.

Properties of Ito integral

First, we should define a class of processes, which we want to integrate. Integrands must be the processes $C = (C_t, t \in \langle 0, T\rangle)$ fulfilling conditions

1. C is adapted to Brownian motion
2. $\int_0^T EC_s^2 \mathrm{d}s < \infty$

We can state most important properties of Ito integral now. We will use a notion $I_t(C) = \int_0^t C_s \mathrm{d}B_s$

1. $I_t(aC_1 + C_2) = aI_t(C_1) + I_t(C_2) \quad a \in \mathbb{R}$
2. $I_t(C)$ is martingale with respect to natural filtration of Brownian motion
3. $EI_t(C) = 0$
4. $E(I_t(C))^2 = \int_0^t EC_s^2 \mathrm{d}s$
5. I_t has continuous trajectories

These properties follow from the construction of Ito integral, so we will not prove them here.

2.2.2 Ito lemma

We know from calculations we made at the beginning of this section

$$\int_0^T B_t(\omega)\mathrm{d}B_t(\omega) = \tfrac{1}{2}(B_T^2 - T)$$

So we can see that classical rules of integration do not hold for Ito integral. We need tools to calculate stochastic integrals analogous to chain rule in classical calculus. Such a tool is called Ito lemma.

Theorem 2.5. *(Ito lemma 1) Let us consider function $f(t,x) \in C^2$ i.e. with continuous second partial derivations then*

$$f(t, B_t) - f(s, B_s) = \int_s^t (f_1(x, B_x) + \tfrac{1}{2} f_{22}(x, B_x))\mathrm{d}x + \int_s^t f_2(x, B_x)\mathrm{d}B_x$$

where f_i means partial derivation $\frac{\partial f}{\partial x_i}$.

Remark. *This theorem can be formally derived using Taylor expansion with use of relation $(\mathrm{d}B_t)^2 = \mathrm{d}t$*

$f(t + \mathrm{d}t, B_t + \mathrm{d}B_t) - f(t, B_t) = f_1(t, B_t)\mathrm{d}t + f_2(t, B_t)\mathrm{d}B_t + \frac{1}{2}(f_{11}(t, B_t)(\mathrm{d}t)^2 + 2f_{12}(t, B_t)\mathrm{d}t\mathrm{d}B_t + f_{22}(t, B_t)(\mathrm{d}B_t)^2) + ...$

and now by neglecting all terms of higher order than $O(\mathrm{d}t)$ we get

$f(t + \mathrm{d}t, B_t + \mathrm{d}B_t) - f(t, B_t) = (f_1(t, B_t) + \frac{1}{2} f_{22}(t, B_t))\mathrm{d}t + f_2(t, B_t)\mathrm{d}B_t$

and we get Ito lemma by integration.

We will calculate few integrals using Ito lemma now.
For example, we can find Ito exponential, i.e. process X fulfilling

$$X_t - X_s = \int_s^t X_y \mathrm{d}B_y$$

we can easily verify using Ito lemma, that such process is $X_t = e^{B_t - \frac{1}{2}t}$ so we have

$$e^{B_t - \frac{1}{2}t} - e^{B_s - \frac{1}{2}s} = \int_s^t e^{B_y - \frac{1}{2}y} \mathrm{d}B_y$$

We can make similar calculation for process $X_t = e^{\sigma B_t + (c - \frac{1}{2}\sigma^2)t}$, $\sigma > 0$ then we will get, with use of Ito lemma

$$X_t - X_s = c\int_s^t X_y \mathrm{d}y + \sigma \int_s^t X_y \mathrm{d}B_y$$

or in differential form

$$\mathrm{d}X_t = cX_t \mathrm{d}t + \sigma X_t \mathrm{d}B_t$$

so we have found a stochastic differential equation solution of which is geometric Brownian motion. We will use this in the chapter on Black-Scholes formula.

We will try to generalize Ito lemma now. Let us consider a process in form

$$X_t = X_0 + \int_0^t A_s^{(1)} \mathrm{d}s + \int_0^t A_s^{(2)} \mathrm{d}B_s \tag{2.12}$$

where coefficients $A^{(i)}$ are adapted to Brownian motion. Such a process is called Ito process and it can be shown that if process X_t has representation in this form then coefficients $A^{(i)}$ are determined uniquely.

Theorem 2.6. *(Ito lemma 2) Let us consider Ito process X_t and function $f(t,x) \in C^2$. Then following formula holds*

$f(t, X_t) - f(s, X_s) = \int_s^t (f_1(y, X_y) + \frac{1}{2}(A_y^{(2)})^2 f_{22}(y, X_y) + A_y^{(1)} f_2(y, X_y))\mathrm{d}y + \int_s^t A_y^{(2)} f_2(y, X_y)\mathrm{d}B_y$

Remark. *This theorem can also be formally proven by Taylor expansion and by neglecting terms of order higher than $O(\mathrm{d}y)$. First, we will get a formula similar to Ito lemma 1.*

$$f(t, X_t) - f(s, X_s) = \int_s^t (f_1(y, X_y) + \tfrac{1}{2}(A_y^{(2)})^2 f_{22}(y, X_y))\mathrm{d}y + \int_s^t f_2(y, X_y)\mathrm{d}X_y$$

and now, with use of

$$\mathrm{d}X_y = A_y^{(1)}\mathrm{d}y + A_y^{(2)}\mathrm{d}B_y$$

we can get the theorem (2.6.).

We can easily generalize Ito lemma in the same manner for functions

$f(t, X_t^{(1)}, X_t^{(2)}, ..., X_t^{(n)})$

where

$$X_t^{(i)} = X_0^{(i)} + \int_0^t A_s^{(1,i)}\mathrm{d}s + \int_0^t A_s^{(2,i)}\mathrm{d}B_s$$

We will only state a formula for derivation of product of two processes, which can be derived by using Ito lemma on function $f(t, x_1, x_2) = x_1 x_2$.

$$\mathrm{d}(X_t^{(1)} X_t^{(2)}) = X_t^{(1)}\mathrm{d}X_t^{(2)} + X_t^{(2)}\mathrm{d}X_t^{(1)} + A_t^{(2,1)} A_t^{(2,2)}\mathrm{d}t \tag{2.13}$$

We will use this formula for some concrete calculations now.
Let us consider function $f(t) \in C^1$ *and processes* $X_t^{(1)} = f(t) - f(0) = \int_0^t f'(s)\mathrm{d}s$ *and* $X_t^{(2)} = B_t = \int_0^t \mathrm{d}B_t$ *then using formula above, we can get*

$$\mathrm{d}(X_t^{(1)} X_t^{(2)}) = (f(t) - f(0))\mathrm{d}B_t + B_t f'(t)\mathrm{d}t$$

by integrating and some rearranging of terms, we get a useful formula

$$\int_0^t f(s)\mathrm{d}B_s = f(t)B_t - \int_0^t f'(s)B_s\mathrm{d}s \tag{2.14}$$

Stratonovich integral

As we already mentioned for any choice of $y_i \in \langle t_{i-1}, t_i\rangle$, we have different stochastic integral. But in reality, only two of these integrals are used. Ito integral was discussed in the previous section. We will introduce Stratonovich integral given by choice $y_i = \frac{1}{2}(t_i - t_{i-1})$ now. Let us for simplicity consider a process C_t in a form $C_t = f(B_t)$ where $f \in C^2$, then Riemann sum has the form

$$\widetilde{S_n} = \sum_{i=1}^n f(B_{\frac{1}{2}(t_i - t_{i-1})})\Delta B_i$$

By some calculations, we can get formula for transformation between Ito and Stratonovich integral from this formula. We will denote Stratonovich integral by $\int_0^T C_t \circ \mathrm{d}B_t$ then

$$\int_0^T f(B_t) \circ \mathrm{d}B_t = \int_0^T f(B_t)\mathrm{d}B_t + \frac{1}{2}\int_0^T f'(B_t)\mathrm{d}t \tag{2.15}$$

Using this formula and Ito lemma we can calculate following Stratonovich integrals

$$\int_0^T f'(B_t) \circ \mathrm{d}B_t = f(B_T) - f(B_0)$$

$$\int_0^T B_t \circ \mathrm{d}B_t = \tfrac{1}{2}B_T^2$$

$$e^{\alpha t} - 1 = \int_0^T e^{\alpha t} \circ \mathrm{d}B_t \quad \alpha \in \mathbb{R}$$

So we can see that rules of classical integration formally holds for Stratonovic integral. This is of course an advantage of Stratonovich integral. However, Ito integral is much richer mathematical structure thanks to its martingale property. Stratonovich integral is clearly not martingale because for example process $\frac{1}{2}B_t^2$ is not martingale.

When we have a stochastic differential equation (SDE), we always have to know, if we should calculate it with respect to Stratonovich or Ito integration. But when this equation arises from a concrete problem, we have to make this decision ourselves. However, there is no universal answer, Ito integral is preferred much more often.

A transformation formula between these two integrals can be generalized for more general function $f(t, X_t)$ where X_t is Ito process. This is only needed when we solve SDE via Stratonovich calculus. It means that we transform Ito SDE to Stratonovich SDE and we can use methods similar to methods of classical integration then. This method will not be further discussed in this thesis. We will focus only on Ito SDE so we will omit more general transformation formulas.

2.2.3 Stochastic differential equation

We will consider Ito SDE with an initial condition in the form

$$\mathrm{d}X_t = a(t, X_t)\mathrm{d}t + b(t, X_t)\mathrm{d}B_t, \quad X_0(\omega) = Y(\omega) \tag{2.16}$$

So a result of this equation is stochastic process $X = (X_t, t \in \langle 0, T\rangle)$. For $b(t, x) = 0$, we clearly have an ordinary differential equation with a random initial condition. SDE above is interpreted as an integral equation,

$$X_t = X_0 + \int_0^t a(s, X_s)\mathrm{d}t + \int_0^t b(s, X_s)\mathrm{d}B_t$$

so in fact, SDE is an integral equation.

Problem is, if integrals on the left side exist in Riemann respectively Ito sense as defined in the previous section. We need to define solution of SDE for that. There exist two different definitions of such solution - strong and weak ones. We start with definition of the strong solution.

Definition 2.18. *Process $X = (X_t, t \in \langle 0, T\rangle)$ obeying SDE (2.16) is a strong solution if*

1. *X is adapted to Brownian motion*
2. *both integrals in (2.16) exists*
3. *X is function of underlying Brownian motion, coefficient $a(t, x), b(t, x)$ and initial condition $X_0(\omega)$*

It is important that X depends on sample paths of given Brownian motion. If we change underlying Brownian motion, a solution will change but it will not change functional form.

Remark. *This is a difference between weak and strong solutions. A weak solution depends only on distribution of X_t. Therefore a knowledge of concrete Brownian sample paths is not essential for a weak solution. We will only mention that there exists SDE with a weak solution only.*

Any solution of Ito SDE is usually called a diffusion. Brownian motion is a typical example of diffusion, of course.
From now on, we will consider only a strong solution of Ito SDE. We will state a following theorem about an existence and a uniqueness of a strong solution.

Theorem 2.7. *Let us assume that an initial condition in SDE (2.16) is independent of Brownian motion and has a finite variance and coefficients a, b fulfils $\forall x, y \in \mathbb{R}, t \in \langle 0, T\rangle$*

1. $|a(t,x) - a(t,y)| + |b(t,x) - b(t,y)| \leq K|x-y| \quad K > 0$
2. *a, b are continuous*

then there exists exactly one strong solution of SDE (2.16).

Linear SDE

We will focus on class of linear SDE now. Non-linear SDE appears quite rarely and are usually analytically unsolvable. We will not discuss them in this thesis.

Let us consider SDE in form

$$X_t = X_0 + \int_0^t (c_1(s)X_s + c_2(s))\mathrm{d}s + \int_0^t (\sigma_1(s)X_s + \sigma_2(s))\mathrm{d}B_s \quad t \in \langle 0, T\rangle \quad (2.17)$$

where coefficients c_i, σ_i are continuous. Such SDE is called a general linear stochastic differential equation.

From the theorem (2.7), we can say that linear SDE has one unique strong solution for a " suitable " initial condition X_0.

As we already mentioned, the main tool for solving SDE is Ito lemma. We will demonstrate a use of Ito lemma on less general examples of linear SDE.
Let us consider an equation

$$X_t = X_0 + c\int_0^t X_s\mathrm{d}s + \sigma\int_0^t X_s\mathrm{d}B_s \quad \sigma > 0, c \in \mathbb{R} \quad (2.18)$$

this SDE is called an equation with multiplicative noise. We already know that geometric Brownian motion is a solution of this equation but for the sake of the example, we will derive the solution using Ito lemma. If we denote $X_t = f(t, B_t)$, we get

$$X_t = X_0 + \int_0^t (f_1(s, B_s) + \tfrac{1}{2}f_{22}(s, B_s))\mathrm{d}s + \int_0^t f_2(s, B_s)\mathrm{d}B_s$$

Since Ito process has a unique representation, we only need to compare terms in integrals. We get the partial differential equations $cf = f_1 + \frac{1}{2}f_{22}$ and $\sigma f = f_2$

solving these equations and using the initial condition, we will get the right result

$$X_t = X_0 e^{(c-\frac{1}{2}\sigma^2)t+\sigma B_t} \tag{2.19}$$

Now let us consider another important example of linear SDE

$$X_t = X_0 + c\int_0^t X_s \mathrm{d}s + \sigma \int_0^t \mathrm{d}B_s \quad \sigma > 0, c \in \mathbb{R} \tag{2.20}$$

This is called Langevin equation or equation with additive noise. For a solution we will use substitution $Y_t = e^{-ct}X_t$. Ito lemma then gives us

$$Y_t - Y_0 = \int_0^t \sigma e^{-cs}\mathrm{d}B_s$$

from that we easily get solution

$$X_t = X_0 e^{ct} + \sigma e^{ct}\int_0^t e^{-cs}\mathrm{d}B_s \tag{2.21}$$

X_t is called Ornstein-Uhlenbeck process and it can be shown that it is Gaussian process.

A solution of general linear SDE can be calculated similarly but calculations are much longer. So, we will state only results.
A solution of general linear SDE (2.17) has a form

$$X_t = Y_t(X_0 + \int_0^t (c_2(s) - \sigma_1(s)\sigma_2(s))Y_s^{-1}\mathrm{d}s + \int_0^t \sigma_2(s)Y_s^{-1}\mathrm{d}B_s) \tag{2.22}$$

where Y_t has a form

$$Y_t = \exp(\int_0^t (c_1(s) - \tfrac{1}{2}\sigma_1^2(s))\mathrm{d}s + \int_0^t \sigma_1(s)\mathrm{d}B_s)$$

Y_t is actually a solution of (2.17) for $\sigma_2 = c_2 = 0$ and the initial condition $Y_0(\omega) = 1$.

Expected value of solution of SDE

When we have linear SDE in the form (2.17), we know a general solution and we can theoretically determine an expected value and other moments from it. However if we are interested only in expected value and/or variance, there is another way to calculate them.
If we calculate an expected value of both sides in (2.17) we get

$$\mu_X(t) = \mu_X(0) + \int_0^t (c_1(s)\mu_X(s) + c_2(s))\mathrm{d}s$$

where constancy of expected value of Ito integral was used. By derivation, we will get an ordinary differential equation

$$\mu_X'(t) = c_1(t)\mu_X(t) + c_2(t) \tag{2.23}$$

So if we can solve this equation, we have an expected value of solution in any time.

We can calculate second moment similarly, we use Ito lemma to calculate X_t^2 from (2.17) and use procedure as above. If we denote $EX_t^2 = q_X(t)$ then we get

$$q'_X(t) = (2c_1(t) + \sigma_1^2(t))q_X(t) + 2(c_2(t) + \sigma_1(t)\sigma_2(t))\mu_X(t) + \sigma_2^2(t) \qquad (2.24)$$

2.3 Levy processes

Definition 2.19. *Stochastic process $X = (X_t, t \geq 0)$ is called Levy process, if it satisfies following properties*

1. $X_0 = 0$ *a.s*
2. *X has stationary and independent increments*
3. $\lim_{h \to 0} P(|X_{t+h} - X_t| > \varepsilon) = 0 \quad \forall \varepsilon > 0, t \geq 0$
4. *X has a.s. right continuous sample paths with left limits*

Remark. *If only the first three conditions are satisfied, the process X is called Levy process in law. For every Levy process in law X, there exists such Levy process $\widetilde{X}$ that $P(X_t = \widetilde{X}_t) = 1, \forall t \geq 0$.*

Typical examples of Levy process is Poisson process and Brownian motion. Another important property of Levy processes follows from this theorem.

Theorem 2.8. *Let X be a Levy process. Then random variable X_t is infinitely divisible for any $t \geq 0$ and a distribution of X_t is given by X_1 by*

$$\varphi_{X_t}(u) = (\varphi_{X_1}(u))^t \qquad (2.25)$$

Proof. Infinite divisibility is clear because using independence and stationarity of increments, we easily get for any $s, t \geq 0$

$$\varphi_{X_{t+s}}(u) = \varphi_{X_t}(u)\varphi_{X_s}(u)$$

so we have

$$\varphi_{X_t}(u) = (\varphi_{X_{t/n}}(u))^n$$

for all n.
The second statement is bit tricky. We will show the main idea and omit technicalities. We can get (2.25) for any rational t using idea as above. Then we use continuity of $\varphi_{X_t}(u)$ in t , which follows from condition 3. in definition of Levy process. We expand validity of relation (2.25) by limit transitions to any $t \in \mathbb{R}$ with use of it. □

A connection between Levy processes and infinitely divisible distributions is underlined by the following theorem.

Theorem 2.9. *A random variable Y is infinitely divisible iff there exists Levy process X such as*

$$Y \stackrel{d}{=} X_1$$

Remark. *The left implication is obvious, the right one has to be proven. This prove is quite technical so we will omit it.*

We will state the most important theorem of this section now - Levy-Khintchine representation of Levy process.

Theorem 2.10. *Let X be a Levy process then there exists triplet (σ^2, γ, ν) such that*

$$\ln \varphi_{X_t}(u) = iu\gamma t - \frac{1}{2}\sigma^2 t u^2 + t\int_{\mathbb{R}} (e^{iux} - 1 - iuxI_{|x|<1})\mathrm{d}\nu(x) \tag{2.26}$$

where I denotes characteristic function, $\gamma \in \mathbb{R}, \sigma > 0$ and ν is a measure on $\mathbb{R}$ such that $\nu(\{0\}) = 0$ and

$$\int_{\mathbb{R}} min\{1, x^2\}\mathrm{d}\nu(x) < \infty$$

Such a triplet is determined uniquely.

Remark. *The condition required for measure ν gives us two pieces of information*

$$\int_{|x|\geq 1} 1\mathrm{d}\nu(x) < \infty$$

and

$$\int_{(-1,1)} x^2\mathrm{d}\nu(x) < \infty$$

It is enough to know that integral in (2.26) exists and is finite, because for $|x| \geq 1$ integrand behaves as $O(1)$ and for $|x| < 1$ both real an imaginary part behaves as $O(x^2)$ (that follows from $1 - \cos x \approx \sin x - x \approx x^2$).

We also know from theorems (2.9) and (2.10) that (2.26) with $t = 1$ is the general form for a characteristic function of a infinitely divisible random variable.

Meaning of parameters in triplet (σ^2, γ, ν) becomes clear with use of decomposition

$$\ln \varphi_{X_t}(u) = \{iu\gamma t - \frac{1}{2}\sigma^2 t u^2\} + \{t\int_{|x|<1} (e^{iux} - 1 - iux)\mathrm{d}\nu(x)\} + \{t\int_{|x|\geq 1} (e^{iux} - 1)\mathrm{d}\nu(x)\} \tag{2.27}$$

So we can see that any Levy process can be written as a sum of three independent Levy processes. Clearly any linear combination of independent Levy processes is Levy process.
The first bracket represents Brownian motion with drift, it is continuous part of original Levy process. So parameters γ and σ are drift and diffusion coefficients.

The last two brackets represent jumps of an original process, so we can see that measure ν actually describes a size and a rate of jumps. The second bracket is a pure jump square integrable martingale. The third bracket is more important, as it represents compound Poisson process. It will be discussed in next paragraph.

To summarize the previous discussion:

Every Levy process is actually composition of Brownian motion, linear drift and jumps, which are represented by compound Poisson process. Therefore Brownian motion with drift is only non-deterministic continuous Levy process.

We will also mention that this decomposition play a key role when defining more general stochastic integral along Levy sample paths.

Compound Poisson process

We will define compound Poisson process. Following calculations should give a better idea of structure of jumps in Levy process and will better explain meaning of measure ν.

Let us consider Poisson process $(N_t, t \geq 0)$ with a rate parameter λ and iid random variables $\{\xi_i, i = 1, 2...\}$ independent of N, with common distribution F that determines the size of jumps. Then process

$$X_t = \sum_{i=1}^{N_t} \xi_i \quad t \geq 0 \tag{2.28}$$

is called compound Poisson process. We can calculate a characteristic function of X_t

$$\sum_{n\geq 0} E(e^{iu\sum_{i=1}^{n}\xi_i})e^{-\lambda t}\frac{(\lambda t)^n}{n!} = \sum_{n\geq 0}(\int_{\mathbb{R}} e^{iux}\mathrm{d}F(x))^n e^{-\lambda t}\frac{(\lambda t)^n}{n!} = e^{-\lambda t\int_{\mathbb{R}}(1-e^{iux})\mathrm{d}F(x)}$$

so we have

$$\varphi_{X_t}(u) = e^{t\int_{\mathbb{R}}(e^{iux}-1)\lambda\mathrm{d}F(x)}$$

We can see that compound Poisson process is a Levy process with a diffusion parameter $\sigma = 0$ and a measure $\nu = \lambda\mathrm{d}F(x)$. If an interval $(-1, 1)$ is null set for measure given by distribution F, then it has exactly the same form as the third bracket.

By adding linear drift to X_t we get centered compound Poisson process

$$X_t = \textstyle\sum_{i=1}^{N_t}\xi_i + ct \quad t \geq 0$$

now if we set $c = \lambda\int_{\mathbb{R}} x\mathrm{d}F(x)$ we get

$$\varphi_{X_t}(u) = e^{t\int_{\mathbb{R}}(e^{iux}-1-iux)\lambda\mathrm{d}F(x)}$$

2.3.1 Stable processes

We will introduce an important subset of Levy processes,

Definition 2.20. *Levy process X is called α-stable, if X_1 is α-stable random variable.*

Remark. *We can see from (2.25) and properties of stable distributions that for α-stable process X and any $t \geq 0$ is X_t α-stable variable.*

So for α-stable process characteristic function has the following form

$$\ln \varphi_{X_t}(u) = it\gamma k - ct|k|^\alpha (1 + i\beta \frac{|k|}{k} \omega(k, \alpha)) \tag{2.29}$$

where function ω and the ranges of parameters are the same as in theorem (2.4). From (2.29) it is obvious that parameter β does not change in time and only scaling parameters γ, c change.

An important question is how we have to choose the triplet $(\sigma^2.\gamma, \nu)$ to get a stable process? There are only two options.

Theorem 2.11. *Non-trivial Levy process with generating triplet (σ^2, γ, ν) is α-stable for some $\alpha \in (0, 2\rangle$, if one of these conditions is fulfilled*

1. *$\alpha = 2$ and $\nu = 0$*
2. *$\alpha \in (0, 2)$, $\sigma = 0$ and $\nu = (c_+ I_{(0,\infty)} + c_- I_{(-\infty,0)})|x|^{-(\alpha+1)} \mathrm{d}x$ where $c_\pm \geq 0$*

Remark. *The first condition leads clearly to Gaussian process. The second condition is more complicated. We can theoretically integrate in (2.26) and we should get a result in form (2.29). That is quite complicated and we will outline a different way to prove this.*

Let Y be an infinitely divisible random variable. We denote a random variable with characteristic function φ_Y^r by Y^{*r}.
Let us consider Levy process X now. We will assume that $X_1 = Y$ is α-stable variable. It implies

$$Y^{*r^\alpha} = rY + c, \quad r > 0$$

where c is a real constant, both sides are infinitely divisible variables so they are uniquely determined by the triplet (σ^2, γ, ν). The previous equation is equivalent to equality of triplets

$$(r^\alpha \sigma^2, r^\alpha \gamma, r^\alpha \nu) = (r^2 \sigma^2, r\gamma + c, \nu \circ S_r^{-1})$$

where $S_r : x \longmapsto xr$, so evidently $\alpha = 2$ or $\sigma = 0$ must hold. From $r^\alpha \nu = \nu \circ S_r^{-1}$ we can derive the condition on ν. Let us denote $G(x) = \nu\{(x, \infty)\}$ respectively $\nu\{(-\infty, x)\}$ then

$$r^\alpha G(rx) = G(x) \quad r, x > 0$$

by choosing $r = 1/x$ we get

$$G(x) = x^{-\alpha}G(1)$$

we know from properties of ν given by theorem (2.10) that $G(1) < \infty$, for $\alpha = 2$ is $G(1) = 0$ (Gaussian case). Finally, for $\alpha \in (0, 2)$ we have

$$G(x) = \int_x^\infty \mathrm{d}\nu(y) = const.\ x^{-\alpha}$$

so $\mathrm{d}\nu(y) = c_+ y^{-\alpha-1} dy$ on $(0, \infty)$ and we proceed with the interval $(-\infty, 0)$ analogically.

2.4 Different approach to stochastic processes: Fokker-Planck equation

In this section, we will introduce a bit different approach to stochastic processes. We will be interested only in time development of PDF $p(x, t)$ of stochastic process X_t. A conditional expectation will be considered as defined by (1.3) without use of σ-algebras. We will also outline useful analogies between these two approaches throughout this section.

Historically, these approaches evolved very differently. The approach, which was used to define Ito integral is more mathematically rigorous but for some applications unnecessarily complicated. An approach, which will be demonstrated in this section was used by physicists when describing diffusion processes and formulating diffusion equations. This approach has some mathematical downsides. For example, conditional PDFs do not always exist and other assumptions are often implicitly required.

2.4.1 Markov processes

We will use a notation

$$p_{1|n-1}(x_n, t_n|x_1, t_1; ...; x_{n-1}, t_{n-1}) = p_{X_{t_n}|X_{t_1},...,X_{t_{n-1}}}(x_n|x_1, ..., x_{n-1})$$

Definition 2.21. *Process X_t is called Markov process, if for all $t_1 < t_2 ... < t_n$*

$$p_{1|n-1}(x_n, t_n|x_1, t_1; ...; x_{n-1}, t_{n-1}) = p_{1|1}(x_n, t_n|x_{n-1}, t_{n-1})$$

Remark. *In other words, Markov processes have no memory. Their future time evolution depends only on a present position.*

In terms of conditional expectation given by σ-algebras, we can define Markov process by

$$E(X_t|\mathcal{F}_s) = E(X_t|\sigma(B_s))$$

for all $t \geq s \geq 0$ and $\mathcal{F}_t = (\sigma(B_s), s \leq t)$ natural filtration of Brownian motion.

The theory behind Markov processes in terms of this definition is quite technical. It is why we omitted it in the previous sections, where we paid attention only to subclass of Markov processes - martingales.

Properties of Markov processes

There is a connection between Markov processes and processes with independent increments. In fact, every process with independent increments is Markov process but the reverse implication does not hold. We will prove this by following calculations. Let X_t be process with independent increments and $X_0 = 0$ then

$p_{X_{t_n}|X_{t_1},...,X_{t_{n-1}}}(x_n|x_1,...,x_{n-1}) =$
$p_{X_{t_n}-X_{t_{n-1}}+X_{t_{n-1}}|X_{t_1}-X_0,...,X_{t_{n-1}}-X_{t_{n-2}}}(x_n|x_1-x_0,...,x_{n-1}-x_{n-2}) =$
$(p_{X_{t_n}-X_{t_{n-1}}} * \delta_{x_{n-1}})(x_n) = p_{X_{t_n}-X_{t_{n-1}}}(x_n - x_{n-1})$

where we used fact that PDF of a sum of independent variables is convolution of individual PDFs, notation $\delta_a(x) = \delta(x-a)$ and we considered x_0=0. So we have proved:

$$\textit{process } X_t \textit{ has independent increments} \Rightarrow X_t \textit{ is Markov}$$

Another important property of Markov processes is a following identity

$$p_n(x_1,t_1;...;x_n,t_n) = \prod_{k=2}^{n} p_{1|1}(x_k,t_k|x_{k-1},t_{k-1})p_1(x_1,t_1) \tag{2.30}$$

this follows directly from using definition of Markov process iteratively on p_n. So for Markov process we only need a knowledge of two functions p_1 and $p_{1|1}$ to describe system completely. This is enormous simplification compared to a general stochastic process.

For $n = 3$ (2.30) reads

$$p_3(x_1,t_1;x_2,t_2;x_3,t_3) = p_{1|1}(x_3,t_3|x_2,t_2)p_{1|1}(x_2,t_2|x_1,t_1)p_1(x_1,t_1)$$

so by integrating over x_2 and dividing by $p_1(x_1,t_1)$ we get

$$p_{1|1}(x_3,t_3|x_1,t_1) = \int_{\mathbb{R}} p_{1|1}(x_3,t_3|x_2,t_2)p_{1|1}(x_2,t_2|x_1,t_1)\mathrm{d}x_2 \tag{2.31}$$

for $t_3 \geq t_2 \geq t_1$. This is Chapman-Kolmogorov equation for PDFs of Markov process.

Master equation

We will approximate Chapman-Kolmogorov equation for stationary Markov processes in this paragraph. For such process we have

$p_1(x,t) = p_1(x,0) = p_1(x)$
$p_{1|1}(x_2,t_2|x_1,t_1) = p_{1|1}(x_2,t_2-t_1|x_1,0) = p_{t_2-t_1}(x_2|x_1)$

so Chapman-Kolmogorov equation has a form

$$p_{t+t'}(x_3|x_1) = \int_{\mathbb{R}} p_t(x_2|x_1)p_{t'}(x_3|x_2)\mathrm{d}x_2 \tag{2.32}$$

Note that this is just matrix multiplication for discrete probability space .
We approximate $p_{t'}(x_3|x_2)$ for small t' by Taylor expansion now

$$p_{t'}(x_3|x_2) = (1 - \nu_{tot}(x_2)t')\delta(x_3 - x_2) + t'\nu(x_3|x_2) \tag{2.33}$$

Where $\nu(x_3|x_2)$ is transition probability per unit time from x_2 to x_3 and $1 - t'\nu_{tot}(x_2)$ is probability to remain in state x_2 up to time t'. So clearly

$$\nu_{tot}(x_2) = \int_{\mathbb{R}} \nu(x_3|x_2)\mathrm{d}x_3 \tag{2.34}$$

by using (2.33) and (2.34) we can rewrite Chapman-Kolmogorov equation (2.32)

$$\frac{p_{t+t'}(x_3|x_1) - p_t(x_3|x_1)}{t'} = \int_{\mathbb{R}} p_t(x_2|x_1)\nu(x_3|x_2)\mathrm{d}x_2 - \int_{\mathbb{R}} p_t(x_3|x_1)\nu(x_2|x_3)\mathrm{d}x_2$$

so in limit $t' \to 0$ we have

$$\frac{\partial}{\partial t}p_t(x_3|x_1) = \int_{\mathbb{R}} p_t(x_2|x_1)\nu(x_3|x_2)\mathrm{d}x_2 - \int_{\mathbb{R}} p_t(x_3|x_1)\nu(x_2|x_3)\mathrm{d}x_2 \tag{2.35}$$

this is master equation for stationary Markov process.

We will not presume a stationarity of Markov process but only time homogeneity of transition probability $p_{1|1}$. This is equivalent to stationary increments. Now we simplify a notation by rewriting $x_2 \to x'$, $x_3 \to x$ and $p_1 \to p$, divide (2.35) by $p_1(x,t)$ and integrate over x_1. We get master equation for probability density itself.

$$\frac{\partial}{\partial t}p(x,t) = \int_{\mathbb{R}} p(x',t)\nu(x,x')\mathrm{d}x' - \int_{\mathbb{R}} p(x,t)\nu(x',x)\mathrm{d}x' \tag{2.36}$$

Notice that we do not need Markov process to have stationary increments to derive this. We need only Taylor expansion (2.33) to hold for small t'. In other words, a condition of stationary increments must be satisfied only on small time scales.

We will derive further approximation of master equation. For sufficiently smooth functions $\nu(x|x')$, $p(x,t)$ we can derive Taylor expansion of (2.36).
Let us for $x = x' + r$ rewrite

$$\nu(x|x') = \nu(x|x - r) \equiv \nu(x - r, r)$$

the first argument is an initial position and second argument r is a jump size. It is important that the first argument is an initial position and it is not position after a jump. This choice is consistent with Markov property.
Now we can write

$$\frac{\partial}{\partial t}p(x,t) = \int_{\mathbb{R}} \nu(x - r, r)p(x - r, t)\mathrm{d}r - p(x,t)\int_{\mathbb{R}} \nu(x, -r)\mathrm{d}r \tag{2.37}$$

we can perform Taylor expansion of the first integral in $x - r$ around $r = 0$ now and we get

$$\frac{\partial}{\partial t}p(x,t) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!}\frac{\partial^n}{\partial x^n}(a_n(x)p(x,t)) \tag{2.38}$$

where we denoted

$$a_n(x) = \int_{\mathbb{R}} r^n \nu(x, r) \mathrm{d}r \tag{2.39}$$

This is Kramer-Moyals expansion of master equation. Notice that if we chose first argument of ν position after a jump we would write $\nu(x, r)$ instead of $\nu(x - r, r)$. We would get a different expansion (2.38) with $p(x, t)$ outside of sum then. This expansion would not have a good physical interpretation.

2.4.2 Fokker-Planck equation

In this section, we will approximate Kramer-Moyals expansion by truncation up to certain order of derivatives. Generally, we cannot neglect any terms in Kramer-Moyals expansion but if $\nu(x, r)$ is slowly varying function of first argument truncation is possible. This problem can be treated much more rigorously of course but we will omit more sophisticated discussions. We will consider a truncation after the second order

$$\frac{\partial}{\partial t} p(x, t) = -\frac{\partial}{\partial x}(a_1(x) p(x, t)) + \frac{1}{2} \frac{\partial}{\partial x^2}(a_2(x) p(x, t)) \tag{2.40}$$

this is famous Fokker-Plank equation, where a_1 is called a drift coefficient and a_2 a diffusion coefficient.

We derived originally master equation for transition PDF $p_{1|1}$ so we can consider Fokker-Plank equation in a form

$$\frac{\partial}{\partial t} p(x, t|x_0, t_0) = -\frac{\partial}{\partial x}(a_1(x) p(x, t|x_0, t_0)) + \frac{1}{2} \frac{\partial}{\partial x^2}(a_2(x) p(x, t|x_0, t_0))$$

this is often called forward Fokker-Planck or forward Kolmogorov equation. Similarly, we can derive backward Fokker-Planck (Kolmogorov) equation.

$$\frac{\partial}{\partial s} p(x, t|z, s) = -a_1(z) \frac{\partial}{\partial z} p(x, t|z, s) - \frac{1}{2} a_2(z) \frac{\partial}{\partial z^2} p(x, t|z, s) \tag{2.41}$$

Physically is forward Fokker-Planck equation much more important because it has an interpretation as a diffusion equation. We will remind that also every solution of Ito stochastic differential equation is called a diffusion. A link between SDE and Fokker-Planck equation will be discussed in the next paragraph.

Stochastic differential equation as diffusion process

Let us consider Ito SDE in a form

$$\mathrm{d}X_t = -\mu(X_t, t)\mathrm{d}t + D(X_t, t)\mathrm{d}B_t \tag{2.42}$$

let $p(x, t)$ be a probability density of X_t then

$$p(x, t + \Delta t) = \int_{\mathbb{R}} p(x - r, t) p_{1|1}(x, t + \Delta t|x - r, t)\mathrm{d}r \tag{2.43}$$

Let us assume time homogeneity of $p_{1|1}$ on small time scales so expansion (2.33) holds, then we get

$$p(x,t+\Delta t) = \Delta t \int_{\mathbb{R}} p(x-r,t)\nu(x-r,r)\mathrm{d}r + p(x,t)(1-\Delta t \nu_{tot}(x)) \quad (2.44)$$

where we have already integrated over $\delta(r)$.
We can use Taylor expansion on (2.44) in the same way as when we derived Kramer-Moyals expansion and truncate this expansion after second order derivatives. We should obtain Fokker-Planck equation. We will consider $p(x,t+\Delta t) - p(x,t) = \Delta t \frac{\partial}{\partial t} p(x,t)$ then

$$\Delta t \frac{\partial}{\partial t} p(x,t) = \Delta t \int_{\mathbb{R}} \{-r \frac{\partial}{\partial x}(p(x,t)\nu(x,r)) + \frac{1}{2} r^2 \frac{\partial}{\partial x^2}(p(x,t)\nu(x,r))\}\mathrm{d}r \quad (2.45)$$

we used identity $\nu_{tot}(x) = \int_{\mathbb{R}} \nu(x,r)\mathrm{d}r$.
We need to calculate moments of $\nu(x,r)$, in other words we need to calculate an average $\Delta X_t = X_{t+\Delta t} - X_t$ for small Δt. We have stationarity of increments on this scale so we can consider coefficients μ and D constant on this interval.

$$\Delta X_t = -\mu(X_t,t)\Delta t + D(X_t,t)\Delta B_t$$

By calculating expected value we get

$$E(\Delta X_t) = \Delta t \int_{\mathbb{R}} r\nu(x,r)\mathrm{d}r = -\mu(X_t,t)\Delta t \quad (2.46)$$

similarly we get

$$E(\Delta X_t)^2 = \Delta t \int_{\mathbb{R}} r^2\nu(x,r)\mathrm{d}r = D(X_t,t)^2\Delta t \quad (2.47)$$

here we used $E(B_t)^2 = t$ and neglected terms of higher order then $O(\Delta t)$.

In conclusion, we got an equation for probability density function of process X_t from (2.45)

$$\frac{\partial}{\partial t} p(x,t) = \frac{\partial}{\partial x}(\mu(x,t)p(x,t)) + \frac{1}{2}\frac{\partial}{\partial x^2}(D^2(x,t)p(x,t)) \quad (2.48)$$

So we have connected Fokker-Planck equation with SDE. So (2.42) and (2.48) describe the same process but SDE gives description on level of sample paths and Fokker-Planck equation shows time evolution of probability density.

We usually want to know both these descriptions, because we do not know path properties of process like fractal dimension from Fokker-Planck equation. This link between them is extremely important.

As a result of this calculations we can see that a solution of SDE (2.42) is Markov process. It has also time homogenous increments on small time scales, so expansion (2.33) works.

We will notice at the end of this section, that if we consider more general coefficients μ, D, which depend not only on present value of X_t, but also on past values or on some other processes, then we do not have to get a solution with Markov property. Also equation for time evolution of PDF might be more complicated than Fokker-Planck equation then.

Chapter 3

Fractal geometry

In this chapter, we will discuss fractals and we will show an important link between fractals and stochastic processes. We will also suggest few ways how fractals can be applied in finance. Proves of all theorems stated in this section can be found in [2].

3.1 Fractals

Sets and functions, that are not sufficiently smooth or regular for methods of classical calculus, were for long time considered just mathematical curiosity. But it turned out that such objects widely appear in nature or in finance and they are often better for a description of real systems than smooth curves or surfaces.

We should start with a definition of fractal but there is actually no satisfactory one. A founder of fractal geometry Mandelbrot has defined fractals as objects with different topological and Haussdorf dimension. A topological dimension is a dimension in classical sense. Haussdorf dimension is one of fractal dimensions, which will be discussed in details further. However, a lot of fractal-like sets would not be fractals by this definition. That is why we will omit a definition of a fractal. We actually do not need one, we can build a mathematical apparat applicable to any set. However, it will give us something new for fractals only. We will state some properties typical for fractals.

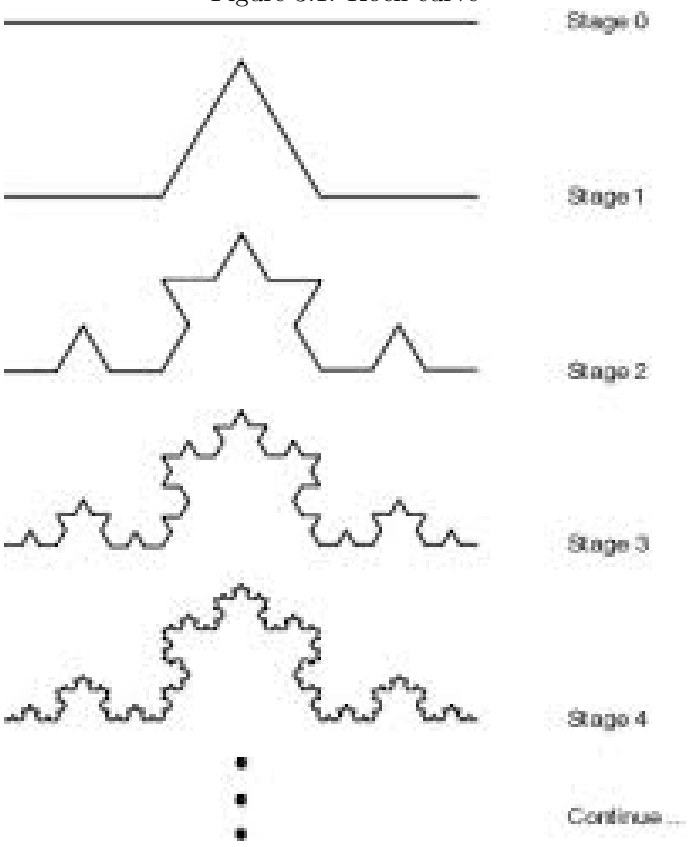
1. fractals have a fine structure i.e. details on every scale
2. fractals are self similar in some sense
3. fractals are too irregular to be described by usual geometrical methods

Fractal dimension belongs among the most important tools for description of fractals. There is a lot of non-equivalent definitions of fractal dimension, so we should always know, which one we are using. Property of self similarity implies

that many fractals can be generated recursively. So they can often be very easily described despite their irregularity by defining recursive procedure.

At the end of this introduction, we will state a well known example of fractal - Koch curve. Recursive procedure that generates Koch curve, is depicted in figure 4.1. Self similarity of Koch curve is obvious. This kind of self-similarity

Figure 3.1: Koch curve



is called strict self similarity. We will encounter also less strict self similarities, for example statistical self similarity, which occurs, when working with random fractals.

From this example, we can also see, why a classical definition of dimension fails in giving us enough information. Topological dimension of Koch curve is of course 1 and the length of this curve is clearly infinite. Further, this curve is always continuous but nowhere differentiable. This itself suggests that we will need more delicate methods for a description of fractals than classical calculus and geometry offer.

3.1.1 Haussdorf dimension and measure

We will start with a definition of the fractal dimension proposed by Haussdorf. His definition is probably the most valuable one from a purely mathematical point of view. However, there are some difficulties with computer calculations of Haussdorf dimension, so other definitions of dimension are often used in practise.

Let us start with a definition of Haussdorf measure.
We define a diameter for any set $U \subset \mathbb{R}^n$ with Euclidean measure (or more generally any metric space) by $|U| = \sup\{|x-y|; x, y \in U\}$. If we consider any set $F \subset R^n$, then δ-cover of F is any countable cover $\{U_i\}_{i\in\mathbb{N}}$ satisfying $|U_i| < \delta, \forall i$. We will define

$$\mathcal{H}^s_\delta(F) = \inf\{\sum_{i=1}^{\infty} |U_i|^s; \{U_i\}\ is\ \delta - cover\ of F\}$$

now we can define an s-dimensional Haussdorf measure by

$$\mathcal{H}^s(F) = \lim_{\delta\to 0} \mathcal{H}^s_\delta(F) \tag{3.1}$$

The definition is correct, the limit always exists because $\mathcal{H}^s_\delta(F)$ increases with $\delta \to 0$. Before we will use Haussdorf measure to define Haussdorf dimension, we will state some properties of $\mathcal{H}^s$.

Properties of Haussdorf measure

We should note that $\mathcal{H}^s$ is in fact a measure on $\mathbb{R}^n$ first. It is not trivial and quite technical to prove, so we will omit it.

$\mathcal{H}^n$ is in $\mathbb{R}^n$ equivalent measure to Lebesque measure, further for every Borel set $F \subset \mathbb{R}^n$ following holds

$$\mathcal{H}^n(F) = 2^n c_n^{-1} \mathcal{L}^n(F)$$

where $\mathcal{L}^n$ is a n-dimensional Lebesque measure and $c_n = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n+2}{2})}$ is a volume of a n-dimensional ball with radius 1.

It is obvious that when we multiply radius of a n-dimensional ball by a factor λ the volume will change with a scaling factor λ^n. We would expect that for $\mathcal{H}^s$, a correct scale factor will be λ^s. This assumption is supported by the following theorem.

For $F \subset \mathbb{R}^n$ a function $f : F \to \mathbb{R}^n$ with property

$$|f(x) - f(y)| = \lambda|x - y| \quad \forall x, y \in F$$

is called a similarity transformation of scale $\lambda > 0$. If $\lambda = 1$, f is called a congruence. Translations and rotations are typical examples of a congruence.

Theorem 3.1. *Let f be a similarity transformation of scale $\lambda > 0$, $F \subset \mathbb{R}^n$ then*

$$\mathcal{H}^s(f(F)) = \lambda^s \mathcal{H}^s(F)$$

Now we will proceed to define Haussdorf dimension.

Haussdorf dimension

We will start with a simple observation, for δ-cover $\{U_i\}$ of $F \subset \mathbb{R}^n$

$$\sum_i |U_i|^l = \sum_i |U_i|^{l-s}|U_i|^s \leq \delta^{l-s} \sum_i |U_i|^s \quad l > s$$

so if $\mathcal{H}^s(F) \in \mathbb{R}$ then $\mathcal{H}^l(F) = 0,\ \forall l > s$ and $\mathcal{H}^t(F) = \infty,\ \forall t < s$ so we can define Haussdorf dimension in the following way

$$dim_H F = \inf\{s \geq 0, \mathcal{H}^s(F) = 0\} = \sup\{s \geq 0, \mathcal{H}^s(F) = \infty\} \tag{3.2}$$

Notice that $\mathcal{H}^s(F) \in \langle 0, \infty\rangle$. Further, if F is Borel set and $\mathcal{H}^s(F) \in (0, \infty)$ F is called a s-set.

Properties of Haussdorf dimension

Haussdorf dimension satisfies following properties that should be expected to hold for an every reasonably defined dimension.

1. topological (Lebesque) dimension is always smaller or equal than Hassdorf dimension
2. $E \subset F \Rightarrow dim_H E \leq dim_H F$, this property is called a monotonicity
3. $dim_H \bigcup_{i\in\mathbb{N}} F_i = \sup\{dim_H F_i\}$
4. all countable sets G have $dim_H G = 0$
5. $F = F^\circ \subset \mathbb{R}^n$ then $dim_H F = n$
6. smooth m-dimensional manifolds have Haussdorf dimension also m

We will also state the following theorem.

Theorem 3.2. *Let $F \subset \mathbb{R}^n$ and we consider a function $f : F \to \mathbb{R}^n$ satisfying condition*

$$|f(x) - f(y)| \leq c|x - y|^\alpha \quad \forall x, y \in F$$

then $dim_H f(F) \leq (1/\alpha) dim_H F$.

This theorem has an important consequence. Let us consider a function $f : F \to \mathbb{R}^n$ satisfying

$$c_1|x - y| \leq |f(x) - f(y)| \leq c_2|x - y| \quad \forall x, y \in F$$

where $0 < c_1 \leq c_2$. Such a function is called a bi-Lipschitz transformation and we have $dim_H f(F) = dim_H F$ from theorem (4.2). It means that Haussdorf dimension is bi-Lipschitz invariant. So we can see bi-Lipschitz transformation plays here a similar role as a homeomorphism in topology.

Calculation of Haussdorf dimension

There are many ways how to calculate a fractal dimension, we have to use a computer for these calculations very often. However, a dimension can be calculated also analytically for a lot of fractals. Here, we will show only few simple examples.

Let F be a Koch curve. We know that F consists of 4 identical disjunct copies of F scaled by factor 1/3. We denote each of those copies by F_i and we get

$$\mathcal{H}^s(F) = 4\mathcal{H}^s(F_i)$$

using the theorem (4.1) we get

$$\mathcal{H}^s(F) = 4(1/3)^s\mathcal{H}^s(F)$$

if we assume that such s exists, for which $\mathcal{H}^s(F)$ is real and positive, we can see that $s = \frac{\ln 4}{\ln 3}$.

Mass distribution principle

We will introduce a very useful method for a calculation of a dimension here.

Theorem 3.3. *Let $\varepsilon > 0$ and $c > 0$ and let μ be a measure on $F \subset \mathbb{R}^n$. Let $\mu(F) \in (0, \infty)$ (such μ is often called a mass distribution on F), further let μ fulfil*

$$\mu(U) \leq c|U|^s \quad (\forall U)(|U| \leq \varepsilon)$$

then $\mathcal{H}^s(F) \geq \mu(F)/c$ which implies

$$s \leq dim_H F$$

A proof of this theorem follows immediately from the following calculation. Let $\{U_i\}$ be a cover of F then

$$0 < \mu(F) \leq \sum_i \mu(U_i) \leq c \sum_i |U_i|^s$$

This theorem is often very useful for a lower estimates of Haussdorf dimension. The next theorem will give us a better understanding of a meaning of Haussdorf dimension.

Theorem 3.4. *Let μ be a mass distribution on $\mathbb{R}^n$, $c > 0$ constant and F Borel subset of $\mathbb{R}^n$. We denote a ball with radius r and a centre in x by $B(x, r)$, then*

1. $\limsup_{r\to 0} \ \mu(B(x,r))/r^s < c \quad \forall x \in F \Rightarrow \mathcal{H}^s(F) \geq \mu(F)/c$
2. $\limsup_{r\to 0} \ \mu(B(x,r))/r^s > c \quad \forall x \in F \Rightarrow \mathcal{H}^s(F) \leq 2^s\mu(\mathbb{R}^n)/c$

The direct and very important consequence of this theorem is

$$\lim_{r\to 0} \frac{\ln \mu(B(x,r))}{\ln r} = s \quad \forall x \in F \Rightarrow dim_H F = s \tag{3.3}$$

3.1.2 Box counting dimension

Let us denote the smallest number of sets with a diameter at most δ needed to cover F by $M_\delta(F)$. We assume for $\delta \to 0$ a behaviour

$$M_\delta(F) \sim c\delta^{-s}$$

so we have

$$s = \lim_{\delta \to 0} \frac{-\ln M_\delta(F)}{\ln \delta} \tag{3.4}$$

This is an idea behind a box counting dimension. This limit does not have to exist always. To be correct, we have to define a lower box counting dimension by

$$\underline{dim}_B F = \liminf_{\delta \to 0} \frac{-\ln M_\delta(F)}{\ln \delta} \tag{3.5}$$

and we define an upper box counting dimension in the same manner. If they are equal to each other, we will denote both of them by $dim_B F$.

There are many equivalent definitions of box counting dimension. We will state only one, which is usually the most useful. Let us consider cubes in $\mathbb{R}^n$ in the form

$$\langle m_1\delta, (m_1+1)\delta\rangle \times ... \times \langle m_n\delta, (m_n+1)\delta\rangle$$

where m_i are integers. We denote by $N_\delta(F)$ a number of these cubes, which intersect $F \subset \mathbb{R}^n$. We can show with a little bit of effort that

$$M_{\delta\sqrt{n}}(F) \leq N_\delta(F) \leq 3^n M_\delta(F)$$

So we can equivalently define a box counting dimension by calculating a number of these cubes that intersect F instead of a number of any sets with a diameter at most δ that cover F.

Thanks to this definition, a box counting dimension is clearly more suitable for computer calculations than Haussdorf dimension. These dimensions are not equivalent but they usually coincide for "reasonable" fractals. A box counting dimension has many non-convenient properties from a mathematical point of view. Its theoretical significance lies in situations, where it is equal to Haussdorf dimension.

Properties of box counting dimension

We will start with a comparison of Haussdorf and a box counting dimension. We use an inequality

$$\mathcal{H}^s_\delta(F) \leq \delta^s M_\delta(F)$$

for $1 < \mathcal{H}^s(F)$ we will get

$$s \leq \lim_{\delta \to 0} \frac{-\ln M_\delta(F)}{\ln \delta}$$

so we have showed $dim_H F \leq dim_B F$. This holds for any set F because the condition $1 < \mathcal{H}^s(F)$ can be achieved by a proper scaling; for $\mathcal{H}^s(F) = 0$ is this inequality trivial.

Now we will summarize basic properties of a box counting dimension.

1. $dim_H F \leq \underline{dim}_B F \leq \overline{dim}_B F$
2. $\underline{dim}_B F, \overline{dim}_B F$ are both monotone
3. $\underline{dim}_B F, \overline{dim}_B F$ are both bi-Lipschitz invariant
4. $\overline{dim}_B(E \cup F) = \max\{\overline{dim}_B E, \overline{dim}_B F\}$, this property hold only for a finite union and in the case of $\underline{dim}_B$ it is not fulfilled at all
5. $\underline{dim}_B(F) = \underline{dim}_B(\overline{F})$ the same identity holds for $\overline{dim}_B$.
   This property is especially inconvenient because the countable dense sets in some area have the same box counting dimension as the whole area. This severely limits a use of a box counting dimension.

We will introduce one of many examples of sets with different box counting and Haussdorf dimension at the end of this section.
The set $F = \{0, 1, \frac{1}{2}, \frac{1}{3}, ...\}$ has clearly $dim_H F = 0$ but it can be shown that $dim_B F = 1/2$.

3.1.3 Self similar sets

As we have already mentioned, many fractals can be generated recursively. We will formalize this notion now.

Let $D = \overline{D} \subset \mathbb{R}^n$ then $S : D \to D$ is called a contraction if

$$|S(x) - S(y)| \leq c|x - y| \quad c \in (0, 1)$$

if the equality holds, S is called a contracting similarity.

A finite set $\{S_1, ..., S_n\}$ of contractions is called an iterated function system (IFS). A compact set $F \subset D$ satisfying $F = \bigcup_{i=1}^{m} S_i(F)$ is called an attractor of IFS. This attractor is very often fractal. It can be shown that such an attractor exists and it is unique. We can use Banach contracting mapping theorem to prove this. We will not go into details. However, we will mention, that from Banach theorem follows also, that for any compact set E fulfilling $S_i(E) \subset E$ following holds

$$F = \bigcap_{k=0}^{\infty} S^k(E)$$

where $S(E) = \bigcup_{i=0}^{n} S_i(E)$.
So if we have IFS, we can get a good approximation of an attractor by simple iterations.

The following theorem is useful for calculations of dimension of self similar sets.

Theorem 3.5. *Let $\{S_1, ..., S_n\}$ be IFS such that*

$$|S_i(x) - S_i(y)| \leq c_i|x - y| \quad c_i \in (0, 1)$$

further let such open set V exists that $\bigcup_{i=1}^{n} S_i(V) \subset V$ where this union is disjoint. Let F be an attractor of this IFS then

$$dim_B F = dim_H F = s$$

where $\sum_{i=1}^{m} c_i^s = 1$.

We actually used this theorem, when we calculated a dimension of Koch curve. The condition on an existence of a set V is required because we need individual similar copies not to overlap much. This was satisfied obviously in the case of Koch curve.

The copies can overlap without an existence of such V so only following would be true $dim_B F = dim_H F \leq s$. We will mention, that if S_i are affine contractions, an attractor is called a self affine set.

3.2 Random fractals

We often encounter fractals in physics or finance, which are generated somehow randomly. We discussed IFS as a possibility to generate fractals in the last section. This procedure can be easily randomize. We will omit general technical theorems and will show a concrete example only.

Let us consider a random Koch curve. A size of the part in the middle, that is replaced by equilateral triangle in each step, is a random variable X with a uniform distribution $X \sim U(0, \frac{1}{3})$. We can similarly as in the theorem (4.5) calculate a dimension of this fractal by

$$1 = E(2(1/2(1-X))^s + 2X^s) = \int_0^{1/3} 3(2(1/2(1-x))^s + 2x^s)\mathrm{d}x$$

se we get $dim_B F = dim_H F = s = 1,44$.

We will omit further generalizations and move to a different more important example of random fractal - Brownian motion.

3.2.1 Brownian motion

We have already discussed many properties of Brownian motion but here we will take another look at it. Let B_t be a Brownian motion in $\mathbb{R}^n$, then we can consider a graph $\{(t, B_t(\omega)); t \in \mathbb{R}\} \subset \mathbb{R}^{n+1}$. We can also consider a trail of Brownian motion $\{B_t(\omega); t \in (t_1, t_2)\}$. Both these sets are fractals. We will determine their fractal dimensions and other properties.

We have already discussed a self similarity of Brownian motion.

$$(T^{1/2}B_{t_1}, ..., T^{1/2}B_{t_n}) \stackrel{d}{=} (B_{Tt_1}, ..., B_{Tt_n})$$

This is the case of a statistical self similarity. A coefficient $H = 1/2$ is often called Hurst index and is directly connected to a fractal dimension. It is obvious that with decreasing H, sample paths of stochastic process will be more dense and fractal dimension will increase. We will state theorems about fractal dimension of Brownian motion now.

Theorem 3.6. *Let B be Brownian motion in $\mathbb{R}^n, n \geq 2$ then a trail $F = \{B_t(\omega); t \in (t_1, t_2)\}$ has a.s. fractal dimension*

$$dim_B F = dim_H F = 2$$

This is very interesting result. A fractal dimension of a trail of Brownian motion does not depend on a dimension of a space, where it is realized. This is not generally true for most of stochastic processes.

Theorem 3.7. *Let B be Brownian motion in $\mathbb{R}$ then a graph $G = \{(t, B_t(\omega)); t \in (t_1, t_2)\}$ has a.s. fractal dimension*

$$dim_B G = dim_H G = 3/2 \tag{3.6}$$

We have only some sample paths to analyse in practice. This theorem gives us a new and powerful tool to determine if an underlying distribution of these sample paths is Gaussian. We only need to calculate a box counting dimension, which is quite easy with a use of computers. We can find out thanks to this theorem, if a underlying distribution is normal. In reality, it is more complicated of course. The main problem is that we do not usually encounter fractals but multifractals. This will be further discussed at the end of this chapter.

3.2.2 Fractional Brownian motion

Definition 3.1. *Stochastic process $(X_t, t \geq 0)$ is called a fractional Brownian motion (FBM) with a parameter $H \in (0, 1)$ if*

1. $X_0 = 0$ *a.s.*
2. $X_t(\omega)$ *is continuous a.s.*
3. $X_{t+h} - X_t \sim N(0, h^{2H})$

We can see that FBM has stationary increments from the property 3. A question is, if FBM has independent increments. From $E(X_{t+h} - X_t)^2 = h^{2H}$ we get

$$E(X_t X_{t+h}) = \frac{1}{2}(t^{2H} + (t+h)^{2H} - h^{2H})$$

so we can write

$$EX_t(X_{t+h} - X_t) = \frac{1}{2}(-t^{2H} + (t+h)^{2H} - h^{2H}) \overset{H \neq \frac{1}{2}}{\neq} 0 \tag{3.7}$$

So we have a standard Brownian motion for $H = \frac{1}{2}$ and FBM without independent increments for $H \neq \frac{1}{2}$. That makes FBM an important concept for generalizations beyond independent increments and Markov processes.

Another interesting property of FBM is that $E(X_t - X_0)(X_{t+h} - X_t)$ is positive for $H > 1/2$ and negative for $H < 1/2$. In conclusion, for $H > 1/2$ if X_t grew in past, it is likely to keep growing in future. This is often called a

super diffusive behaviour. On the other hand, for $H < 1/2$ X_t is kept oscillating close to a concrete value. This indicates a sub-diffusive behaviour.

The scaling property of FBM is obvious

$$(T^H X_{t_1}, ..., T^H X_{t_n}) \stackrel{d}{=} (X_{Tt_1}, ..., X_{Tt_n})$$

We will also state a theorem about a fractal dimension of FBM.

Theorem 3.8. *Let X_t be fractional Brownian motion with a parameter H then a.s. fractal dimension of a graph $G = \{(t, X_t(\omega)); t \in (t_1, t_2)\}$ is*

$$dim_H G = dim_B G = 2 - H$$

We will mention a possible generalization of FBM to a multifractal Brownian motion. Multifractals differ from fractals by a diverse behaviour on different scales, i.e. we need more than one Hurst index to describe them. So if we consider the parameter H as a function $H : t \to (0, 1)$, we will get a process with a multifractional behaviour.

We will briefly introduce fractional Brownian surfaces.

Definition 3.2. *Set of random variables $(X_{t,s}; t, s \geq 0)$ fulfilling following conditions*

1. *$X_{0,0} = 0$ a.s.*
2. *$X_{t,s}(\omega)$ is continuous a.s.*
3. *$X_{t+h,s+k} - X_{t,s} \sim N(0, (h^2 + k^2)^{2H})$*

is called a fractional Brownian surface with a parameter H.

A similar theorem as for FBM holds here.

Theorem 3.9. *Let $X_{t,s}$ be a fractional Brownian surface with a parameter H then a.s. fractal dimension of the graph $G = \{(t, s, X_{t,s}(\omega)); t \in (t_1, t_2), s \in s_1, s_2\}$ is*

$$dim_H G = dim_B G = 3 - H$$

3.2.3 Levy stable processes

We have already known that a general form of a characteristic function of stable process is given by (2.29). If we consider the parameter $\gamma = 0$ in (2.29), we have scaling property for α-stable process X_t

$$(T^{1/\alpha} X_{t_1}, ..., T^{1/\alpha} X_{t_n}) \stackrel{d}{=} (X_{Tt_1}, ..., X_{Tt_n})$$

so clearly Hurst index $H = 1/\alpha$.
We will consider symmetric stable processes (with the parameter $\beta = 0$). For calculations of their fractal dimension, we have the following theorem.

Theorem 3.10. *A graph G of α-stable symmetric process X_t has a.s. fractal dimension given by*

$$dim_H G = dim_B G = \max\{1, 2 - 1/\alpha\}$$

So for $\alpha \leq 1$ α-stable process X_t has fractal dimension 1. This is consistent with the fact that X_t has an infinite number of small jumps on every interval for $\alpha \leq 1$. For $1 < \alpha < 2$ X_t, it combines a continuous component with jumps.

As in the case of Brownian motion, this theorem is a very important tool for us. If we have a graph and expect underlying distribution to be stable, this theorem helps us determine the parameter α of distribution.

3.3 Multifractals

We have discussed fractals with only one scaling (Hurst) index in this section till now. We will consider a very important generalization to fractals with the whole spectrum of scaling indexes - multifractals. We will use a mass distribution principle for mathematical description of these sets. If $\mu(B(x,r)) \simeq r^s$ holds for every point of given set, then the theorem (4.4) says that this set has Haussdorf dimension equal to s. However, if a mass distribution μ is spread irregularly and this power law behaviour differs for individual points, a knowledge of just fractal dimension is not enough. We need to know the whole multifractal spectrum. These mass distributions are called multifractal measures.

From now on, we will not consider sets but directly multifractal measures. In reality, we have a set, for which computers can calculate corresponding multifractal measure. Then we can apply some other methods, few of them will be described in the next section.

3.3.1 Multiftactal spectrum

We will consider multifractal measure $0 < \mu(\mathbb{R}^n) < \infty$, naturally a support of μ is an original multifractal set. We will define two kinds of multifractal spectrums - coarse and fine ones. A connection between them is very similar to a connection between Haussdorf and a box counting dimension. We will also discuss Legendre transform method for a calculation of multifractal spectrum.

Coarse multifractal analysis

Let us consider r-mesh cubes in the form

$$\langle m_1\delta, (m_1+1)\delta\rangle \times ... \times \langle m_n\delta, (m_n+1)\delta\rangle$$

where m_i are integers. For $\alpha \geq 0$ we write

$$N_r(\alpha) = |\{r - mesh\ cubes\ with\ \mu(C) \geq r^\alpha\}|$$

where $|\{...\}|$ denotes a number of elements in a set. Now we can define a coarse multifractal spectrum by

$$f_C(\alpha) = \lim_{\varepsilon \to 0} \lim_{r \to 0} \frac{\ln^+ \left(N_r(\alpha + \varepsilon) - N_r(\alpha - \varepsilon) \right)}{-\ln r} \tag{3.8}$$

where we write $\ln^+ x = max\{0, \ln x\}$ to ensure non-negativity of $f_C(\alpha)$. To ensure an existence of a coarse spectrum, we define an upper coarse spectrum $\overline{f_C}(\alpha)$ and a lower coarse spectrum $\underline{f_C}(\alpha)$ in the usual way.

For small ε, we have a power law behaviour

$$N_r(\alpha + \varepsilon) - N_r(\alpha - \varepsilon) \approx r^{-f_C(\alpha)} \tag{3.9}$$

so we can see that $f_C(\alpha)$ is a power law exponent for a number of r-mesh cubes B that behave as $\mu(B) \simeq r^\alpha$. Some estimates of a coarse spectrum can be numerically calculated, if we choose r small enough and we box-count a number of cubes that fulfil $\mu(B)/\ln r \simeq \alpha$. This calculations are however numerically quite awkward, so we will introduce a different method.

We can start by defining qth moments for $q \in \mathbb{R}, r > 0$

$$M_r(q) = \sum_{C_r} \mu(B)^q \tag{3.10}$$

where a sum is over such r-mesh cubes that $\mu(B) > 0$. We will define a function $\beta(q)$ by

$$\beta(q) = \lim_{r \to 0} \frac{\ln M_r(q)}{-\ln r} \tag{3.11}$$

Now let us consider $q \geq 0$, then clearly the following inequality holds

$$M_r(q) \geq r^{q\alpha} N_r(\alpha)$$

so for a sufficiently small r and any ε we get

$$(\varepsilon + \alpha)q + \frac{\ln M_r(q)}{-\ln r} \geq \frac{\ln N_r(\alpha + \varepsilon)}{-\ln r}$$

so by sending $r \to 0$ and using (3.9) we have $\forall q > 0$

$$\overline{f_C}(\alpha) \leq \beta(q) + \alpha q \tag{3.12}$$

Similarly, this can be shown for $q < 0$. So (3.12) actually holds for any real q. This motivates us to define Legendre transform of β by

$$f_L(\alpha) = \inf_{q \in \mathbb{R}} \{\beta(q) + \alpha q\} \tag{3.13}$$

From calculations we made, we have a following relation between Legendre transform of β and a coarse spectrum.

Theorem 3.11. *Let μ be a finite measure on $\mathbb{R}$ then $\forall\alpha \geq 0$ if limit (3.11) exists*

$$\underline{f_C}(\alpha) \leq \overline{f_C}(\alpha) \leq f_L(\alpha)$$

So Legendre transform of β gives us an upper estimate of coarse spectrum but the equality $f_L(\alpha) = f_C(\alpha)$ holds very often. It is why Legendre transform $f_L(\alpha)$ is sometimes taken as a coarse spectrum.

Fine multifractal analysis

A coarse spectrum provides us information about fluctuations of the measure μ. A fine multifractal spectrum will tell us more about a limiting behaviour of the multifractal measure.

We can start by defining a local dimension of the measure μ

$$dim_{loc}\mu(x) = \lim_{r\to 0} \frac{\ln \mu(B(x,r))}{\ln r} \tag{3.14}$$

now, if we denote $F_\alpha = \{x \in \mathbb{R}^n, dim_{loc}\mu(x) = \alpha\}$, we can define a fine spectrum as

$$f_H(\alpha) = dim_H(\alpha) \tag{3.15}$$

From the theorem (3.4), we have an inequality

$$0 \leq f_H(\alpha) \leq \alpha \tag{3.16}$$

Calculating a box counting dimension of F_α is not very useful because in the most of interesting cases F_α is dense in support of μ. We know that $dim_B F = dim_B \overline{F}$ and so we would get a measure of the whole multifractal.

As in the case of a box counting and Haussdorf dimension, a fine analysis is more valuable from a theoretical point of view but a coarse spectrum is more practical for real calculations. Also the same inequality holds between them.

Theorem 3.12. *For μ finite measure on $\mathbb{R}^n$ and every $\alpha \geq 0$ we have*

$$f_H(\alpha) \leq \underline{f_C}(\alpha)$$

We also get an equality between coarse and fine spectra very often.

So far, we have defined a multifractal spectrum and discussed ways to calculate it. Let us consider a situation now, where we have a spectrum and we will outline a way to use it on a simple example.

We consider a graph of the price evolution of an asset. If we find out, that it is a multifractal, then we can calculate its spectrum $(f(\alpha), \alpha)$. We believe, based on a generalized central limit theorem and some other empirical data, that an underlying distribution is stable. So, using the theorem (3.10), we know corresponding stable distribution for every α from spectrum. We can calculate an underlying distribution as a weighted mixture of these stable

distributions, where their weights is given by $f(\alpha)$. We can consider these individual stable distributions independent, so we only need to calculate their convolution. Notice that because these distributions do not have the same tail parameter, the resulting distribution does not have to be stable.

Our approach was based on a generalized central limit theorem, which gives stable distributions. However, we can get resulting distribution non stable. It sounds suspicious but there is a good reason for it. Multifractal behaviour suggests some memory effects, so the price evolution process is not Markov. Respectively, individual variables are not independent, so none of limit theorems we discussed so far can work. On the other hand, an independence can hold on individual scales represented by multifractal spectrum. So on these scales, our limit theorems can work. For example, we can usually expect CLT to work on very small scales, so corresponding distribution will be Gaussian. We will also mention that we can assess a stability of the market based on a weight of Gaussian distribution in this mixture of stable distributions. We know that an underlying distribution should be Gaussian and a graph would be only monofractal on a completely stable market.

Chapter 4

Black-Scholes formula

In this chapter, we will explain basic terminology used on the financial markets. Then we will derive a famous Black-Scholes formula for pricing of European options and discuss its limitations and possible generalizations.

4.1 Financial markets

Trading on financial markets has so many aspects that any mathematical model cannot address all of them. It is why we always have to make assumptions about a market, which are not completely fulfilled. However, they should be close enough to reality to keep our models well working. We will always consider a risky asset (stock). A price of this asset will be given by a stochastic process. Further, we will consider bonds, risk free assets, which provide fixed interests. An example of such bonds is a bank account. We will omit possibilities like a bankruptcy or so. We will consider only losses/gains caused by price evolutions of risky assets.

In last decades, some traders shifted their focus from trading stocks to trading derivatives - assets, of which prices are derived from a price of some underlying assets. Mathematically speaking, if a price of the underlying asset is given by a stochastic process X_t, then the price of this derivative will be a function $f(X_t)$. We will focus on one of the most popular derivatives now.

4.1.1 Options

A put/call option is a derivative that gives us the right to sell/buy an underlying asset at certain time for an agreed price. The price is usually called a strike price. We also distinguish between European and American options. European options give us a right to buy/sell only at given time, which is usually called expiration time of option. American options give us a right to do so in any time between a moment we bought the option till its expiration. From now on we will consider only European options.

4.2 Black-Scholes option pricing

We will be interested in determining right prices of options. For deriving Black-Scholes formula, we need to make some assumptions about markets first. Let us assume that financial markets fulfils following conditions

1. all information relevant to the trading can be quickly obtained by all participants
2. a market is liquid, this means that we can in any time buy or sell any stocks, which we want
3. there is a low friction on the market - this mean that all trading costs and fees are negligible

Such markets are called efficient markets. The efficient market hypothesis states that asset prices on such markets follow Markov stochastic processes.

We can move to a mathematical formulation of an option pricing problem. We have to find a suitable stochastic process that will describe a time evolution of risky assets first. We will choose a geometric Brownian motion as such a process. So the price evolution is given by following SDE

$$\mathrm{d}X_t = cX_t\mathrm{d}t + \sigma X_t \mathrm{d}B_t \tag{4.1}$$

and we already know a solution of this SDE

$$X_t = X_0 e^{(c-\frac{1}{2}\sigma^2)t+\sigma B_t} \tag{4.2}$$

A deterministic part of the equation (4.1) suggests that prices of assets grow exponentially and fluctuations are given by Brownian motion. Many economists believe in exponential growth so this assumption seems reasonable.

With use of Ito lemma, we get

$$\mathrm{d}\ln X_t = (c - \frac{1}{2}\sigma^2)\mathrm{d}t + \sigma \mathrm{d}B_t$$

so we can see that a logarithm of the price follows Brownian motion. It means that a price of assets is log-normally distributed.

The parameters $c, \sigma > 0$ are called a rate of return and volatility. The volatility determines how "dangerous" trading on the market is. For now, we will consider these parameters as constants.

We will define bonds mathematically - we will presume an bank account, where interests are paid continuously

$$\beta_t = \beta_0 e^{rt} \tag{4.3}$$

where β_0 is an initial deposit and $r > 0$ is an interest rate.

4.2.1 Building portfolio

We can define a portfolio now. For simplicity, we assume that we have only one type of stock and one kind of bond. a_t will be a number of stocks that we have. b_t will be a number of bonds. Both a_t, b_t are stochastic processes and we presume that they are adapted to Brownian motion. We will call

$$((a_t, b_t), t \in \langle 0, T \rangle)$$

a trading strategy. So our fortune at any moment can be written as

$$V_t = a_t X_t + b_t \beta_t \tag{4.4}$$

We should note that $a_t, b_t \in \mathbb{R}$ so they can be negative. It corresponds with a situation, when we borrow money or make a deal with someone to buy stock in future for in advance agreed price. This kind of deal is called taking a short position.

Processes a_t, b_t are special because we have a control over them and logically we will change them only in reaction to changes on the market. Mathematically speaking, it means

$$\mathrm{d}V_t = a_t \mathrm{d}X_t + b_t \mathrm{d}\beta_t \tag{4.5}$$

This is often called a self-financing condition.

4.2.2 Option pricing problem

We explain how we can determine prices of options by building so called risk less portfolios now. We will consider European call options with expiration time T and a strike price K. It means that in time T a value of this option is $(X_T - K)^+$.

We need to calculate how much money we have to invest in order to secure stocks and bonds with value $(X_T - K)^+$ in the time T. If a price of the option is $p > q$ (where q is the price we will calculate), then we can sell a call option, invest q according to investment strategy we will calculate and lock a risk-less profit $p - q$.

If we denote $V_t = u(T - t, X_t)$ then with a use of Ito lemma, we get

$$V_t - V_0 = \int_0^t (-u_1 + cX_s u_2 + \frac{1}{2}\sigma^2 X_s^2 u_{22})\mathrm{d}s + \int_0^t (\sigma X_s u_2)\mathrm{d}B_s \tag{4.6}$$

where $u_1 = \frac{\partial u}{\partial t}$ and u_2 denotes $\frac{\partial u}{\partial x}$.

Now we use the self financing property (4.5) and get

$$V_t - V_0 = \int_0^t a_s \mathrm{d}X_s + \int_0^t b_s \mathrm{d}\beta_s \tag{4.7}$$

with a use of $\mathrm{d}\beta_t = r\beta_t \mathrm{d}t$ and (4.1), we can compare integrands in (4.6) and (4.7) and we get following equations

$$a_t = u_2(T - t, X_t) \tag{4.8}$$

and partial differential equation for $u = u(x,t)$

$$u_1 = \frac{1}{2}\sigma^2 x^2 u_{22} + rxu_2 - ru \qquad for\ x > 0 \tag{4.9}$$

Notice that the constant c was eliminated in calculations and do not play role in building self financing portfolio.

The partial differential equation (4.9) can be solved analytically, which is a big advantage of Black-Scholes model. With a use of a boundary condition $u(0,x) = (x-K)^+$ we can write the solution in a form

$$u(t,x) = x\Phi(g(t,x)) - Ke^{-tr}\Phi(g(t,x) - \sigma\sqrt{t}) \tag{4.10}$$

where Φ is a cumulative distribution function of normal distribution $N(0,1)$ and

$$g(t,x) = \frac{\ln\frac{x}{K} + (r + (1/2)\sigma^2)t}{\sigma\sqrt{t}} \tag{4.11}$$

We can finally state that the right price of European call options with a strike price K and expiration time T with a value of underlying stock given by X_t is

$$V_0 = u(T, X_0) \tag{4.12}$$

where u is given by (4.10). This follows from the fact that if we invest V_0 and we use the strategy given by

$$a_t = u_2(T-t, X_t) \tag{4.13}$$

$$b_t = \frac{u(T-t,X_t) - a_t X_t}{\beta_t} \tag{4.14}$$

then the value of our portfolio in time T will be $V_T = u(0,X_T) = (X_T - K)^+$. So if value of an option is different than this estimate, we can use this strategy and lock a risk-free profit. Of course, it is a more complicated in reality. Limitations of this formula will be discussed later in this chapter.

We can derive a right price similarly for European put options. However, there is a direct connection between a price of call and put options. It is called put-call parity. Let us denote a price of a put option by $P(X_t,t)$ and a price of a call option by $C(X_t,t)$. Let us presume that both of them have same strike prices K and expiration time T, then we can consider portfolio

$$V_t = X_t + P(X_t,t) - C(X_t,t) \tag{4.15}$$

So we bought one stock and one put option and sell one call option. It is obvious that a value of this portfolio in time T is always $V_T = K$. So this portfolio guarantees a risk-free profit. The only other way to gain such a profit is to invest $Ke^{-r(T-t)}$ in time t in bonds, so we know that a value of this portfolio in time t is

$$V_t = Ke^{-r(T-t)}$$

so we have a relation between prices of put and call options

$$P(X_t,t) = C(X_t,t) - X_t + Ke^{-r(T-t)} \tag{4.16}$$

4.2.3 Change of measure

We will introduce a very useful tool of stochastic calculus - a change of underlying probability measure.

We have always considered some probability space $(\Omega, \mathcal{F})$ with given probability measure P till now. Let us consider a process

$$\widetilde{B}_t = B_t + qt \qquad t \in \langle 0, T \rangle \tag{4.17}$$

where B_t is a standard Brownian motion under probability measure P. We can ask now, if a probability measure Q, such as $\widetilde{B}_t$ is standard Brownian motion under Q, exists. An answer is given by the following Girsanov's theorem.

Theorem 4.1. *Let us consider a process*

$$M_t = \exp\left(-qB_t - \frac{1}{2}q^2 t\right) \quad t \in \langle 0, T \rangle$$

and a probability measure Q defined by

$$Q(A) = \int_A M_T(\omega) \mathrm{d}P(\omega) \quad A \in \mathcal{F} \tag{4.18}$$

then a process $\widetilde{B}_t$ defined by (4.17) is standard Brownian motion under probability measure Q. Q is called equivalent martingale measure and Q and P are in fact equivalent measures.

We remind that the measures P, Q are equivalent if

$$P(A) = 0 \iff Q(A) = 0$$

or alternatively, it can be defined with a use of Radon-Nikodym theorem. One implication follows immediately from (4.18). Further, it can be shown that M_t, as defined in the theorem (4.1), is martingale with respect to natural Brownian filtration.

Black-Scholes formula revisited

Here we will use a change of measure technique to derive Black-Scholes formula.

A naive idea how to calculate a price of European call options would be

$$V_t = e^{-r(T-t)}(X_T - K)^+ \quad t \in \langle 0, T \rangle \tag{4.19}$$

where r is an interest rate, T is expiration time and K is a strike price. This formula does not work well but we will show how it can be corrected. Let us consider a process

$$\widetilde{X}_t = e^{-rt} X_t$$

then with a use of Ito lemma, we get

$$\mathrm{d}\widetilde{X}_t = \widetilde{X}_t((c-r)\mathrm{d}t + \sigma \mathrm{d}B_t) \tag{4.20}$$

now if we denote $\widetilde{B}_t = B_t + \frac{c-r}{\sigma}t$ we get

$$\mathrm{d}\widetilde{X}_t = \sigma \widetilde{X}_t \mathrm{d}\widetilde{B}_t \tag{4.21}$$

We will denote a probability measure given by the theorem (4.1), under which $\widetilde{B}_t$ is standard Brownian motion, by Q.

Now we will consider self financing portfolio V_t given by (4.4) and portfolio $\widetilde{V}_t = e^{-rt}V_t$ then using (4.20) we get

$$\mathrm{d}\widetilde{V}_t = a_t \mathrm{d}\widetilde{X}_t \tag{4.22}$$

so by integration we get

$$\widetilde{V}_t = V_0 + \int_0^t a_s \mathrm{d}\widetilde{X}_s = V_0 + \sigma \int_0^t a_s \widetilde{X}_s \mathrm{d}\widetilde{B}_s \tag{4.23}$$

Under the measure Q is $\widetilde{B}_t$ standard Brownian motion, so $\widetilde{V}_t$ is martingale with respect to natural Brownian filtration $\mathcal{F}_t$. Obviously, a natural filtration of B_t and $\widetilde{B}_t$ is the same, so we do not have to distinguish between them. From martingale property, we have

$$\widetilde{V}_t = E_Q(\widetilde{V}_T|\mathcal{F}_t) \quad t \in \langle 0, T\rangle$$

so in conclusion we derived

$$V_t = E_Q(e^{-r(T-t)}V_T|\mathcal{F}_t) \tag{4.24}$$

This equation gives us a value of self financing portfolio in time t. It will be useful in particular, if we know a boundary condition in the form $V_T = h(X_t)$. Clearly, this is a case of risk-less portfolios we built in the last section for option pricing.

We will use this to determine a correct price of European call options. A boundary condition for them reads

$$V_T = h(X_T) = (X_T - K)^+$$

so for a price of option in time t we have

$$V_t = E_Q(e^{-r(T-t)}h(X_T)|\mathcal{F}_t) = E_Q(e^{-r(T-t)}h(X_t e^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma(\widetilde{B}_T - \widetilde{B}_t)})|\mathcal{F}_t) \tag{4.25}$$

We can use the fact that $\sigma(X_t) \subset \mathcal{F}_t$ and that $\widetilde{B}_T - \widetilde{B}_t$ is independent of $\mathcal{F}_t$ so we can treat X_t as a constant and calculate just a non-conditional expected value over $\widetilde{B}_T - \widetilde{B}_t \sim N(0, T-t)$ and we get

$$V_t = \int_{\mathbb{R}} e^{-r(T-t)} h(X_t e^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma y(T-t)^{1/2}}) \varphi(y) \mathrm{d}y \tag{4.26}$$

where $\varphi(y)$ is PDF of a distribution $N(0,1)$. Now we can use $h(x) = (x-K)^+$ and after some calculations we obtain

$$V_t = X_t \Phi(g(T-t, X_t)) - K e^{-r(T-t)} \Phi(g(T-t, X_t) - \sigma\sqrt{T-t}) \tag{4.27}$$

where $g(t,x)$ is defined by (4.11). This result clearly corresponds with the formula we obtained above.

4.3 Beyond Black-Scholes formula

In this section, we suggest some generalizations of Black-Scholes model. We do not have ambitions to go here into details. We just want to explain limitations of this model and outline ways how it can be overcome.

4.3.1 Implied volatility

Black-Scholes price depends on four parameters and a price of an underlying asset. A strike price and expiration time of an option are given by contract and an interest rate, which can be easily determined and considered as a constant. Volatility σ is a bit more complicated to determine. We can try to estimate it from a behaviour of a risky asset X_t by measuring

$$\mu = \frac{1}{N\Delta t} \sum_{k=0}^{N-1} \ln\left(\frac{X_{(k+1)\Delta t}}{X_{k\Delta t}}\right)$$

and then define it by

$$\sigma_{his}^2 = \frac{1}{N\Delta t} \sum_{k=0}^{N-1} \left(\ln\left(\frac{X_{(k+1)\Delta t}}{X_{k\Delta t}}\right) - \mu\right)^2 \tag{4.28}$$

Volatility defined in this manner is called historic volatility. It is not very often used because it fluctuates a lot in time.

An implied volatility σ_{imp} is used more frequently. This volatility is obtained by inverting Black-Scholes option pricing formula, where we use a price of option on the market as a correct option price. It means that we believe that right volatility is given by an interplay of supply and demand that determines a market price of option.

If theoretically Black-Scholes formula would be completely rigorous then $\sigma_{his} = \sigma_{imp}$ should hold. However, this is usually not true. We can actually

estimate how off Black-Scholes price is by comparing these two volatilities. A problem of σ_{imp} is that it depends on a strike price and expiration time, not just on an underlying asset. Implied volatility as a function of $X_t - K$ where K is a strike price, forms a convex function. A graph of this function is usually called volatility smile.

We should also mention that a price calculated by Black-Scholes formula works much better in situations, where the price X_t of an asset is close to a strike price K. An estimate given by this formula is getting less reliable with an increasing difference $X_t - K$. This can be easily explained - a real distribution of a price of an underlying asset is not purely Gaussian but more likely stable one. There are jumps in these stable processes, i.e. bigger price changes are much more probable than in the case of Gaussian distribution. For example, if we have a call option and a price of underlying asset is much lower than a strike price, then the Gaussian model will almost exclude a possibility that this asset will grow enough, so an option could not be exercised. Therefore such an option would be worthless according to this model. A model based on stable distributions will assign much higher probability to it and also a higher price to the option. As it turns out, a stable model is closer to reality.

We will briefly discuss generalizations of Bl.-Sch. formula beyond constant volatility. If we consider volatility σ to be known function of time, we can just take its mean value

$$\sigma^2 = \frac{1}{T}\int_0^T \sigma^2(t)\mathrm{d}t$$

More substantial generalization is to consider volatility as a stochastic process. Most usually we require volatility to follow Langevin equation

$$\mathrm{d}\sigma_t = \alpha\sigma_t\mathrm{d}t + \beta\mathrm{d}B'_t$$

and we already know that a solution of this equation is Ornstein-Uhlenbeck process.

The price evolution X_t of assets is still given by (4.1) and we consider a correlation between B_t and B'_t. This approach is called Heston model and we can proceed similarly as when we derived Bl.-Sch. formula. But a derivation here is more complicated and we will not get a solution in an analytical form so we will omit further calculations.

4.3.2 Beyond geometric Brownian motion

In this section, we look on assumptions of Bl.-Sch. model more generally.

The most important assumptions of the efficient market are liquidity and no friction. No friction is hard to arrange for a common trader but it can be close to reality for banks and other financial institutions. Liquidity is also quite real for some stocks. An efficient market hypothesis suggests then that a price evolution process should be Markov. A motivation behind a choice of Gaussian distribution is naturally CLT. Empirical observations then prefer geometric Brownian motion over Brownian motion with drift.

Problems arise when markets are heading towards crises. Underlying processes stop being Markov and underlying distributions will not be Gaussian. Assumptions of Bl.-Sch. model will not be fulfilled at all and its predictions will be naturally very far from reality. This exactly happened to the hedge fund led by Black and Scholes after few very successful years. They lost almost everything during the crisis in 1998.

Many attempts to go beyond Black-Scholes model were made. We can try to calculate a right option price with a use of stable or infinitely divisible distributions instead of geometric Brownian motion. These calculations are possible but they usually do not provide results in terms of elementary functions, which is impractical, of course. Another possibility is to use a multifractal analysis as described in the previous chapter. A multifractal spectrum can help to determine situations when Bl.-Sch. formula works well and when we need a different model. These are just few of many possible approaches to this problematic. There is a huge demand for new sophisticated models nowadays and also a huge number of problems to be solved.

Chapter 5

From thermodynamics to finance

We will start this chapter with a review of Boltzmann-Gibbs statistics that leads to standard equilibrium (extensive) thermodynamics. We will move to non-extensive thermodynamics with a use of generalize Tsallis entropy then. We will also demonstrate a mathematical correspondence between extensive and non-extensive thermodynamics given by q-calculus. We will discuss a generalize entropy as way to explain a presence of heavy tailed distributions. Finally, we will use an analogical procedure to generalize Bl.-Sch. option pricing formula. This parallel between thermodynamics and finance demonstrates how close physics and finance can be and why physical methods are useful in finance so often.

5.1 Boltzmann-Gibbs statistics

We will derive a classical Boltzmann-Gibbs distribution. It is stated in most textbooks that B.-G. statistics work when a system is in thermal equilibrium. We will discuss conditions required for B.-G. to work a bit more thoroughly at the end of this section.

5.1.1 MaxEnt principle

We will start with a brief review of concept of entropy. Entropy was first introduced in thermodynamics as a new state function that never decreases spontaneously. Modern entropy is defined axiomatically and it measures a level of uncertainty is systems. In other words, it quantifies a level of disorder in a system. Entropy nowadays has huge applications in statistical physics, information theory and in dynamical system theory.

Let us consider a probability space $(\Omega, \mathcal{F}, P)$. We will define an information function I, which we require to fulfil these natural conditions. Let $A, B \in \mathcal{F}$

1. $I(A) = 0 \Leftrightarrow P(A) = 1$
2. $P(A) \to 0 \Rightarrow I(A) \to \infty$
3. for A, B independent $I(A \cap B) = I(A) + I(B)$

From these conditions, we can define I only by a following relation

$$I(A) = -k \ln P(A) \quad k > 0 \tag{5.1}$$

where k is any constant. By comparison with results of classical thermodynamics we can determine that k is Boltzmann constant in statistical physics. Here we use $k = 1$ for simplicity.

We can define entropy of a discrete random variable X as a mean value of I. Let us assume that for any $x_i \in Ran\ X$ we have $P(X = x_i) = p_i$ then

$$S = \sum_i p_i I(X^{-1}(x_i)) = -\sum_i p_i \ln p_i \tag{5.2}$$

This is called Shannon information entropy. For continuous case we only use integration instead of a sum.

Entropy has two main uses. The first one is to calculate or measure entropy of a random variable or a physical system. This number tells us how much information random variable or given system contains. This look at entropy is closely connected with information theory. Here we will not go further into details.

The other use is more important for us - it is a maximal entropy principal. Let us consider a system and assume that we know its mean value of energy. The goal is to find PDF that maximize entropy (5.2) and also gives a right mean value of energy. This PDF contains the most information, i.e. maximize information we need to describe the system. We take this distribution as a right one for description of fluctuations of energy. This is far from a rigorous approach and the result we will obtain is often considered as a postulate. Mathematically speaking, if we denote a mean energy by E and E_i will be energy in state with probability p_i, we have to find an extreme of function S with these constraints

$$S = -\sum_i p_i \ln p_i \tag{5.3}$$

$$\sum_i p_i = 1 \tag{5.4}$$

$$\sum_i E_i p_i = E \tag{5.5}$$

we use a method of Lagrange multipliers and construct a function

$$\Lambda = -\sum_i p_i \ln p_i + \alpha(\sum_i p_i - 1) + \beta(\sum_i E_i p_i - E) \tag{5.6}$$

Finding an extreme of Λ and using conditions (5.4) yields

$$p_i = \frac{1}{Z(\beta)} \exp\left(-\beta E_i\right) \tag{5.7}$$

where Z is a normalization constant

$$Z(\beta) = \sum_i \exp\left(-\beta E_i\right) \tag{5.8}$$

a multiplier β could be theoretically calculated with a use of (5.5) but in general, it is not analytically possible. However, in statistical physics, we can derive by comparison with classical thermodynamics $\beta = \frac{1}{k_B T}$.

If we consider energy of a free particle $E_i = \frac{1}{2} m v_i^2$ then with a use of the theorem (1.1), we get Maxwell-Boltzmann distribution for velocity v_i. So we obtained Gaussian distribution. It is no coincidence, a connection between CLT and Shannon entropy is very deep. In fact, one of proofs of CLT is based on Shannon entropy.

5.1.2 Limitations of Boltzmann-Gibbs statistics

Here we will briefly discuss conditions required for Boltzmann-Gibbs statistics to work well and typical examples of systems where it does not work at all.

To determine exact conditions, where B.-G. statistics can be used, is very complicated and still an open issue. Ergodicity is a very important concept in this problematic. Simply said, systems, whose time averages coincide with space averages, are ergodic. A mathematical formulation of ergodicity is rather complicated, so we will omit it here. An important fact is that ergodicity naturally leads to B.-G. statistics. A problem is that ergodicity is almost impossible to verify in real systems.

Anyway B.-G. statistics work well very often but systems with properties like long-range particle interactions, long-term memory or (multi)fractal behaviour, cannot usually be well described in framework of B.-G. statistics. Financial markets are the most important example of such systems for us, of course.

5.2 Non extensive statistical mechanics

In this section, we will discuss Tsallis generalize entropy, which leads to non extensive statistical mechanics. We will use MaxEnt principle on it and obtain heavy tailed distributions, which obviously play an important role in finance.

5.2.1 Tsallis entropy

There are many generalizations of entropy. We will discuss only one of them - Tsallis entropy. Tsallis entropy is non additive, i.e. we do not require information function to fulfil the property 3.. This is a reason, why we use a term non-extensive statistical mechanics - entropy of two independent systems

is not a sum of individual entropies. This sounds strange but in open systems (like financial markets or in quantum mechanics), a notion of independence is complicated and the condition 3. is not necessarily reasonable.

We will define informational function I by

$$I_q(A) = \frac{1}{q-1}(1 - P(A)^{q-1}) \tag{5.9}$$

Notice that for $q \to 1$ we obtain informational function $I_1 = I$ that leads to Shannon entropy. Further condition $P(A) \to 0 \Rightarrow I_q(A) \to \infty$ does not hold for $q > 1$. However, it is important that $I(A)$ increases with $P(A) \to 0$. Instead of additivity, we get following property for independent $A, B \in \mathcal{F}$

$$I_q(A \cap B) = I_q(A) + I_q(B) + (1-q)I_q(A)I_q(B) \tag{5.10}$$

We will consider a discrete random variable X. If we denote $P(X = x_i) = p_i$, then we get Tsallis entropy of X analogically as in previous section

$$S_q = \frac{1}{q-1}(1 - \sum_i p_i^q) \quad q \in \mathbb{R} - \{0\} \tag{5.11}$$

where we used $\sum_i p_i = 1$.

Problematic is that Tsallis entropy does not have a solid background in information theory as Shannon entropy. This means that calculating Tsallis entropy for a concrete random variable yields a number but we do not know how to interpret it. So the only way to use Tsallis entropy is MaxEnt principal. However, this principal will give us a very important class of distributions.

We can consider q as a biasing parameter, $q > 1$ privileges common events, $q < 1$ privileges rare events. So B.-G. ($q = 1$) is unbiased statistics.

We will use MaxEnt principal. We can calculate constraint extremes as in the last section and we get following the most probable distribution

$$p_i = \frac{1}{Z_q}(1 + (q-1)\beta_q E_i)^{1/(1-q)} \tag{5.12}$$

where Z_q is a normalization constant given by

$$Z_q = \sum_i (1 + (q-1)\beta_q E_i)^{1/(1-q)} \tag{5.13}$$

Clearly for $q = 1$ we obtain B.-G. distribution. We have power law behaviour $p_n \sim E_n^{1/(1-q)}$ for other $q > 1$. Meaning of inverse temperature β_q is not very well understood with an exception of $\beta_1 = \beta = 1/k_B T$. We have obtained heavy tailed distributions, which coincide with empirical observations in finance and in many other complex dynamical systems. However, it is still not very clear, if this way to obtain them is really fundamental or if it is more likely coincidence.

5.2.2 Quantum calculus

Another interesting interpretation of Tsallis entropy is given via quantum calculus also known as q-calculus.

Q-calculus is "calculus without limits". Instead of it, q-calculus introduces a new parameter q. For example, q-derivation is given by

$$D_q(f(x)) = \frac{f(qx) - f(x)}{qx - x} \tag{5.14}$$

Notice that in a limit $q \to 1$ we get $D_1(f(x)) = f'(x)$. This is characteristic for q-calculus. Q-calculus corresponds in limit $q \to 1$ with classical calculus. In fact, we can build corresponding q-version to almost every field of mathematics. This is usually very difficult and quite technical, so we will not go into details. Anyway, we will define q-versions of basic operations and functions. We will also show that Tsallis entropy is q-deformation of Shannon entropy. We will outline the notion of q-dependent random variables and q-version of CLT.

Let us start with definition of q-exponential and q-logarithm

$$exp_q(x) = [1 + (1-q)x]_+^{\frac{1}{1-q}} \tag{5.15}$$

$$ln_q(x) = \frac{x^{1-q} - 1}{1-q} \tag{5.16}$$

where $[x]_+ = x$ for $x \geq 0$ and $[x]_+ = 0$ for $x \leq 0$.
Again, (5.15) and (5.16) are normal exponential respectively logarithm in limit $q \to 1$. Naturally, q-exponential and q-logarithm are inverse functions. We will define q-sum, subtraction, product and division

$$x \oplus_q y = x + y + (1-q)xy$$

$$x \ominus_q y = \frac{x - y}{1 + (1-q)y}$$

$$x \otimes_q y = [x^{1-q} + y^{1-q} - 1]_+^{\frac{1}{1-q}}$$

$$x \oslash_q y = (x^{1-q} - y^{1-q} - 1)^{\frac{1}{1-q}}$$

It can be easily shown that q-sum and q-product are commutative and associative. All these operations recover a usual form for $q = 1$, also $x \oplus_q 0 = x$ and $x \otimes_q 1 = x$ holds. Further, q-subtraction/division was naturally derived as an inverse operation to q-sum/product.

These operations are defined in the way that following relations would hold.

$$exp_q(x \oplus_q y) = exp_q(x)exp_q(y) \qquad exp_q(x + y) = exp_q(x) \otimes exp_q(y)$$

$$ln_q(x \otimes_q y) = ln_q(x) + ln_q(y) \qquad ln_q(xy) = ln_q(x) \oplus ln_q(y)$$

Further, we can find Taylor expansion of these functions and exp_q can be analytically extended to a complex plane. We will not need these results and we omit them here.

We can demonstrate a role of Tsallis entropy as q-analogue of Shannon entropy now. Indeed

$$S_q = \frac{1}{q-1}(1 - \sum_i p_i^q) = \sum_i p_i ln_q(1/p_i) \tag{5.17}$$

this corresponds with Shannon entropy $S = \sum_i p_i ln(1/p_i)$ but because $ln_q(1/x) \neq -ln_q(x)$ we can see that we can define q-analogue to Shannon entropy in more then one way.

We can notice another analogies between Tsallis entropy and q-calculus. Pseudo-additivity of Tsallis entropy given by the relation (5.10) obviously corresponds with our definition of q-sum. Also the most probable distribution derived via maximization of q-entropy has form

$$p_i = \frac{1}{Z_q}(1 + (q-1)\beta_q E_i)^{1/(1-q)} = \frac{1}{Z_q} exp_q(-\beta_q E_i) \tag{5.18}$$

this distribution is called q-Gaussian, so now we can state that q-entropy leads to q-Gaussian distribution, which is heavy tailed with an exception of the case $q = 1$, where we get Shannon entropy and normal distribution.

We can proceed to q-dependence of random variables to q-CLT and to q-stable distributions. Following theory is very technical, so we omit a rigorous mathematical formulation of some statements. We also have no ambitions to go into details here.

A very important tool for establishing q-analogues in probability theory is q-Fourier transform.

Definition 5.1. *For $q \geq 0$ we define q-Fourier transform by*

$$\mathcal{F}_q[f](\xi) = \int_{suppf} exp_q(ix\xi) \otimes_q f(x)\mathrm{d}x$$

Remark. *Notice that $exp_q(ix\xi) \otimes_q f(x)$ is not zero outside of support of f, as it is in the case of normal product.*

If f is PDF, then we will call $\mathcal{F}_q[f]$ its q-characteristic function. We will introduce a notion of q-independence now.

Definition 5.2. *Random variables X, Y are (q', q, q'') independent, if for their PDF p_X,p_Y following holds*

$$\mathcal{F}_{q'}[p_{X+Y}] = \mathcal{F}_q[p_X] \otimes_{q''} \mathcal{F}_q[p_Y]$$

If $q = q' = q''$, we call random variables just q-independent. It can be easily seen that for $q = q' = q'' = 1$ this definition is equivalent to a normal definition od independent variables.

Before stating q-CLT, we also need to define q-convergence.

Definition 5.3. *A sequence of random variables X_n is q-convergent to a random variable X, if*

$$\lim_{N\to\infty} \mathcal{F}_q[X_n](\xi) = \mathcal{F}_q[X](\xi) \quad \forall \xi \in \mathbb{R}$$

Remark. *This is equivalent to convergence in distribution For $q = 1$, because it is well known that*

$$X_n \xrightarrow{\mathcal{D}} X \iff \varphi_{X_n}(y) \to \varphi_X(y) \quad \forall y \in \mathbb{R}$$

where φ as usually denotes characteristic function.

We will define q-moments. Let $p(x)$ be PDF then corresponding q-PDF has a form

$$p_q(x) = \frac{|p(x)|^q}{\nu_q(p)} \tag{5.19}$$

where ν_q is normalization given by

$$\nu_q(p) = \int_{-\infty}^{\infty} |p(x)|^q \mathrm{d}x < \infty \tag{5.20}$$

where we have to assume a finiteness of ν_q. We can define q-mean by

$$\mu_q = \int_{-\infty}^{\infty} x p_q(x) \mathrm{d}x \tag{5.21}$$

and q-variance by

$$Var_q(x) = \int_{-\infty}^{\infty} (x - \mu_q)^2 p_q(x) \mathrm{d}x \tag{5.22}$$

and we can define every other q-moment in the same manner.

We will state q-CLT. It can be stated in more general form but we would need much more preliminary results for it.

Theorem 5.1. *Let X_n be a sequence of identically distributed and q-independent random variables with $1 \leq q < 2$. Let X_n have a finite μ_q and Var_{2q-1}. We define a sequence*

$$S_N = \frac{\sum_{n=1}^{N} X_n - N\mu_q}{C_q(N)}$$

where $C_q(N)$ is proper normalization. Then S_N is q-convergent to $\widetilde{q}$-Gaussian distribution , where $\widetilde{q}$ is given by q.

Remark. *Even though this theorem gives a very interesting possibility to generalize limit theorems even for strongly correlated variables, its applicability is questionable. This type of correlation was not observed in any real system so far.*

We will also briefly introduce (q, α) stable symmetric distributions.

Definition 5.4. *Random variable X is said to be (q,α)- stable, if its q-characteristic function has a form*

$$\varphi_q(x) = exp_q(-\beta|x|^\alpha)$$

where $\beta > 0$, $q \in \mathbb{R}$ and $\alpha \in (0,2\rangle$.
A class of these distributions is usually denoted by $\mathcal{L}_q[\alpha]$

Remark. *From the theorem (1.4), we can see that these distributions really are just stable symmetric distributions with peak in 0 for $q = 1$. These distributions were recently used for generalized q-CLT (only for $q > 1$). We will not go into details here, for details see [13]. Anyway, many important questions are still to be answered in this area of q-calculus.*

We will end this section with a following theorem.

Theorem 5.2. *Let $X_j \in \mathcal{L}_q[\alpha]$ be q-independent random variables and let a_i be arbitrary real constants, then*

$$\sum_{i=1}^{n} a_i X_i \in \mathcal{L}_q[\alpha]$$

Proof. Let $\mathcal{F}_q[p_{X_j}](\xi) = exp_q(-\beta_j|\xi|^\alpha)$ where p_{X_j} is PDF of a random variable X_j. Now with a use of $\mathcal{F}_q[aX](\xi) = \mathcal{F}_q[X](a^{2-q}\xi)$ we obtain $\mathcal{F}_q[\sum_{i=1}^n a_i X_i](\xi) = exp_q(-\beta|\xi|^\alpha)$ where $\beta = \sum_{i=1}^n \beta_j |a|^{\alpha(2-q)}$. □

So a sum of q-independent (q,α)-stable random variables is (q,α)-stable.

5.3 Non-Gaussian option pricing

In this section, we will derive a generalization of Bl.-Sch. option pricing formula in the framework of Tsallis generalized thermodynamics. We will assume that underlying noise follows q-generalized Brownian motion. We will derive analogous formulas as in the chapter about ordinary Bl.-Sch. option pricing.

Let S_t be a price of a risky asset, we will consider a process

$$Y_t = \ln \frac{S_{t+t_0}}{S_{t_0}} \tag{5.23}$$

now we will assume that Y is given by

$$\mathrm{d}Y_t = \mu \mathrm{d}t + \sigma \mathrm{d}\Omega_t \tag{5.24}$$

where a process Ω_t with PDF $P(x,t)$ is given by

$$\mathrm{d}\Omega_t = P(\Omega_t, t)^{\frac{1-q}{2}} \mathrm{d}B_t \tag{5.25}$$

Ω_t is called a statistical feedback process and it connects a macroscopic level given by P and a microscopic level represented by Ω. It is just standard Brownian motion for $q = 1$.

This definition of Ω_t also leads to non-linear Fokker-Planck equation

$$\frac{\partial}{\partial t}P(\Omega,t|\Omega',t') = \frac{1}{2}\frac{\partial^2}{\partial\Omega^2}P^{2-q}(\Omega,t|\Omega',t') \tag{5.26}$$

A solution of this equation can be found only for unconditional PDF $P(\Omega,t)$. Then we can replace Ω by $\Omega-\Omega'$ and t by $t-t'$. It follows from calculations we made on page 49 when deriving that every process with independent increments is also Markov. A solution is given by

$$P_q(\Omega,t|\Omega',t') = \frac{1}{Z(t-t')}(1-\beta(t-t')(1-q)(\Omega-\Omega')^2)^{\frac{1}{1-q}} \tag{5.27}$$

Notice that this is equivalent to

$$P_q(\Omega,t|\Omega',t') = \frac{1}{Z(t-t')}\exp_q(-\beta(t-t')(\Omega-\Omega')^2)$$

q-dependent constants Z,β are chosen so we would get usual autocorrelation $E(\mathrm{d}\Omega_t\mathrm{d}\Omega_{t'}) = \delta(t-t')\mathrm{d}t\mathrm{d}t'$. Obviously, this holds for normal Brownian motion. A concrete form of these constants is not important for us. It can be found in [11] .

We will also consider only a certain range of the parameter q. We want Ω_t to be heavy tailed distribution, so we want $q>1$. Otherwise, P_q can have only a finite support. Further, we want Ω_t to have a finite variance, which gives the range of $1<q<5/3$. As usually, standard model is recovered for $q=1$.

With use of (5.23) and (5.24), we can rewrite P_q in terms of an asset price S_t

$$P_q(\ln(S_{t+t_0})|\ln(S_{t_0})) = \frac{1}{Z(t-t_0)}(1-\widetilde{\beta}(t-t_0)(1-q)(\ln\frac{S_{t+t_0}}{S_{t_0}}-\mu t)^2)^{\frac{1}{1-q}} \tag{5.28}$$

where $\widetilde{\beta} = \frac{\beta}{\sigma^2}$.

From now on, we will presume $t_0=0$ and $\Omega_0=0$ for simplicity and without loss of generality. This choice makes Ω generalized Brownian motion.

By (5.24) is a given form of $\mathrm{d}\ln S_t$ from that we can derive

$$\mathrm{d}S_t = (\mu+\frac{\sigma^2}{2}P_q^{1-q}(\Omega_t,t))S_t\mathrm{d}t + \sigma S_t\mathrm{d}\Omega_t := \widetilde{\mu}S_t\mathrm{d}t + \sigma S_t\mathrm{d}\Omega_t \tag{5.29}$$

This can be easily verified by calculating $\mathrm{d}\ln S_t$ with a use of Ito lemma.

Notice that we started with the equation (5.29) instead of (5.24) as in the chapter about standard Bl.-Sch. formula. Equivalence of these two approaches is clear from the fact that the parameter $\widetilde{\mu}$ (called rate of return) will not appear here as well as in normal model in final formula.

We will derive a generalization of Bl.-Sch. equation (4.9). We will use a little short cut in this derivation. Anyway, we will derive similar formula more rigorously later with use of equivalent martingale measures and we will show equivalence between these two approaches.

Let us consider a derivative with a price given by function $u = u(t, S_t)$ then with a use of Ito lemma we get

$$\mathrm{d}u = u_2 \mathrm{d}S + (u_1 + \frac{1}{2}\sigma^2 S^2 P_q^{1-q} u_{22})\mathrm{d}t \tag{5.30}$$

where u_i denotes partial derivations.
In the previous chapter, we derived that a price u can be "modelled" by risk-free portfolio consisting of bonds and shares S_t. We also derived that the right number of shares we need to hold for one derivative is given by $u_2(t, S_t)$, so portfolio

$$\Pi = u - u_2 S \tag{5.31}$$

should be risk-free with interest rate r. For little changes of Π we have

$$\Delta\Pi = \Delta u - u_2 \Delta S = (u_1 + \frac{1}{2}\sigma^2 S^2 P_q^{1-q} u_{22})\Delta t \tag{5.32}$$

but Π is also risk-free. Following relation analogous to properties of bonds given by (4.3) holds

$$\Delta\Pi = r\Pi\Delta t \tag{5.33}$$

combined together, we can get generalized Black-Scholes equation

$$u_1 + \frac{1}{2}\sigma^2 S^2 P_q^{1-q} u_{22} = r(u - u_2 S) \tag{5.34}$$

It can be easily verified that we obtain classical Bl.-Sch. equation (4.9) in limit $q \to 1$.

We will try to derive a price $u(t, S_t)$ with a use of an equivalent martingale measure. In other words, we try to find a measure Q, under which a discounted price

$$G_t = e^{-rt} S_t \tag{5.35}$$

is martingale. We will need a more general version of theorem (4.1) for that. If we have a process

$$X_t = q(t, X_t)\mathrm{d}t + \mathrm{d}B_t \tag{5.36}$$

and we assume

$$\int_0^T q^2(t, X_t)\mathrm{d}t < \infty \tag{5.37}$$

then equivalent martingale measure Q , under which X_t is standard Brownian motion, exists. Q is given by

$$Q(A) = \int_A M_T(\omega)\mathrm{d}P(\omega) \tag{5.38}$$

where

$$M_T(\omega) = \exp\left(-\int_0^T q\mathrm{d}B_t - \frac{1}{2}\int_0^T q^2\mathrm{d}t\right) \tag{5.39}$$

We can proceed with deriving option pricing formula in similar manner as in the previous chapter. With a use of Ito lemma, we can derive from (5.29)

$$\mathrm{d}G_t = (\widetilde{\mu} - r)G\mathrm{d}t + \sigma G\mathrm{d}\Omega \tag{5.40}$$

We already know that a solution of this equation has the form

$$G_T = G_0 \exp\left((\widetilde{\mu} - r)T - \frac{\sigma^2}{2}\int_0^T P_q^{1-q}\mathrm{d}t + \sigma\int_0^T P_q^{\frac{1-q}{2}}\mathrm{d}B_t\right) \tag{5.41}$$

The relation for an original price S_t can be obtained with use of definition of $\widetilde{\mu}$ given by (5.29)

$$S_T = \exp\left(\mu T + \sigma\int_0^T P_q^{\frac{1-q}{2}}\mathrm{d}B_t\right) \tag{5.42}$$

We need to find a measure Q, under which is G_t martingale, so if we rewrite (5.40)

$$\mathrm{d}G_t = \sigma G P_q^{\frac{1-q}{2}}\mathrm{d}\widetilde{B}_t \tag{5.43}$$

where

$$\mathrm{d}\widetilde{B}_t = \left(\frac{\widetilde{\mu} - r}{\sigma P_q^{\frac{1-q}{2}}}\right)\mathrm{d}t + \mathrm{d}B_t \tag{5.44}$$

we want $\widetilde{B}_t$ to be standard Brownian motion under Q. P_q is non-vanishing bounded function of Ω_t so we know that such Q exists and is given by (5.38, 5.39) .

Now we need a form of price S analogical to (5.42) but under a new measure Q. We can easily see from (5.43)

$$\mathrm{d}\ln G = -\frac{\sigma^2}{2}P_q^{1-q}\mathrm{d}t + \sigma P_q^{\frac{1-q}{2}}\mathrm{d}\widetilde{B}_t \tag{5.45}$$

from this we can find a form of G_t under Q by stochastic integration. Because $S_t = e^{rt}G_t$ we get

$$S_T = S_0 \exp\left(\sigma\int_0^T P_q^{\frac{1-q}{2}}\mathrm{d}\widetilde{B}_t + \int_0^T (r - \frac{\sigma^2}{2}P_q^{1-q})\mathrm{d}t\right) \tag{5.46}$$

or equivalently

$$\mathrm{d}S_t = rS\mathrm{d}t + \sigma S P_q^{\frac{1-q}{2}}\mathrm{d}\widetilde{B}_t \tag{5.47}$$

The difference in form of S_t under original measure and under Q is only in replacement of a time dependent rate of return $\widetilde{\mu}$ by an interest rate r.

To calculate integrals in (5.46), we have to realize first that under Q following holds

$$\Omega_T = \int_0^T P_q^{\frac{1-q}{2}}\mathrm{d}\widetilde{B}_t \tag{5.48}$$

further from the theorem (1.1) and the form of P_q we obtain

$$\Omega_t \stackrel{d}{=} \sqrt{\frac{\beta(T)}{\beta(t)}}\Omega_T \tag{5.49}$$

where $\stackrel{d}{=}$ denotes equality of distributions.
We will use $\int_0^T \frac{1}{Z(t)}\mathrm{d}t = \alpha T^{\frac{2}{3-q}}$, where α is q-dependent constant (its form can be found in [11]). We get following equality now

$$S_T = S_0 \exp\left(\sigma\Omega_T + rT - \sigma^2\alpha T^{\frac{2}{3-q}}(1-(1-q)\beta(T)\Omega_T^2)\right) \tag{5.50}$$

Let us assume that V_t is a price of a derivative of an underlying asset S_t. Let us assume further that $V_T = h(S_T)$. This is obviously case of European call option, where $h(S_T) = (S_T - K)^+$ where K is a strike price.

We will use conditional probability in similar manner as in the previous chapter to derive the correct formula. We will work with natural filtration of generalized Brownian motion I_t and use the fact that $\Omega_T = \Omega_T - \Omega_0$ is independent of Ω_0. From martingale property follows

$$V_0 = E_Q(e^{-rT}h(S_T)|I_0) \tag{5.51}$$

With a use of (5.50) and independence of Ω_T and I_0, we can calculate only a normal mean value of function $h(S_T)$ and we obtain

$$V_0 = \frac{e^{-rT}}{Z(T)}\int_{\mathbb{R}} h(S_T)(1-\beta(T)(1-q)(\Omega_T)^2)^{\frac{1}{1-q}}\mathrm{d}\Omega_T \tag{5.52}$$

where S_T is given by (5.50).

Notice that in the previous chapter, we derived this formula more generally for V_t at any time. It would be possible here too but we would have to integrate over (t,T) in (5.46) so resulting integrals would be more complicated and dependent on Ω_T and Ω_t. We would use independence of $\Omega_T - \Omega_t$ with Ω_t and the fact that Ω_t does not obtain more information than I_t. We could continue similarly as we did above then.

A closed form solution of V_0 for European call option (where $h(S_T) = (S_T - K)^+$) can be analytically found, see [11]. However, we usually cannot obtain a closed form solution of (5.52) for more complicated derivatives. So we have to calculate it numerically or to try to solve generalized Bl.-Sch. equation (5.34) with given boundary conditions. We will prove equivalence of these two approaches in the next paragraph.

We will start with a solution we have obtained via equivalent martingale measure. We will show that the solution also fulfils generalized Bl.-Sch. equation.
We define a function $v(t,x)$ by

$$v(t,S_t) = E_Q(h(S_T)|I_t) \tag{5.53}$$

Clearly $v(t, S_t)$ is martingale under Q with respect to filtration $I_t,\ 0 \leq t \leq T$. A form of $\mathrm{d}S_t$ under Q is given by (5.47), so we can use Ito lemma and get

$$\mathrm{d}v(t, S_t) = (v_1 + rS_t v_2 + \frac{1}{2}\sigma^2 S_t^2 P_q^{1-q} v_{22})\mathrm{d}t + \sigma S P_q^{\frac{1-q}{2}} v_2 \mathrm{d}\widetilde{B}_t \tag{5.54}$$

because $v(t, S_t)$ is martingale following must hold

$$v_1 + rS_t v_2 + \frac{1}{2}\sigma^2 S_t^2 P_q^{1-q} v_{22} = 0 \tag{5.55}$$

if we denote a price of a derivative $u(t, S_t)$, we get

$$u(t, S_t) = E_Q(e^{-r(T-t)} h(S_T)|I_t) = e^{-r(T-t)} v(t, S_t) \tag{5.56}$$

if we insert $v(t, S_t) = e^{r(T-t)} u(t, S_t)$ into (5.55), we will obtain Bl.-Sch. generalized equation (5.34). So we have proved desired equivalence.

Above derived formulas give very good predictions for values of q around $q = 1.5$. They seem to be much more reliable then results achieved by classical Bl.-Sch. option pricing. This coincides with the fact that heavy tailed distributions are better for describing financial markets then Gaussian distribution. On the other hand, it is theoretically quite unclear, if derivations of these formulas really follow the "right direction". The obtained solution corresponds with empirical observations but it might be caused just by a correct assumption to start with a heavy tailed distributions.

In conclusion, we derived formulas with a real-world applicability with a use of advanced mathematics and concepts from statistical physics. This should be an enough motivation to study q-calculus and other related fields of mathematics in order to improve or generalize these results. Also finding other parallels between physics and finance should be very promising.

Appendix A

We will state an important theorem for probability theory here - Radon-Nykodum theorem. It will give us an equivalent definition of equivalent measures. Two measures μ, ν on same measurable space are equivalent if

$$\mu(A) = 0 \iff \nu(A) = 0$$

we write $(\mu >> \nu) \Leftrightarrow (\mu(A) = 0 \Rightarrow \nu(A) = 0)$, we also need following definition.

Definition A.1. *A measure μ on measurable space $(X, \mathcal{F})$ is σ-finite if X is a countable union of measurable sets with finite measure μ.*

Clearly any probability measure and Lebesque's measure are σ-finite.

Theorem A.1. *For σ-finite measures μ, ν on the same measurable space $(X, \mathcal{F})$ holds $\mu << \nu$ if and only if there exists measurable non negative function f fulfilling*

$$\mu(A) = \int_A f(\omega) \mathrm{d}\nu(\omega) \qquad A \in \mathcal{F}$$

further f is given uniquely with an exception of sets with ν measure 0.

If $\mu = P \circ X^{-1}$ where X is random variable and P probability measure and ν is Lebesque's measure then f is just PDF as we know it.

Appendix B

We briefly introduce main modes of convergence of random variables that we used throughout this thesis.

Convergence in distribution

A sequence of random variables (A_n) converges in distribution (converges weakly) to A, if for all bounded and continuous functions f following holds

$$\lim_{n\to\infty} Ef(A_n) = Ef(A)$$

convergence in distribution is equivalent to

$$\lim_{n\to\infty} F_{A_n}(x) = F_A(x) \quad \forall x \in D$$

where F denotes cumulative distribution and D set of all continuity points of F_A.

Also equivalent to $A_n \xrightarrow{\mathcal{D}} A$ is

$$\lim_{n\to\infty} \varphi_{A_n}(x) = \varphi_A(x) \quad \forall x \in \mathbb{R}$$

where φ denotes characteristic function.

Convergence in probability

A sequence A_n converges to A in probability ($A_n \xrightarrow{P} A$) if

$$\lim_{n\to\infty} P(|A - A_n| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

An important fact is that convergence in probability implies convergence in distribution.

L^p convergence

Let $p > 0$ then we say that sequence A_n converges to A in L^p if all A_n and A are L^p integrable variables and

$$\lim_{n\to\infty} E|A - A_n|^p = 0$$

Further L^p convergence imply convergence in probability. L^2 convergence is usually called mean square convergence.

Appendix C

We will show main steps of construction of stochastic integral in this appendix. We will not go into all details and will not prove most of the statements here. For a completely rigorous construction see [9].

As in the case of Daniell's construction of integral, we will start with simple processes - an analogue of step functions from non-random case.

Definition C.1. *Process $C = (C_t, t \in \langle 0, T\rangle)$ is called a simple process, if there exists partition $0 = t_0 < ... < t_n = T$ such as*

$$C_t = Z_i \quad for \quad t_{i-1} \leq t < t_i$$

where Z_i are adapted to Brownian filtration $\mathcal{F}_{t_{i-1}}$ and $EZ_i^2 < \infty$

So for a simple process we can easily calculate

$$I_t(C) = \int_0^t C_s \mathrm{d}B_s = \sum_{i=1}^{k-1} Z_i(B_{t_i} - B_{t_{i-1}}) + Z_k(B_t - B_{t_{k-1}}) \tag{C.1}$$

for $t \in (t_{k-1}, t_k\rangle$.

We will move to properties of stochastic integral for simple processes now. Of course, as in case of general Ito stochastic integral, following properties hold

1. $I_t(aC_1 + C_2) = aI_t(C_1) + I_t(C_2) \quad a \in \mathbb{R}$
2. $I_t(C)$ is martingale with respect to natural filtration of Brownian motion
3. $EI_t(C) = 0$
4. $E(I_t(C))^2 = \int_0^t EC_s^2 \mathrm{d}s$
5. I_t has continuous trajectories

The properties (1), (5) are in this case obvious. The property (3) follows directly from property (2) because $EI_0(C) = 0$. We will prove that $I_t(C)$ is martingale with respect to Brownian filtration. So we have to prove three things that $I_t(C)$ is adapted to Brownian filtration, which is obvious. Then we have to prove

$$E|I_t(C)| < \infty$$

which follows from the property (4) because $E(I_t(C))^2 < \infty$ implies $E|I_t(C)| < \infty$. The last property defining martingale reads

$$E(I_t(C)|\mathcal{F}_s) = I_s(C) \quad for \quad t > s \tag{C.2}$$

we will prove this for $s, t \in \langle t_{k-1}, t_k \rangle$ but a proof for any t, s is exactly the same just bit more lengthy. It follows from calculations

$$I_t(C) = I_s(C) + Z_k(B_t - B_s) \tag{C.3}$$

so

$$E(I_t(C)|\mathcal{F}_s) = I_s(C) + Z_k E(B_t - B_s) = I_s(C) \tag{C.4}$$

We will prove the property (4). We can assume $t = t_k$ without loss of generality. We denote $\Delta B_i = B_{t_i} - B_{t_{i-1}}$ then by direct calculations we can get

$$E(I_t(C))^2 = \sum_{i=1}^{k}\sum_{j=1}^{k} E(Z_i \Delta B_i Z_j \Delta B_j) = \sum_{i=1}^{k} E(Z_i \Delta B_i)^2 = \sum_{i=1}^{k} E Z_i^2 (t_i - t_{i-1}) \tag{C.5}$$

in the last term we can recognize Riemann sum so we have

$$E(I_t(C))^2 = \int_0^t E C_s^2 \mathrm{d}s \tag{C.6}$$

We have proved all desired properties of stochastic integral for simple processes now. The most complicated part is to generalize Ito integral for much larger class of integrable processes. We will only outline how it can be done.

We want to integrate processes $C_t, t \in \langle 0, T \rangle$ fulfilling conditions

1. C is adapted to Brownian motion
2. $\int_0^t E C_s^2 \mathrm{d}s < \infty$

A crucial point of construction is to prove that for every such process C sequence of simple processes $C^{(n)}$ exists that following holds

$$\int_0^T E|C_s - C_s^{(n)}|^2 \mathrm{d}s \to 0 \quad n \to \infty \tag{C.7}$$

In other words, simple processes are dense in space $L^2[\Omega \times \langle 0, T \rangle, \mathrm{d}P \times \mathrm{d}t]$. Further, it can be proven that $I_t(C^{(n)})$ converges uniformly to $I_t(C) = \lim_{n\to\infty} I_t(C^{(n)})$, which implies

$$E \sup_{0 \le t \le T} |I_t(C) - I_t(C^{(n)})|^2 \to 0 \quad n \to \infty \tag{C.8}$$

It follows from this that general stochastic integral $I_t(C)$ is correctly defined for every C fulfilling required conditions, because (C.7) implies that its value does not depend on a choice of $C^{(n)}$. Continuity of sample path is also verified thanks to uniform convergence. All other properties, which we derived for simple processes, can be proven for general stochastic integral similarly by limit transitions .

Bibliography

[1] W.Paul, J. Baschnagel: Stochastic Processes: From physics to finance. Springer, 2000.

[2] K. Falconer: Fractal geometry, Mathematical foundations and applications. Wiley, 1989.

[3] J.-P. Bouchard, M. Potters: Theory of financial risks, From statistical physics to risk management. CUP, Cambridge, 2001.

[4] B. V. Gnedenko, A. N. Kolmogorov: Limit distributions for sums of independent random variables. Adison-Wesley, 1968

[5] R. N. Mantegna, H. E. Stanley: An introduction to econophysics. CUP, Cambridge, 2000.

[6] T. Mikosch: Elementary stochastic calculus with finance in view. World scientific, 1998.

[7] H.Risken: The Fokker-Planck equation. Springer, 1988.

[8] J. C. Hull: Options, futures and other derivatives. Prentice Hal, Upper Saddle River, 1997.

[9] B.Oksendal: Stochastic differential equations. Springer, 1994.

[10] B.B. Mandelbrot: Fractals and scaling in finance. SELECTA VOLUME E, 1996.

[11] L. Borland: Option pricing formulas based on a non-Gaussian stock price model. Physical review letters, 2002.

[12] C. Tsallis, C. Anteneodo, L. Borland, R. Osorio: Non-extensive statistical mechanics and economics. Physica A, 2003.

[13] S.Umarov, C. Tsallis, M. Gell-Mann, S. Steinberg: Generalization of symmetric α-stable distributions for $q > 1$. APS/123-QED, 2010.

[14] A. E. Kyprianou: Levy processes and continuous-state processes. University of Bath, Claverton Down, 2000.

[15] L. Borland: Microscopic dynamics of non-linear Fokker-Planck equation. Physical review E, 1998.

[16] S.Umarov, C. Tsallis, S. Steinberg: On a q-central limit theorem consistent with nonextensive statistical mechanics. Milan journal of mathematics, 2008.

[17] J.Korbel: Application of generalized statistics in econophysics, 2010.