Mathematical Engineering

Sergey S. Stepanov

Stochastic World

Springer

Mathematical Engineering

Series Editors

For further volumes:
http://www.springer.com/series/8445

Sergey S. Stepanov

Stochastic World

Springer

Sergey S. Stepanov
Dneropetrovsk
Ukraine

ISSN 2192-4732 ISSN 2192-4740 (electronic)
ISBN 978-3-319-00070-1 ISBN 978-3-319-00071-8 (eBook)
DOI 10.1007/978-3-319-00071-8
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013939583

Printed on acid-free paper

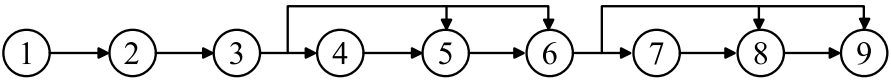
Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book is based on the expanded lecture notes formed during the statistics course for the employees of the company "Altus Assets Activities", which was organized by the Center of Fundamental Research. The main aim of the lectures was to give a quick and simple introduction to the stochastic differential equations, at the same time keeping the argumentation conclusive.

Stochastic processes appear in different financial, biological and physical systems. The corresponding mathematical approach, even though it deals with highly non-trivial entities, is quite simple. We believe that, for the practical applications, the informal understanding of the practical methods is more important on the initial stage than their strict axiomatic justification. Unlike most common methodology we will only occasionally use stochastic integration. This will simplify the reasoning significantly, and we will be able to proceed to the practical applications directly.

The recommended way of studying the material by chapters can be presented with a following diagram:



The first six chapters form the core of the book and cover the basics of the stochastic math. The seventh and eighth chapters (as well as their individual sections) are dedicated to the applications and can be read in any order. In the ninth chapter the numerical computer modeling of stochastic processes is considered and it would help if the reader has basic knowledge of some programming language.

Throughout the text small problems are scattered; they are denoted with the character ($\lessdot \mathrm{H}_i$), where i is the number of the solution in the Appendix "*Help*". There are also links, marked by ($\lessdot \mathrm{C}_i$), which should be followed *only* in case some questions arise while reading; the answer could be found in the

Appendix "*Endnotes*" with the number i. The asterisk* marks the sections that can be skipped on the first reading.

In addition to the appendixes "*Help*" and "*Endnotes*" the book contains "*Mathematical appendix*" and "*Stochastic manual*". The first one collects the actively used definitions and formulas of probability theory, mathematical and tensor analysis; the second one presents various formulas of stochastic math.

The "*Stochastic manual*" may also be useful for the Reader who is already familiar with stochastic differential equations. However, it is strongly recommended to first read the page 40 and look through the sections §2.8, p. 53, and §5.1, p. 109.

Contents

Chapter 1
Random Events

Absolutely deterministic events and processes do not exist. The Universe speaks to us in the language of probability theory. We assume that the Reader is familiar with the basics of probability, therefore, only those terms and concepts that are necessary for the understanding of further material are introduced.

The first section is introductory; it is concluded with the necessity of using stochastic differential equations when studying various systems. After that the concept of probability density is discussed, which allows to compute the mean values of observable variables. The Gaussian probability distribution lies in the basis of noise, which influences the deterministic dynamics. Both stochastic dependencies and independencies between random variables are important when searching for the relation between different objects and their characteristics. "*The Model of Additive Random Walk*" is the key section of the chapter. The generalization of this simple model leads to the stochastic differential equations in the next chapter. The last section "*Martingales and the Free Cheese*" includes a number of formal definitions that can be skipped if necessary. Before reading the chapter it would be useful to look through the elements of the probability theory in the Mathematical Appendix on p. 239.

1.1 Stochastic World

• Thanks to the works of Newton and Leibnitz the scientists have differential equations at their disposal. If some values vary with time, there is the set of equations which describes this dynamics.

A situation when the quantity change is proportional to the quantity itself is commonly encountered. This is the simplest example:

$$\frac{dx}{dt} = \alpha\, x \qquad => \qquad x(t) = x_0\, e^{\alpha t}. \tag{1.1}$$

The function $x(t) > 0$ can describe the quantity of rabbits for which the rate of reproduction increases along with the increase in their present quantity.

S.S. Stepanov, *Stochastic World*, Mathematical Engineering,
DOI: 10.1007/978-3-319-00071-8_1,

Another example can be taken from the economics. The dynamics of means of production growth increases along with their accumulated quantity. The increase of population according to Malthus can also serve as an example. If $\alpha > 0$, this equation is called a *growth equation*, otherwise it is called a *decay equation*. There is an arbitrary constant x_0 in the solution. In order to determine it, it is necessary to set the initial quantity of the rabbits: for example, $x_0 = x(0) > 0$ at time $t_0 = 0$.

The exponential function grows very quickly. If rabbits were to reproduce according to this equation all the time, the Earth would have soon become completely white and fluffy. In reality they are not only reproducing but also dying. The *relative* change of the population size $dx/x = A\,dt$ can be a function of x in the general case. Let's expand it in the series $A(x) = \alpha - \beta\,x + ...$ up to the order of linear dependency. The second term has the meaning of relative deceleration of rabbit population due to natural resource depletion (the lack of grass). The process is getting even more intensive along with the population size growth. As a result the more realistic equation leads to the *logistic function*, which reaches the stationary value α/β at a certain moment in time (when $\alpha > 0$):

$$\frac{dx}{dt} = \alpha x - \beta x^2 \qquad => \qquad x(t) = \frac{\alpha}{\beta - (\beta - \alpha/x_0)\, e^{-\alpha t}}. \tag{1.2}$$

The solution to the equation (1.2) can be obtained ($\lessdot \mathrm{H}_1$) after the substitution $x(t) = 1/y(t)$. Asymptotically ($t \to \infty$), the equilibrium value $x_\infty = \alpha/\beta$ can be easily found from the equation, where $dx/dt = 0$ ($\lessdot \mathrm{C}_1$). It is worth mentioning that (1.2) can also be applied to the primates who consider themselves sapient and live on the planet with limited resources. However, the logistic equation itself has the flavor of cannibalism ($\lessdot \mathrm{C}_2$).

• Differential equations were first used in classical mechanics. The force $F(x)$ applied to a particle changes its momentum $p = m\dot{x}$:

$$\begin{cases} \dot{p} = F(x) \\ \dot{x} = p/m, \end{cases} \tag{1.3}$$

where a dot over a variable denotes the time derivative $\dot{x} = dx/dt$ and m is the particle mass. For example, in case the force is linear $F(x) = -kx$, the coordinate exhibits oscillatory behaviour $x(t) = x_0 \cos(wt) + (p_0/\omega m)\, \sin(wt)$ with the frequency $w = \sqrt{k/m}$ ($\lessdot \mathrm{H}_2$). As there are two equations, the solution contains two constants, and it is necessary to set two initial conditions for the coordinate $x_0 = x(0)$ and the momentum $p_0 = p(0)$.

Most economical, biological and physical systems can be described by the system of differential equations:

$$d\mathbf{x} = \mathbf{a}(\mathbf{x}, t)\, dt, \tag{1.4}$$

where $\mathbf{x}(t) = \{x_1(t), ..., x_n(t)\}$ is a vector of variables that describe the state of the system. The vector function $\mathbf{a}(\mathbf{x}, t)$ determines its dynamics.

Any differential equations that contain second and higher derivatives can be reduced to the system (1.4) by introducing new dynamical variables. Differential equations of motion in Hamiltonian form (1.3) can be a good example of this statement.

Equation (1.4) describes the change of vector $\mathbf{x}(t)$ during the infinitesimal time period dt. This representation gives a simple algorithm of numerical integration of the equations (1.4) in the situation when the analytical solution cannot be obtained. For this purpose the infinitesimal changes are replaced with small but finite ones $\Delta\mathbf{x} = \mathbf{x}_{k+1} - \mathbf{x}_k$, $\Delta t = t_{k+1} - t_k$. As a result, Eq. (1.4) corresponds to the discrete *iteration scheme*:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{a}(\mathbf{x}_k, t_k)\,\Delta t. \tag{1.5}$$

Given the initial vector $\mathbf{x}_0$ we can obtain its new value $\mathbf{x}_1$ after the time period Δt. Then $\mathbf{x}_1$ substitutes $\mathbf{x}_0$ and we get $\mathbf{x}_2$. By repeating this procedure it is possible to obtain the sequence of vector values $\mathbf{x}(t)$ in discrete points in time t_0, $t_1 = t_0 + \Delta t$, $t_2 = t_0 + 2\Delta t$, etc. The smaller the time period Δt is, the closer the numerical values of the scheme (1.5) approach the "true" solution of equation (1.4).

If the vector function $\mathbf{a}(\mathbf{x}, t)$ is smooth, the solution of the system of equations is also a set of smooth functions. Here we use the term "smoothness" quite informally. It means that such functions are differentiable on the whole domain of their arguments.

• Differential equations are ubiquitous in natural sciences. The progress of science within the last three centuries is amazing. However, the thorough comparison of theoretical results with experimental data shows that ordinary differential equations are only part of the truth.

In most situations the studied systems are subjected to unpredictable external influences which make their dynamics less smooth. The stone which flies along a parabola complies with the mathematical curve as a rough approximation only. Its inevitable contact with the air results in certain fluctuations around this trajectory. Even bigger irregularity can be found when we consider small objects like Brownian pollen which are exposed to the irregular molecular strokes. They follow the broken-line trajectory. The complexity of this trajectory $\mathbf{x}(t)$ is so great that its time derivative does not exist.

The importance of stochastic processes increases along with the structural complexity of natural and social systems. Rabbits multiply according to the logistic equation as a rough approximation only. The fluctuations of the population size caused by the internal and external random factors, which are not considered in the simple model (1.2), are significant indeed. Correspondingly, the economic growth has an exponential form only as first approximation. In reality the function $x_0 e^{\alpha t}$ is significantly distorted by economic ups and downs that have stochastic nature and are difficult to predict. Thus, randomness is dominant in the financial world. It determines the inner character of markets. Therefore stochastics is not a negligible correction but the main approximation to the reality like that for the Brownian movement.

Thus, our world is not deterministic. Its real nature is stochastic:

The ordinary differential equations are only a rough approximation of the reality. The stochastic equations provide more adequate research instrument ($\lessdot C_3$).

The corresponding mathematical formalism will be described further in these lectures. It allows combining two completely different entities: deterministic smooth dynamics and uneven broken-line random processes.

In case "the noise component" of a stochastic differential equation is small, its solution will be smooth enough. The gradual increase of noise part makes the contribution of stochastic dynamics dominating.

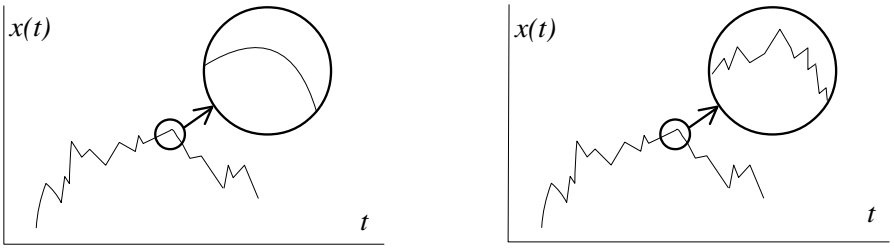
As for the external noise that breaks the smooth dynamics, we assume that the following *stochastic equation* is satisfied:

$$d\mathbf{x} = \mathbf{a}(\mathbf{x}, t)\, dt + \mathbf{Noise}(\mathbf{x}, t, dt). \tag{1.6}$$

It describes both the deterministic (the first term) and the random (the second one) change of system state variables $\mathbf{x}$. As $d\mathbf{x}$ is considered to be small, the noise will decrease along with the time period dt. Our discussion will be devoted to the introduction of noise with certain properties $\mathbf{Noise}(\mathbf{x}, t, dt)$ into differential equations.

The noise can change in time (see the dependency on t) and may also depend on the value of dynamic variables $\mathbf{x}$. The explicit functional dependency on $\mathbf{x}$ and t is specific to the given problem, and its determination often requires some complicated empirical research.

Let us consider *the random function* $\mathbf{x}(t)$ which is a solution to the stochastic equation. Usually it would be very different from the "well-behaved" functions of mathematical analysis. In we look at a strongly irregular ordinary function under "magnifying glass", we can see that it turns out to be smooth at small scale. A stochastic, random function would stay broken at any scale:



Despite the fact that the random function $\mathbf{x}(t)$ is assumed to be continuous, as a rule it is non-differentiable. Indeed, since the derivative is defined as $[\mathbf{x}(t + \Delta t) - \mathbf{x}(t)]/\Delta t$ as Δt tends to zero, in case of stochastic functions no matter how small the time period is, the direction of the function *change* can have an unpredictable *sign* due to random factors. Therefore no convergence to the definite limit can be obtained. Other facts from mathematical analysis have to be reconsidered for such $d\mathbf{x}$ as well.

The methods of solving the equations like (1.6) are of great importance. In cases when the exact solution cannot be found, we can use the numerical modeling or approximate analytical methods. There is no need to remind that any mathematical tool is developed to get more powerful means of studying the surrounding world. So it is necessary to see the real random process in finance, physics or biology behind each equation or its solution.

1.2 Random Variables

• Consider a random variable with its observed values which give us the set of numbers $x_1, x_2, ...$ These can be the quotes of daily stock prices or the coordinates of a Brownian particle. The numbers x_1, x_2, ... can be considered as possible realizations of the *random variable* x. In the first stage of investigation the ordering of numbers x_i doesn't matter. For example, this sequence can be randomly shuffled.

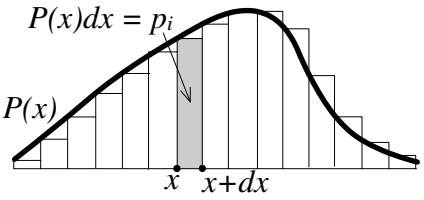
Assume that x_i occurs n_i times and the total quantity of numbers is equal to n. The following expression is called *the mean value* of random variable x:

$$\bar{x} = \langle x \rangle = \frac{1}{n} \sum_i x_i n_i = \sum_i x_i p_i = \int_{-\infty}^{\infty} x \; P(x)\, dx, \tag{1.7}$$

where $p_i = n_i/n$ are the relative *frequencies* (or *probabilities*) of x_i occurrence. If all x_i are different, the sample mean is equal to their sum divided by n. The more probable x_i is, the bigger contribution to the mean it gives due to its more frequent occurrence.

The majority of financial or physical quantities are continuous. For the infinite number of observations the sum is replaced by the integral. *Probability density function (pdf)* is such a function $P(x)$ which, when multiplied by the interval dx, gives the probability p_i that corresponds to the event that the value x falls into the segment from x to $x + dx$.

The probability of finding the value x at some point on a interval $[-\infty..\infty]$ is equal to the area under the curve $P(x)$. As this is *a certain event*, it has a unit probability:



$$\sum_i p_i = \int_{-\infty}^{\infty} P(x)dx = 1. \tag{1.8}$$

This equation is called a *normalization condition.*

Sometimes the random variable has "forbidden" values. For example, the price or the number of rabbits are always positive. In this case the probability of finding x in the range $x < 0$ is equal to zero. When calculating the mean we will often integrate from negative to positive infinity. Therefore probability density function must be equal to zero in the "forbidden" intervals of the random variable.

• If the probability density function is known, it is possible to find the mean of an arbitrary deterministic function $F(x)$ of random variable x:

$$\langle F(x) \rangle = \overline{F(x)} = \int_{-\infty}^{\infty} F(x)\ P(x)\, dx.$$

Going forward in this book we will be denoting *the procedure* of averaging by two *equivalent* notations – braces or overline. The notation $\mathbf{E}F(x)$ is also common in mathematical and financial literature.

Since the mean is a sum (or an integral), the mean of the sum of two functions is equal to the sum of their means. In addition, it is possible to move a multiplicative constant out of the mean sign:

$$\langle \alpha\, f(x) \rangle = \alpha\ \langle f(x) \rangle\,, \qquad \langle f(x) + g(x) \rangle = \langle f(x) \rangle + \langle g(x) \rangle\,.$$

But that's all! In general, non-linear functions cannot be moved out of the mean sign: $\langle x^2 \rangle \neq \langle x \rangle^2$.

• *Volatility* σ is another important characteristics of a random variable:

$$\sigma^2 = \left\langle (x - \bar{x})^2 \right\rangle = \int_{-\infty}^{\infty} (x - \bar{x})^2\ P(x)\, dx.$$

In "non-financial" applications volatility σ is usually called a standard deviation. Its square is *a variance*: $\sigma^2 = \mathbf{Var}(x)$. Being a constant, the mean $\bar{x}$ can be moved out of the mean sign, therefore,

$$\sigma^2 = \left\langle (x - \bar{x})^2 \right\rangle = \left\langle x^2 - 2x\bar{x} + \bar{x}^2 \right\rangle = \left\langle x^2 \right\rangle - 2\bar{x}\, \langle x \rangle + \bar{x}^2 = \left\langle x^2 \right\rangle - \langle x \rangle^2\,.$$

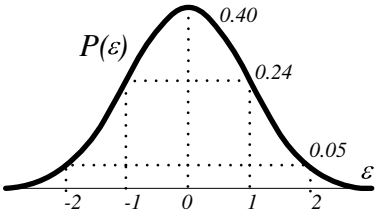
If the probability density function of a continuous random variable has a single symmetric peak, then its mean is a good characteristics of "the most typical" value of x. Volatility is the typical deviation of x from its mean. The smaller is σ, the narrower is the probability density function $P(x)$. As $\sigma \to 0$, the random variable is becoming almost completely deterministic with the value $x = \bar{x}$.

It is possible to define the higher *moments* by analogy with variance. Thus, the dimensionless expressions

$$asym = \left\langle (x-\bar{x})^3 \right\rangle / \sigma^3, \qquad excess = \left\langle (x-\bar{x})^4 \right\rangle / \sigma^4 - 3 \tag{1.9}$$

are called *skewness* (asymmetry) and *kurtosis* (excess). Skewness is the characteristics of the "asymmetry" of probability density. It is equal to zero for a symmetric function $P(x)$. When kurtosis takes a big positive value, $P(x)$ decreases slower while moving away from the mean than when it is negative.

• *Gaussian probability density*, or *the normal distribution*, occurs very often. Bellow, we denote the corresponding random variable by ε. We will not distinguish the notation for the random variable ε and the variable in its probability density. It looks as follows for the normal distribution:



$$\boxed{P(\varepsilon) = \frac{e^{-\frac{1}{2}\varepsilon^2}}{\sqrt{2\pi}}} \tag{1.10}$$

The mean of ε is equal to zero $\langle \varepsilon \rangle = 0$; the mean of its square is equal to a unit correspondingly: $\left\langle \varepsilon^2 \right\rangle = 1$. Therefore, the variance is also a unit: $\sigma_\varepsilon^2 = 1$. Further, let us introduce the following notation: $\varepsilon \sim N(0,1)$. If we consider the random variable $x = \mu + \sigma\,\varepsilon$, it will have mean μ and volatility σ, thus $x \sim N(\mu, \sigma^2)$. ($\ll \mathrm{C}_4$)

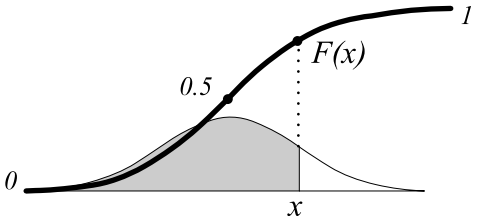
It is useful to know the form of the *moment-generating function* for Gaussian quantities, which is defined as the mean of the exponential function [see (14), p. 255]:

$$\langle e^{\alpha\,\varepsilon} \rangle \;=\; \int_{-\infty}^{\infty} e^{\alpha\,\varepsilon}\, P(\varepsilon)\, d\varepsilon \;=\; e^{\alpha^2/2}. \tag{1.11}$$

Series expansion by parameter α of the left and right sides of (1.11) makes it possible to easily find the means of arbitrary powers $\langle \varepsilon^n \rangle$ ($\ll \mathrm{H}_3$).

Particularly, $\left\langle \varepsilon^4 \right\rangle$ is equal to 3 and so $excess = 0$. Substraction of 3 from the dimensionless fourth moment in the definition of kurtosis (1.9) is related to our desire to compare everything to the normal distribution. If $excess > 0$, then the distribution is likely to have "fat tails", i.e. it lies above the normal distribution line (for $x \to \pm\infty$). If kurtosis is negative, the situation is opposite and the distribution tails lie below the normal plot.

The integral distribution:



$$F(x) = \int\limits_{-\infty}^{x} \frac{e^{-\varepsilon^2/2}}{\sqrt{2\pi}}\, d\varepsilon \qquad (1.12)$$

is the probability of the following event: the random variable is equal to or less than x.

• If probability density function $P(x)$ of x is known, then it is possible to find the probability density of another random variable y that is related to x according to a certain functional dependency $y = f(x)$. For this purpose the mean of an *arbitrary* function $F(y)$ is calculated. This can be done by averaging with the known probability density function $P(x)$:

$$\langle F(y)\rangle = \int\limits_{-\infty}^{\infty} F\big(y\big)\, \tilde{P}(y)\, dy = \int\limits_{-\infty}^{\infty} F\big(f(x)\big)\, P(x)\, dx. \qquad (1.13)$$

As $\tilde{P}(y)$ is not known, let us integrate with $P(x)$ and substitute $y = f(x)$ into $F(...)$. It is possible to convert the second integral into the first one with the reverse substitution. The multiplier at $F(y)$ in the integrand appears to be the required probability density function $\tilde{P}(y)$ for y.

Consider the random variable $r = \mu + \sigma\,\varepsilon$ as an example. Let it have the normal distribution with the mean μ and the volatility σ. Let us find the distribution for $x = x_0\, e^r$, where x_0 is constant:

$$\langle F(x)\rangle = \int\limits_{-\infty}^{\infty} F\big(x_0 e^{\mu+\sigma\,\varepsilon}\big)\; e^{-\varepsilon^2/2} \frac{d\varepsilon}{\sqrt{2\pi}} = \int\limits_{0}^{\infty} F(x)\; e^{-[\ln(x/x_0)-\mu]^2/2\sigma^2} \frac{dx}{x\sigma\sqrt{2\pi}}.$$

The first integral gives the expression for the mean with the normal distribution. The following substitution is made there: $x = x_0\, e^{\mu+\sigma\varepsilon}$, $dx = \sigma x d\varepsilon$. As a result, for $x \geqslant 0$ we get:

$$P_L(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(x/x_0)-\mu)^2}{2\sigma^2}\right]. \qquad (1.14)$$

The probability $P_L(x)$ is called *lognormal distribution*. As an exercise, one can calculate the mean $\langle x\rangle$ using $P_L(x)$ or Gaussian density $P(\varepsilon)$ ($\ll \mathrm{H}_4$).

• When using the random variables in the expressions like $x = \mu + \sigma\varepsilon$, we don't perform arithmetic operations with specific numbers. Instead *a potential* calculations are demonstrated: "if ε appears to be equal to some values, then x ..." Sometimes a distinction is made in notation when calculating the mean between a random variable denoted with the capital letter X and an integration variable which is denoted with the small letter x. We are not going to do it here.

1.3 Conditional and Joint Probability

• Consider two random variables x and y and the pairs of their observed values $\{x_1, y_1\}$, $\{x_2, y_2\}$, and so on, that occur with a certain frequency. We can now define *the joint probability density* $P(x, y)$ of the event that the quantities take certain values in the neighborhood of x and y. Joint probability allows us to calculate the mean of an arbitrary two-variable function:

$$\langle F(x, y)\rangle = \int_{-\infty}^{\infty} F(x, y)\, P(x, y)\, dx\, dy. \tag{1.15}$$

If we are not interested in y value, $P(x, y)$ can be integrated over all possible realizations of this quantity. As a result, we receive the probability density only for x:

$$\int_{-\infty}^{\infty} P(x, y)\, dy = P(x). \tag{1.16}$$

If after that we integrate the left and the right sides by x we get a unit. For this reason *a condition of normalization* has the form of a double integral. It can be obtained from (1.15) if $F(x, y)$ is set to 1, as $\langle 1\rangle = 1$.

Simultaneous studying of x and y doesn't necessarily mean their obligatory coincidence in time. For example, in finance x may be the daily price change of a European stock index and y is the corresponding change of an American stock index which is traded *after* the European one. There is the causal relation between them separated in time. On the other hand the daily price change of two stocks x and y happens simultaneously and depends on external synchronizing factors (news, macroeconomics).

As one can see in the following section the joint probability density $P(x, y)$ is extremely important if there is a certain dependency between two random variables. This relation can be expressed as a function $y = f(x)$. Then, in case a certain value is realized for x, the quantity y will be fully predetermined. However, the following case is more frequent: $y = f(x, \xi)$, where ξ is another "non-observable" random variable. It might be an unpredictable external impact that changes the parameters of functional dependency $y = f(x)$, or the dynamic variable which was not taken into consideration in the simpler model.

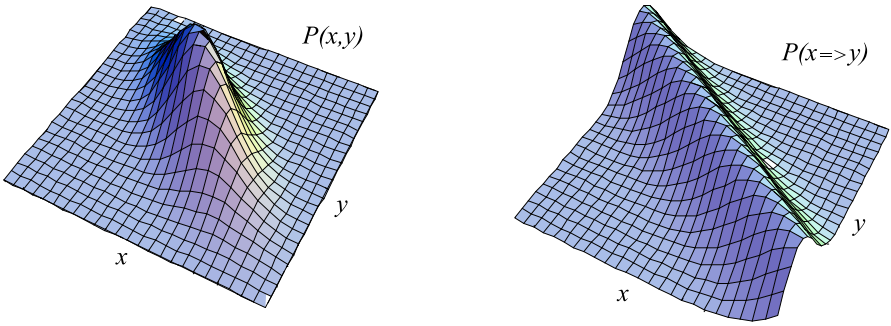
• In addition to the joint probability of two quantities x and y it is convenient to introduce *a conditional probability density*. It answers the following question: what is the probability of y if the value of x is already known? The conditional density is equal to the joint density $P(x, y)$ normalized by the probability of already available information $P(x)$ (see p. 242 in appendix M):

$$\boxed{P(x \Rightarrow y) = \frac{P(x, y)}{P(x)}}. \tag{1.17}$$

Let us take the normal distribution (1.10) as an example of $P(x)$, while for joint probability density $P(x, y)$ let us take the "two-dimensional displaced" normal plot:

$$P(x, y) = \frac{e^{-(x^2+y^2+\sqrt{2}\,xy)}}{\pi\sqrt{2}}, \qquad P(x \Rightarrow y) = \frac{e^{-(x^2/2+y^2+\sqrt{2}\,xy)}}{\sqrt{\pi}}.$$

The joint and conditional probabilities are shown in the figure below:



The volume under $P(x, y)$ is equal to a unit while the volume under $P(x \Rightarrow y)$ is equal to infinity. Normalization of conditional probability has the meaning of obtaining any value y for the a given x:

$$\int_{-\infty}^{\infty} P(x \Rightarrow y)\, dy = 1. \tag{1.18}$$

It is worth verifying that the formula (1.18) agrees with (1.16).

Let us note that the following notation is more common for conditional probability: $P(y|x)$. But it will be shown further that $P(x \Rightarrow y)$ appears to be more natural notation when describing the chains of events connected to each other. In any case $P(x \Rightarrow y)$, like $P(x, y)$, is the function of *two* real arguments.

Conditional probability is important as it allows to link different events with each other, manifesting their cause-effect relation.

• Consider probabilistic properties of the English language. Each of 28 letters including space "_" and apostrophe has its own probability of occurrence in a text:

$$p(\_) = 0.192, \quad p(\mathbf{e}) = 0.099, \quad p(\mathbf{t}) = 0.074, ..., \; p(\mathbf{z}) = 0.0005.$$

If we need to find the probability of certain substring, e.g. "**th**" to appear in a random place, we have to count occurrences of such substrings and divide it by total number of all substrings like "******", where asterisk is any character. In order to calculate the conditional probability $P(\mathbf{t} \Rightarrow \mathbf{h})$ of the occurrence of letter "**h**" conditioned on the fact that there is letter "**t**" before it, we will need to select all substrings that satisfy the pattern "**t***" ("**t**", then any character "*****") and find out how many "**th**" there are among them. This results in:

$$p(\mathbf{th}) = N(\mathbf{th})/N(**) = 0.024, \qquad p(\mathbf{t} \Rightarrow \mathbf{h}) = N(\mathbf{th})/N(\mathbf{t}*) = 0.328,$$

where N is the number of substrings that satisfy the corresponding mask. For a text that has n characters: $N(**) = n - 1$ and $N(\mathbf{t}*) = p(\mathbf{t})\, n$. The number of both joint and conditional probabilities for two letters equals to $28^2 = 784$.

The probability of finding a specific letter in a text depends on the prehistory (previous letters). For example, the probability of "**h**" occuring after "**t**" is 6.5 times higher than the absolute probability of "**h**" occur: $p(\mathbf{h}) = 0.050$. Vice versa, some combinations of letters are very difficult to pronounce. E.g. "**z**" is very unlikely to appear after "**q**".

Knowing the conditional probabilities we can create synthetic texts. Thus, the new letter "**x**" will be generated with probability

$$p(...\mathbf{cba} \Rightarrow \mathbf{x})$$

according to the known prehistory "**...cba**". The longer is the history on which we condition, the more "euphonious" combinations will appear in such text:

- ▷ $P(x)$: teiesgeo sn lsdupeaguylohsnnr a soontwe as ihh leoaanhe lttaea iv a ebtaeeeedcfroi oone shcw nihdeeetn h gelhrrnn;
- ▷ $P(a \Rightarrow x)$: hescer man plyocerkn'ma mofind maritathes hilel whed inond as hr s threr ath ruprout win modangeno at oun y d ct o;
- ▷ $P(ba \Rightarrow x)$: and bea me mor laysid usal barry inted is that se pand wit com a saught i'm the drupoing on to yession the as rettly;
- ▷ $P(cba \Rightarrow x)$: wild the sames resough wered in his do in him for i do said in ver through this can one oth pret be the dog frient.

In the first case only unconditional probabilities are used and prehistory is not taken into consideration at all. In the second case a single previous letter is used to determine the next, etc.

• As a second example let us use the data of daily closing prices x_t of the stock index S&P500. First calculate its log returns

$$r_t = \ln(x_t/x_{t-1})$$

in percentage terms ($\ll \mathrm{C}_6$). Then discretize the values into the five intervals:

$$(-\infty...-3\%),\ [-3\%...-1\%),\ [-1\%...+1\%],\ (+1\%...+3\%],\ (+3\%...+\infty).$$

Therefore, the market state is characterized by one of the five possibilities: starting from "panic" $(-\infty...-3\%)$ and ending with "euphoria" $(+3\%...\infty)$. As such, each r_t turns into a discrete random variable which can take five values. This values are not returns anymore. These are the market states numbers, e.g. -2,-1,0,1,2.

It is possible to consider *the joint probability* $p(r_{t-1}, r_t)$ of an event that two consecutive days have the states r_{t-1} and r_t. Every day one of five possibilities realizes, so there will be $25 = 5^2$ different combinations of the following states for two consecutive days: {(0,0); (0,1); (0,-1);...}. Between 1990 and 2007 there were $n = 4531$ trading days. Occurrence of each state revealed the following values for the probabilities:

$$p(r) = \begin{pmatrix} 0.007 & 0.110 & 0.761 & 0.125 & 0.007 \end{pmatrix}.$$

In order to calculate them we have to count the number of trading days in every state; after that we need to divide them by n. The calm days are the most typical for the market $[-1\%...+1\%]$. They took place $3451 = 0.76 \cdot 4531$ times during the considered period. Similarly to the letters from the previous example, conditional probabilities can be calculated:

$$p(r_{t-1} \Rightarrow r_t) = \begin{pmatrix} 0.067 & 0.167 & 0.400 & 0.267 & 0.100 \\ 0.022 & 0.146 & 0.651 & 0.168 & 0.014 \\ 0.004 & 0.107 & 0.783 & 0.102 & 0.004 \\ 0.006 & 0.084 & 0.759 & 0.138 & 0.013 \\ 0.000 & 0.303 & 0.515 & 0.152 & 0.030 \end{pmatrix}.$$

The first row in this matrix corresponds to the state transition from "panic" yesterday to one of five possible states today. In the same way the last row includes the conditional probabilities of state transition from "euphoria". It is notable that the probabilities of state transition from "calm" market (the middle row) are very close to the unconditional probabilities $p(r)$. In case the market was not calm yesterday, probabilities differ from the daily ones. It is especially clear ($\ll \mathrm{C}_5$) for the extreme states of "panic" and "euphoria". Since the probability that some state will happen next is equal to a unit, the sum of the numbers in each row is also equal to one [see (1.18)].

1.4 Dependency and Independency

• Quantities are statistically *independent* if their joint probability density is equal to the product of the probability distributions of each quantity:

$$P(x,y) = P_1(x)\,P_2(y)\ .$$

We will often be omitting the indexes here and will be using the same letter to denote different functions distinguishing them by an argument. It follows from the definition (1.17) that for independent events the conditional probability density $P(x \Rightarrow y) = P(y)$ depends only on y. This equation can be considered as another definition of event independency. If the occurrence of event y doesn't depend on the fact whether the event x took place or not, then they are independent.

The mean of a product of several independent quantities is equal to the product of their means:

$$\langle x\,y\rangle = \int\limits_{-\infty}^{\infty} x\,y\ P(x)P(y)\ dxdy = \langle x\rangle\ \langle y\rangle\,.$$

So, *the covariance*

$$\mathrm{cov}(x,y) = \langle (x-\bar{x})(y-\bar{y})\rangle = \langle xy\rangle - \langle x\rangle\,\langle y\rangle \tag{1.19}$$

between independent quantities is equal to zero. Note, that the opposite statement can be false ($\lessdot \mathrm{C}_7$).

• The function $z = f(x,y)$ of two random variables x and y is also the random variable with a certain distribution $P(z)$. In order to find $P(z)$, we need to transform the formula for the mean of an arbitrary function $F(z)$ in such a way that it turns into an integral by z only:

$$\langle F(z)\rangle = \int\limits_{-\infty}^{\infty} F\big(f(x,y)\big)P(x,y)\,dxdy = \int\limits_{-\infty}^{\infty} F(z)P(z)\,dz. \tag{1.20}$$

E.g., if x and y are *independent* Gaussian numbers with the arbitrary volatilities σ_x, σ_y, then the quantity $z = x + y$ is also Gaussian:

$$\langle F(z)\rangle = \int\limits_{-\infty}^{\infty} F\big(x+y\big)\,e^{-x^2/2\sigma_x^2 - y^2/2\sigma_y^2}\ \frac{dxdy}{2\pi\sigma_x\sigma_y} = \int\limits_{-\infty}^{\infty} F(z)e^{-z^2/2\sigma^2}\ \frac{dz}{\sigma\sqrt{2\pi}},$$

where $\sigma^2 = \sigma_x^2 + \sigma_y^2$. The following substitution is performed in the double integral: $z = x + y$, $u = x$. Integration by u is carried out according to the formula (14) on p. 255 of the Appendix M. Thus, *the sum of two normal quantities is also a normal quantity.*

• Let x and y be two *independent* random variables with *arbitrary* distributions. Consider z which is their sum: $z = x + y$. As we discused before, the mean is equal to the sum of means: $\bar{z} = \bar{x} + \bar{y}$. Find the variance:

$$\sigma_z^2 = \left\langle (z - \overline{z})^2 \right\rangle = \left\langle (x - \overline{x} + y - \overline{y})^2 \right\rangle = \sigma_x^2 + \sigma_y^2 + 2 \left\langle (x - \overline{x})\,(y - \overline{y}) \right\rangle ,$$

where the square was expanded under the mean sign and the volatility of each quantity was introduced, for example, $\sigma_x^2 = \left\langle (x - \bar{x})^2 \right\rangle$. *If* (!) x and y are *independent*, their covariance (the last term) is equal to zero: $\left\langle (x - \overline{x})\,(y - \overline{y}) \right\rangle = \left\langle x - \overline{x} \right\rangle \left\langle y - \overline{y} \right\rangle = 0$. So,

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2.$$

In general the following is true for the sum of n independent quantities:

$$z = x_1 + ... + x_n \qquad => \qquad \sigma_z^2 = \sigma_1^2 + ... + \sigma_n^2. \tag{1.21}$$

To prove this we have to consider $x_1 + x_2$ as one random variable and obtain

$$\sigma_z^2 + \sigma_3^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2$$

after adding x_3 to it, and so on.

If the volatilities of each independent x_i are the same and equal to σ_0, the volatility of their sum will increase with the number of terms as $\sigma_z = \sigma_0 \sqrt{n}$. This square root relation between n and σ is extremely important. It lies in the basis of all the **Noise** properties which we will add to deterministic differential equations.

We emphasize that the obtained result (1.21) doesn't depend on the distribution of quantities x_i; they can even be different. The only condition is that they must be independent.

We earlier derived the same result for the sum of two independent Gaussian numbers. However, in that case the probability density turned out to be Gaussian as well. A random variable z is called *infinitely divisible* if it can be represented as a sum of *independent* random variables, which have *the same* distribution as z (perhaps, with different parameters). The Gaussian probability density is one example of an infinitely divisible distribution, others are Cauchy distribution and gamma function, which are considered in the next section.

In fact, it is enough for *infinite* divisibility that all three quantities in $z = x + y$ have the same distribution. Though the same *functional* form of the distribution is implied, the parameters (in particular, volatility) can be different. In general case, the sum of arbitrary distributed numbers has a distribution that is different from the distribution of each term in the sum. However, Eq.(1.21) is always true for independent quantities and turns out to be a universal relation.

• The simplest relation between two random variables x and y is a linear dependence $y = \alpha + \beta\, x$. In more general case there can be a third random variable ξ, which is interpreted as an "external" *random noise*. The resulting model (regression) parametrized by constants α and β looks as follows:

$$y = \alpha + \beta\, x + \xi. \tag{1.22}$$

The search of the relation between empirical quantities usually starts from this equation.

Usually the mean value of the noise is assumed to be zero:

$$\langle \xi \rangle = 0.$$

Otherwise, it can be included in the parameter α. We require that the "noise" variance ξ (the model error) is minimized:

$$\sigma_\xi^2 = \left\langle \xi^2 \right\rangle = \left\langle (y - \alpha - \beta\, x)^2 \right\rangle = min. \tag{1.23}$$

Taking the derivatives by α and β, it is possible ($\lessdot \mathrm{H}_5$) to find the equation of the *regression line*. Its slope β is equal to

$$\beta = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2} = \frac{\langle (x - \bar{x})(y - \bar{y}) \rangle}{\sigma_x^2}. \tag{1.24}$$

Finally, we rewrite the equation in a symmetrical form of ratios of dimensionless deviations from the means:

$$\frac{y - \bar{y}}{\sigma_y} \;=\; \rho(x, y)\, \frac{x - \bar{x}}{\sigma_x} \;+\; \frac{\xi}{\sigma_y}. \tag{1.25}$$

The coefficient of this linear law is called *correlation*:

$$\rho_{xy} = \rho(x, y) = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y}, \tag{1.26}$$

and contains the covariance (1.19) in the numerator.

A non-vanishing correlation ($\rho \neq 0$) between two quantities x, y doesn't necessarily mean the presence of a *causal relation* $y = f(x)$ or $x = g(y)$. For example, there can be the third quantity z which influences both x and y synchronizing their behavior. Thus, a recession in the world economy has the same impact on two industries that are not connected to each other and are both export-oriented. *A "false" correlation* also occurs when two quantities have an explicit ascending or descending *trend* (a systematic increase or decrease). In this case a significant correlation will appear between them. This correlation shows the presence of a deterministic growth component ($\lessdot \mathrm{C}_8$).

• The correlation coefficient determines the slope of the regression line. However, it is more important that it can be a measure of the linear model forecast capability. Let us show this by substituting the initial equation (1.22) into the Eq. (1.24) and take into account that $\langle\xi\rangle = 0$ and $\bar{y} = \alpha + \beta\,\bar{x}$:

$$\beta = \frac{\langle (x-\bar{x})(\beta\,(x-\bar{x})+\xi)\rangle}{\sigma_x^2} = \beta + \frac{\langle x\xi\rangle}{\sigma_x^2}.$$

Thus, $\langle x\xi\rangle = 0$, which allows us to calculate the variance y:

$$\sigma_y^2 = \left\langle (y-\bar{y})^2\right\rangle = \left\langle (\beta\,(x-\bar{x})+\xi)^2\right\rangle = \beta^2\sigma_x^2 + \left\langle \xi^2\right\rangle.$$

As $\beta = \rho(x,y)\sigma_y/\sigma_x$, it is possible to obtain the expression for *the relative error* of the model:

$$E = \frac{\sigma_\xi}{\sigma_y} = \sqrt{1-\rho^2(x,y)}. \tag{1.27}$$

The value of the noise volatility $\sigma_\xi^2 = \left\langle \xi^2\right\rangle$ can be considered as the error of the linear model $y = \alpha + \beta x$. It is useful to compare it with the volatility σ_y, which is the typical error of the trivial model $y = \bar{y}$. Let us note that such relative error E depends on the correlation coefficient. The closer is its square to 1, the smaller is the error. For vanishing ρ the relative error is equal to 1. Correspondingly, the linear model has the same forecasting power as the trivial statement that the best forecast for y is its mean. *The coefficient of determination*

$$R^2 = 1 - E^2 = \rho^2$$

is also frequently used. Note that the absolute value of the correlation coefficient is always less than one: $|\rho| \leqslant 1$.

• In conclusion, let us mention that the linear model (1.22) can be interpreted in different ways.

1) First, this is a forecasting model for y if x is known (like $P(x \Rightarrow y)$). In this case ξ is an external noise, or the model error when the "true" dependence between x and y is not that simple. In this case y always appears to be a random variable because of the noise. As for x, there are different possibilities. For example, when studying a demand curve, x can be a product price which is controlled and set by the researcher (e.g. equally spaced price points). In this case it is deterministic. However, the dispersion of its values allows formally determining the mean $\bar{x}$ and the volatility σ_x.

2) It often happens that both x and y are equivalent random variables. For example, the daily stock price changes of two companies x and y are stochastically related to each other in the stock market. Both quantities are random and aren't affected by the researcher.

1.5 The Characteristic Function

• *The characteristic function* $\Phi(q)$ is the Fourier transform (p. 257) of the probability density of a random variable x:

$$\Phi(q) = \int_{-\infty}^{\infty} e^{\imath q x}\ P(x)\ dx, \qquad P(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\imath q x}\ \Phi(q)\ dq.$$

Using this function, it is easy to get the mean values of arbitrary powers of x. Doing one Fourier integration and finding the characteristic function, one can receive $\langle x^n \rangle$ by simple differentiation:

$$\frac{1}{\imath^n} \frac{d^n \Phi(q)}{dq^n}\bigg|_{q=0} = \int_{-\infty}^{\infty} x^n\ P(x)\ dx\ =\ \langle x^n \rangle\,.$$

The characteristic function can be represented as the mean of the exponent, i.e.: $\Phi(q) = \langle e^{\imath q x} \rangle$. It is obvious that $\Phi(0) = 1$. The coefficients of the series expansion of $\Phi(q)$ by q are the means of the powers of x:

$$\Phi(q) = \langle e^{\imath q x} \rangle = \sum_{n=0}^{\infty} \frac{\imath^n \langle x^n \rangle}{n!}\, q^n = 1 + \imath \langle x \rangle\ q - \frac{1}{2}\ \langle x^2 \rangle\ q^2 + ... \qquad (1.28)$$

Sometimes it is necessary to consider the real-valued variant of the characteristic function obtained by the substitution $q \to q/\imath$ and called the *moment-generating function*: $\Phi(q/\imath) = \phi(q) = \langle e^{q x} \rangle$.

• Assume that a random variable y is connected with x according to a linear dependency $y = a + b\,x$. Then its characteristic function is the following:

$$\Phi_y\big(q\big) = \langle e^{\imath q y} \rangle = \Big\langle e^{\imath q (a + b x)} \Big\rangle = e^{\imath q a} \left\langle e^{\imath q b x} \right\rangle.$$

Thus, an additional phase factor appears in the linear transformation and the argument q of Φ is scaled:

$$y = a + b\,x \qquad => \qquad \Phi_y(q) = e^{\imath q a}\ \Phi_x\big(b\,q\big). \qquad (1.29)$$

If $b = 0$, then $\Phi_y(q) = e^{\imath q a}$. If we take into account the integral expression for the Dirac delta function (p. 258), this leads to the probability density $P(y) = \delta(y - a)$. We see that y is no longer a random variable, but a deterministic constant $y = a$.

• Let us give some examples of characteristic functions for several important probability distributions:

$$\textit{Gauss}: \qquad P(x) = \frac{e^{-(x-x_0)^2/2\sigma^2}}{\sigma\sqrt{2\pi}}, \qquad \Phi(q) = e^{\imath x_0 q - \sigma^2 q^2/2}.$$

$$\textit{Cauchy}: \qquad P(x) = \frac{a/\pi}{(x-x_0)^2 + a^2}, \qquad \Phi(q) = e^{\imath x_0 q - a|q|}.$$

$$\textit{Gamma}: \qquad P(x) = \frac{1}{\gamma\Gamma(\mu)} \left(\frac{x}{\gamma}\right)^{\mu-1} e^{-x/\gamma}, \qquad \Phi(q) = \frac{1}{(1-\imath\gamma q)^\mu}.$$

In order to find $\Phi(q)$ for the Gauss distribution it is necessary to complete the square in the exponent. The Cauchy function $\Phi(q)$ is easier to check in opposite direction, calculating $P(x)$ from this function. In the third case a straightforward integration is performed for the *Gamma function* according to the formula (16), p. 256. Let us note that the Cauchy characteristic function $\Phi(q)$ is not analytic and the distribution doesn't have any finite moments $\langle x^m \rangle$ for $m > 1$.

• Consider two *independent* random numbers x, y with *the arbitrary* distributions $P_1(x)$, $P_2(y)$, and their sum $z = x + y$. Let us find the probability distribution $P(z)$ for the random variable z. For this purpose we calculate the mean of an arbitrary function (the integration is from $-\infty$ to ∞):

$$\langle F(z) \rangle = \int F(x+y)\ P_1(x)P_2(y)\ dx\,dy = \int F(z)\ \underbrace{P_1(x)P_2(z-x)\ dx}_{P(z)}\ dz,$$

where the following substitution is made: $y = z - x$. So

$$P(z) = \int\ P_1(x)P_2(z-x)\ dx.$$

The characteristic function for the sum of two independent quantities is equal to the product of their characteristic functions:

$$\Phi_z(q) = \left\langle e^{\imath q(x+y)} \right\rangle = \langle e^{\imath qx} \rangle\, \langle e^{\imath qy} \rangle = \Phi_x(q)\, \Phi_y(q),$$

where the fact that x and y are independent was used. It is obvious that in general case of n independent random variables x_i, the characteristic function of their sum is equal to the product of the characteristic functions of every term:

$$z = x_1 + ... + x_n \qquad => \qquad \Phi_z(q) = \Phi_1(q) \cdot .. \cdot \Phi_n(q).$$

If the distributions of every x_i are the same, we have $\Phi_z(q) = \Phi^n(q)$. Now it is possible to show that Gauss, Cauchy and gamma distributions are infinitely divisible ($\lessdot \mathrm{H}_6$).

• When studying random processes we will often use the fact of normal distribution infinite divisibility. In particular, if $\varepsilon_1, ..., \varepsilon_n$ are independent Gaussian quantities with zero mean and unit variance $\varepsilon_i \sim N(0,1)$, then their sum is also Gaussian:

$$\varepsilon_1 + ... + \varepsilon_n = \varepsilon \sqrt{n}. \tag{1.30}$$

The factor $\sqrt{n}$ is extracted for convenience, so that $\varepsilon \sim N(0,1)$ [(1.21), p. 14]. As a result, ε_i and ε have *the same distribution* with the same parameters (mean, moments, etc.). The characteristic function for ε satisfies the equation $\Phi(q)^n = \Phi(\sqrt{n}\, q)$ and is equal to

$$\Phi(q) = e^{-q^2/2}.$$

In general case a distribution $P(x)$ is called *stable* if for any n there exist such constants a_n and b_n that

$$x_1 + ... + x_n = a_n + b_n\, x, \tag{1.31}$$

where $x_1, ..., x_n$ and x have the same distribution $P(x)$. If $a_n = 0$, then such distribution is called *strictly stable*. The Gauss distribution with the constant $b_n = \sqrt{n}$ is strictly stable.

Let us remark that the condition (1.31) restricts the class of acceptable distributions more strongly than the simple requirement of infinite divisibility. The reason is that the left and right hand side in the definition (1.31) contain random variables that have distributions with *the same parameters*, whereas it is not necessary for divisibility.

Similarly to the linear scaling (1.29), the following functional equation is true for the characteristic function of a stable distribution:

$$\Phi^n(q) = e^{iqa_n}\Phi(b_n q). \tag{1.32}$$

It is not difficult to verify that both Gauss and Cauchy distributions satisfy this equation. At the same time Gamma distribution, which is infinitely divisible, is not stable. General functions satisfying (1.32) are called *Levy-Khinchin distributions*:

$$\Phi(q) = e^{\imath q\beta - \gamma[1 + \imath\theta\, \mathrm{sign}(q) \tan(\pi\alpha/2)]\,|q|^\alpha}, \qquad \Phi(q) = e^{\imath q\beta - \gamma|q| - \imath\gamma\theta\, q \ln|q|},$$

where $\mathrm{sign}(q) = q/|q|$ is the sign of q, and $0 < \alpha \leqslant 2$. In addition, $|\theta| \leqslant 1$, $\gamma \geqslant 0$. The first distribution is four-parametric. The second one is three-parametric and appears to be the limit of the first one as $\alpha \to 1$. These distributions can describe random numbers with "fat tails" (large excesses) if appropriate parameter values are set. This property is widely used when modeling the returns of financial instruments.

• Consider n independent random variables $x_1, ..., x_n$ which have arbitrary identical distributions. Let us study the properties of the following sum:

$$u = \frac{x_1 + ... + x_n}{\sqrt{n}}$$

as $n \to \infty$. Without loss of generality it can be assumed that $\langle x_i \rangle = 0$ as this is always possible to accomplish with the substitution $x \to x - \langle x \rangle$. In this case the mean of u is also equal to zero. As x_i are independent, the mean of u^2 is equal to the mean of x^2:

$$\left\langle u^2 \right\rangle = \frac{\left\langle x_1^2 \right\rangle + ... + \left\langle x_n^2 \right\rangle}{n} = \left\langle x^2 \right\rangle = \sigma^2.$$

For the identical *arbitrary* distributed x_i with $\Phi(q)$ and large n, the characteristic function for u looks as follows:

$$\Phi_u(q) \;=\; \left[\Phi\left(\frac{q}{\sqrt{n}}\right)\right]^n \;=\; \left[1 - \frac{\sigma^2}{2}\,\frac{q^2}{n} + ..\right]^n,$$

where we have used the equation (1.29) and expanded $\Phi(q/\sqrt{n})$ into the series up to the second infinitesimal order. The term proportional to q is equal to zero as $\langle x \rangle = 0$. By definition, the Euler number is the limit of

$$e^x = (1 + x/n)^n \quad as \quad n \to \infty.$$

Therefore, the characteristic function and distribution for u are tending to the Gaussian form:

$$\Phi_u(q) \to e^{-\sigma^2 q^2/2}. \tag{1.33}$$

As an exercise ($\lessdot \mathrm{H}_7$), it would be useful to find the asymmetry and excess for the characteristic function $\Phi_z(q) = \Phi^n(q)$ for large n.

The result (1.33) is extremely important. It can be formulated in the following way:

> *"the distribution of the sum of a large number of independent random variables tends to the normal distribution".*

For example, if some physical quantity is subjected to external independent random effects, then in most cases the distribution of its values obeys the Gauss distribution. A stock price is also subjected to the random background of supply and demand fluctuations in financial markets. However, its distribution is not Gaussian. Mainly, this is caused by two reasons: 1) the correlation of market players activity (as a result of synchronizing informational channels) and 2) their slow re-evaluation of the stock risk (volatility). We will get back to these issues in the Chapter 8.

1.6 Multidimensional Gaussian Distribution*

• When studying systems of stochastic equations we will often use matrix and tensor notation. In order to simplify the notation of matrix multiplication the following two conventions are used:

$$\eta_\alpha = \sum_{i=1}^{n} S_{\alpha i}\ \varepsilon_i \ = \ S_{\alpha i}\ \varepsilon_i = (\mathbf{S}\cdot\boldsymbol{\varepsilon})_\alpha. \tag{1.34}$$

Summation is always implied over the *repeated* index and the summation sign is *omitted*. Above, the index "i" in the second equality is of such kind. The repeated indexes which are used for summation are called "*dummy*". During the calculation they can be replaced by any letter that is not otherwise used in the expression. The third equality in the equation (1.34) is the matrix form of the same sum. The matrix $\mathbf{S} = S_{\alpha\beta}$ and the vector $\boldsymbol{\varepsilon} = \{\varepsilon_1, ..., \varepsilon_n\}$ are multiplied without explicit mentioning of indexes and a summation sign.

Consider n independent Gauss random variables, which have zero mean and unit variance. The mean value of their product $\langle \varepsilon_i \varepsilon_j \rangle$ is equal to one for same indexes and to zero for different ones. We will denote such a matrix by the Kronecker delta:

$$\langle \varepsilon_i \varepsilon_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

For example, let us calculate the covariance matrix of random variables η_α:

$$\langle \eta_\alpha \eta_\beta \rangle = S_{\alpha i} S_{\beta j} \langle \varepsilon_i \varepsilon_j \rangle = S_{\alpha i} S_{\beta j} \delta_{ij} = S_{\alpha i} S_{\beta i} = S_{\alpha i} S^T_{i\beta} = (\mathbf{S}\mathbf{S}^T)_{\alpha\beta}. \tag{1.35}$$

When summing up with Kronecker delta δ_{ij}, only the terms with $i = j$ survive in the sum. So, one of the sums (by j) and the Kronecker delta cancel each other out. Only the summation index i is left. Then *the new* matrix $S^T_{i\beta} = S_{\beta i}$ with the transposed indexes is introduced. This operation is called *transposition*. In the tabular presentation it corresponds to transposition of the matrix rows and columns.

A matrix $\mathbf{S}$ has the inverse one $\mathbf{S}^{-1}$, if the following equation holds:

$$\mathbf{S}\cdot\mathbf{S}^{-1} = \mathbf{S}^{-1}\cdot\mathbf{S} = \mathbf{1},$$

where $\mathbf{1} = \delta_{ij}$ is the unit matrix (the Kronecker delta). Thus, it is possible to write the following for the vector $\boldsymbol{\eta} = (\eta_1, ..., \eta_n)$ defined above:

$$\boldsymbol{\eta} = \mathbf{S}\cdot\boldsymbol{\varepsilon} \qquad => \qquad \boldsymbol{\varepsilon} = \mathbf{S}^{-1}\cdot\boldsymbol{\eta},$$

where we multiplied the left and right parts by $\mathbf{S}^{-1}$.

• Let $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)$ be standard independent Gaussian random variables $\varepsilon_i \sim N(0,1)$, and the quantities $\boldsymbol{\eta} = (\eta_1, ..., \eta_n)$ are obtained from them (1.34) by mixing the coefficients $S_{\alpha\beta}$. The mean of the product $\eta_\alpha \eta_\beta$ is determined by *the variance matrix* (1.35):

$$D_{\alpha\beta} = \langle \eta_\alpha \eta_\beta \rangle, \qquad \mathbf{D} = \mathbf{S} \cdot \mathbf{S}^T,$$

which is *symmetric*: $D_{\alpha\beta} = D_{\beta\alpha}$.

Let us find the moment-generating function for random variables η. To this end, we introduce the vector $\mathbf{b} = (b_1, ..., b_n)$ and calculate the exponent mean of the scalar product $\mathbf{b} \cdot \boldsymbol{\eta} = b_1\eta_1 + ... + b_n\eta_n$ (there is *no* sum by n!):

$$\left\langle e^{\mathbf{b}\cdot\boldsymbol{\eta}} \right\rangle = \left\langle e^{\mathbf{b}\cdot\mathbf{S}\cdot\boldsymbol{\varepsilon}} \right\rangle = \left\langle e^{b_i S_{i1}\varepsilon_1} \right\rangle \cdot ... \cdot \left\langle e^{b_i S_{in}\varepsilon_n} \right\rangle = e^{\frac{1}{2}\{(b_i S_{i1})^2 + ... + (b_i S_{in})^2\}}.$$

We have used the independence of quantities ε_i to split the mean of the product into the product of means, and the equation (1.11), p. 7. In the exponent, we have the following matrix expression:

$$(b_i S_{i1})^2 + ... + (b_i S_{in})^2 = b_i S_{ik}\, b_j S_{jk} = b_i\, S_{ik}\, S^T_{kj}\, b_j = \mathbf{b} \cdot \mathbf{S} \cdot \mathbf{S}^T \cdot \mathbf{b}.$$

Finally, the moment-generating function is:

$$\phi(\mathbf{b}) = \left\langle e^{\mathbf{b}\cdot\boldsymbol{\eta}} \right\rangle = e^{\frac{1}{2}\,\mathbf{b}\cdot\mathbf{D}\cdot\mathbf{b}}. \tag{1.36}$$

Taking the partial derivatives by b_α, it is not difficult to find any mean of any product η_α. Let us verify that the mean $\langle \eta_\alpha \eta_\beta \rangle$ is equal to $D_{\alpha\beta}$. Differentiate the moment-generating function by b_α. Considering that $\mathbf{b} \cdot \mathbf{D} \cdot \mathbf{b}$ is equal to $b_i D_{ij} b_j$, we have the following:

$$\frac{\partial \phi(\mathbf{b})}{\partial b_\alpha} = \frac{1}{2}\,(D_{\alpha j} b_j + b_i D_{i\alpha})\,\phi(\mathbf{b}) = D_{\alpha i} b_i\,\phi(\mathbf{b}),$$

where in the second equality we have used the fact that $D_{\alpha\beta} = D_{\beta\alpha}$. The second derivative can be found in the same way:

$$\frac{\partial^2 \phi(\mathbf{b})}{\partial b_\alpha \partial b_\beta} = D_{\alpha\beta}\,\phi(\mathbf{b}) + D_{\alpha i} b_i\, D_{\beta j} b_j\,\phi(\mathbf{b}).$$

Assuming $\mathbf{b} = 0$ and considering that

$$\frac{\partial^2 \left\langle e^{\mathbf{b}\cdot\boldsymbol{\eta}} \right\rangle}{\partial b_\alpha \partial b_\beta}\Big|_{\mathbf{b}=0} = \langle \eta_\alpha \eta_\beta \rangle,$$

we come to the equation $D_{\alpha\beta} = \langle \eta_\alpha \eta_\beta \rangle$. As an exercise, we encourage the reader to verify the following tensor expression:

$$\langle \eta_\alpha \eta_\beta \eta_\gamma \eta_k \rangle = D_{\alpha\beta} D_{\gamma k} + D_{\alpha\gamma} D_{\beta k} + D_{\alpha k} D_{\beta\gamma}.$$

Thus, the mean of any power of η is fully defined by the variance matrix $\mathbf{D}$.

• Now let us find the explicit form of the joint probability density for the quantities $\eta_1, ..., \eta_n$. First let us write the probability density for $\varepsilon_1, ..., \varepsilon_n$:

$$P(\varepsilon_1, ..., \varepsilon_n) = P(\varepsilon_1) \cdot ... \cdot P(\varepsilon_n) = \frac{e^{-\frac{1}{2}(\varepsilon_1^2 + ... + \varepsilon_n^2)}}{(2\pi)^{n/2}}.$$

When we make the substitution $\eta_\alpha = S_{\alpha\beta}\varepsilon_\beta$ in the integral we must change the element of the integration volume $d^n\varepsilon = d\varepsilon_1 ... d\varepsilon_n$ multiplying it by the Jacobian:

$$d^n\eta = \det\left|\frac{\partial\eta_\alpha}{\partial\varepsilon_\beta}\right| d^n\varepsilon = (\det \mathbf{S})\, d^n\varepsilon.$$

As the determinant doesn't change when the matrix is transposed, $\mathbf{D} = \mathbf{S}\mathbf{S}^T$, and the determinant of the matrix product is equal to the product of the determinants, we have $\det \mathbf{D} = (\det \mathbf{S})^2$. Correspondingly:

$$P(\eta_1, ..., \eta_n) = \frac{e^{-\frac{1}{2}\,\boldsymbol{\eta}\cdot\mathbf{D}^{-1}\cdot\boldsymbol{\eta}}}{(2\pi)^{n/2}\sqrt{\det \mathbf{D}}},$$

where the substitution $\boldsymbol{\varepsilon} = \mathbf{S}^{-1} \cdot \boldsymbol{\eta}$ is made in the exponent:

$$\boldsymbol{\varepsilon}^2 = S^{-1}_{i\alpha}\eta_\alpha\, S^{-1}_{i\beta}\eta_\beta = \eta_\alpha {S^{-1}}^T_{\alpha i}\, S^{-1}_{i\beta}\eta_\beta = \boldsymbol{\eta} \cdot {\mathbf{S}^{-1}}^T \cdot \mathbf{S}^{-1} \cdot \boldsymbol{\eta} = \boldsymbol{\eta} \cdot (\mathbf{S} \cdot \mathbf{S}^T)^{-1} \cdot \boldsymbol{\eta}$$

and the property of inverse matrices was also used: $(\mathbf{A} \cdot \mathbf{B})^{-1} = \mathbf{B}^{-1} \cdot \mathbf{A}^{-1}$ (see p. 247). As any probability density, $P(\eta_1, ..., \eta_n)$ is normalized to one. Thus, it is possible to find the value of the following n-dimensional Gaussian integral taking into account the expression (1.36) for the moment-generating function $\left\langle e^{\mathbf{b}\cdot\boldsymbol{\eta}}\right\rangle$:

$$\int_{-\infty}^{\infty} e^{\mathbf{b}\cdot\boldsymbol{\eta} - \frac{1}{2}\boldsymbol{\eta}\cdot\mathbf{D}^{-1}\cdot\boldsymbol{\eta}}\, d^n\eta = (2\pi)^{n/2}\, \sqrt{\det \mathbf{D}}\; e^{\frac{1}{2}\,\mathbf{b}\cdot\mathbf{D}\cdot\mathbf{b}}. \tag{1.37}$$

Until now we worked with mixed quantities with a zero mean: $\left\langle\boldsymbol{\eta}\right\rangle = \mathbf{S}\cdot\left\langle\boldsymbol{\varepsilon}\right\rangle = 0$. It is possible to add some constant vector $\bar{\eta}_\alpha$ to them, which will have the meaning of the mean values of η_α:

$$\eta_\alpha = \bar{\eta}_\alpha + S_{\alpha\beta}\varepsilon_\beta.$$

Then the general n-dimensional Gauss distribution will be the following:

$$P(\eta_1, ..., \eta_n) = \frac{e^{-\frac{1}{2}\,(\boldsymbol{\eta}-\bar{\boldsymbol{\eta}})\cdot\mathbf{D}^{-1}\cdot(\boldsymbol{\eta}-\bar{\boldsymbol{\eta}})}}{(2\pi)^{n/2}\sqrt{\det \mathbf{D}}},$$

where $\boldsymbol{\varepsilon} = \mathbf{S}^{-1}\cdot(\boldsymbol{\eta}-\bar{\boldsymbol{\eta}})$ is substituted into the probability density $P(\varepsilon_1, ..., \varepsilon_n)$.

• Consider the case $n = 2$ as an example. Let us denote the components of the symmetric matrix $D_{\alpha\beta}$ by three independent constants σ_1, σ_2 and ρ:

$$\mathbf{D} = \begin{pmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

It is not difficult to verify that the determinant $\mathbf{D}$ is equal to

$$\det \mathbf{D} = \sigma_1^2\sigma_2^2(1-\rho^2),$$

and the matrix inverse to $\mathbf{D}$ looks as follows:

$$\mathbf{D}^{-1} = \frac{1}{\det \mathbf{D}} \begin{pmatrix} \sigma_2^2 & -\rho\,\sigma_1\sigma_2 \\ -\rho\,\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

As a result, the joint probability density for η_1, η_2 can be written in the following way:

$$P(\eta_1, \eta_2) = \frac{\exp\{-(x_1^2 - 2\rho\,x_1x_2 + x_2^2)/2(1-\rho^2)\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}},$$

where $x_i = (\eta_i - \bar\eta_i)/\sigma_i$ are the relative deviations of η_i from their means $\bar\eta_i$. The parameters σ_i are volatilities: $\big\langle(\eta_1 - \bar\eta_1)^2\big\rangle = D_{11} = \sigma_1^2$, and ρ is the correlation coefficient: $\rho = \langle x_1x_2\rangle$.

The matrix $\mathbf{D} = \mathbf{S}\mathbf{S}^T$ is symmetric, whereas $\mathbf{S}$ is not symmetric in general case. Therefore, $\mathbf{D}$ depends on three parameters and $\mathbf{S}$ depends on four ones. Several different matrices $\mathbf{S}$ can correspond to the same variance matrix. Thus, we can write:

$$\mathbf{S} = \begin{pmatrix} \sigma_1\cos\alpha & \sigma_1\sin\alpha \\ \sigma_2\sin\beta & \sigma_2\cos\beta \end{pmatrix},$$

where $\rho = \sin(\alpha+\beta)$. Therefore, the same correlation coefficient ρ can be produced by different combinations of angles α and β.

If $\alpha = -\beta$, then $\rho = 0$, and $\mathbf{D}$ is diagonal. For $\sigma_1 = \sigma_2 = 1$ it is a unit matrix. The matrix $\mathbf{S}$, which satisfies the equation $\mathbf{S}\mathbf{S}^T = \mathbf{1}$, is called *orthogonal*.

If $\alpha = 0$, $\rho = \sin\beta$ and $\sigma_1 = \sigma_2 = 1$, then

$$\mathbf{S} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}, \qquad \mathbf{D} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{1.38}$$

Such component mixing transforms the independent standard variables $\varepsilon_1, \varepsilon_2 \sim N(0,1)$, $\langle\varepsilon_1\varepsilon_2\rangle = 0$ into correlated ones, so that $\eta_1, \eta_2 \sim N(0,1)$:

$$\begin{cases} \eta_1 = \ \ \varepsilon_1 \\ \eta_2 = \rho\,\varepsilon_1 + \sqrt{1-\rho^2}\ \varepsilon_2 \end{cases} \qquad => \qquad \big\langle\eta_1\cdot\eta_2\big\rangle = \rho, \qquad \big\langle\eta_1^2\big\rangle = \big\langle\eta_2^2\big\rangle = 1.$$

This allows us, for example, to generate correlated quantities from the uncorrelated ones for computer simulation.

1.7 Additive Random Walk Model

• The coordinate of a Brownian particle in the water or a price in financial market x has a completely irregular trajectory. *The additive independent discrete random walk* is its simplest description. The four adjectives in the model name describe the basic process properties.

Suppose that the initial value $x = x_0$. Then x undergoes $t = 1, 2, ...$ independent random Gaussian changes ("impacts"), each with the volatility σ. As a result, x turns to be equal to the accumulated sum of such changes:

$$x_t = x_0 + \sigma\,(\varepsilon_1 + ... + \varepsilon_t), \tag{1.39}$$

where $\varepsilon_i \sim N(0,1)$ are Gaussian numbers with zero mean and unit variance. For now the index t is an integer, but below we will proceed to a continuous time limit.

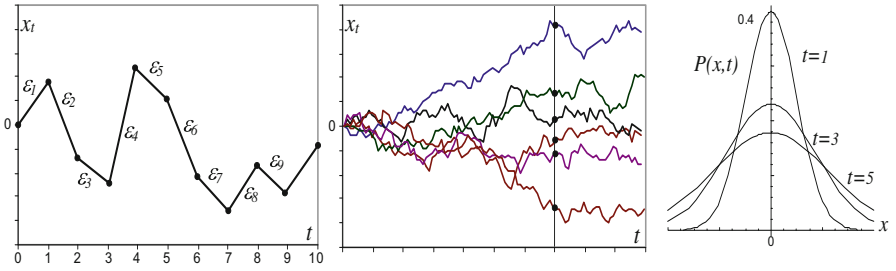
It is convenient to introduce *the discrete Wiener variable*:

$$W_t = \varepsilon_1 + ... + \varepsilon_t = \varepsilon\,\sqrt{t}. \tag{1.40}$$

The second equality above reflects the fact that the sum of t Gaussian numbers is equal to another Gaussian number with the volatility $\sqrt{t}$ (pp. 13–14). Random numbers, both with indexes ε_i, and without them ε, are assumed to be normalized: $\langle\varepsilon\rangle = 0$, $\langle\varepsilon^2\rangle = 1$, i.e. as $\varepsilon \sim N(0,1)$. The model (1.39) now looks as follows:

$$x_t = x_0 + \sigma\, W_t.$$

Let us model the following walk with the help of a computer. Starting from $x_0 = 0$, we generate the random numbers ε_1, ε_2, ... and build their accumulated sum. Such a trajectory is called a *sample trajectory* of the random process (the first figure):



As the changes ε_k will be new every time, the trajectories $x_t = x(t)$ of the walk will be different too (see the second figure). Various realizations of the walk process cross the vertical line $t = const$ at certain values of x. The set of all these numbers is a random variable.

So, when talking about *the process* $x(t)$, we mean that *in the given moment of time*, $x = x(t)$ has certain distribution $P(x)$. The distribution might be different in another moment of time. Thus, the probability density $P(x,t)$, the mean $\bar{x}(t)$ and volatility $\sigma(t)$ are functions of time. The dependence of these characteristics of the random value x on time motivates introduction of the term "process".

The volatility of the random walk increases as $\sqrt{t}$. This can be clearly seen from several realizations of x_t in the second figure above. Their "bunch" widens gradually. As a result, the uncertainty of the future value of x_t grows. We can find x_t quite far from the initial value $x_0 = 0$. This is also shown in the third figure where the probability densities $P(x,t)$ are given. They "spread" gradually in the course of time. Though the maximum of $P(x,t)$ is always at $x = x_0$, nevertheless, the coordinate of the Brownian particle (or the price) might soon be found arbitrary far from its initial value because of the "spreading" of the probability density.

Random walk trajectories start from a certain initial value $x_0 = x(t_0)$ in the moment of time t_0. Therefore, when talking about probabilities we mean the conditional density

$$P(x_0, t_0 \Rightarrow x, t).$$

For now, the moments of time t_0 and t are integers that correspond to the jump number ε_k on the next stage.

It is important to understand that $x_t = x(t)$ is not a specific sample trajectory. It comprises the whole *set* of all trajectories of a random process. By analogy, a random number x doesn't represent a specific value. It contains all possible realizations obeying some distribution $P(x)$. The probability of getting x_t on the t-th step is determined by the probabilities of all changes ε_i. Thus, the discrete Wiener process W_t is determined by the probability density:

$$P(\varepsilon_1, ..., \varepsilon_t) = P(\varepsilon_1) \cdot ... \cdot P(\varepsilon_t),$$

where the equality shows the mutual independence of all ε_i. Therefore, W_t is, in fact, a multidimensional random variable.

Let us pay attention to the meaning of the formula:

$$\varepsilon_1 + ... + \varepsilon_t = \varepsilon \sqrt{t}.$$

Assume that we generate t independent Gaussian numbers ε_i and add them together. The result will have *the same statistical properties* as the single Gaussian number ε with a unit volatility multiplied by the factor $\sqrt{t}$. When studying the properties of the accumulated sum it is sufficient to use the quantity ε instead of the joint density $P(\varepsilon_1, ..., \varepsilon_t)$. In particular, if the mean of the sum of Gaussian numbers is required, its calculation can be simplified by using only one random number. However, some tricks are necessary if we are interested in relations between sums obtained in different moments of time.