5

MONOGRAPH SERIES: M.R. P.N. - MONTE CARLO METHODS: THEORY, ALGORITHMS AND APPLICATIONS TO SELECTED FINANCIAL PROBLEMS

5

MONOGRAPH SERIES

MACIEJ ROMANIUK
PIOTR NOWAK

MONTE CARLO METHODS:
THEORY, ALGORITHMS, AND APPLICATIONS TO SELECTED FINANCIAL PROBLEMS

INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES

MONOGRAPH SERIES
INFORMATION TECHNOLOGIES: RESEARCH
AND THEIR INTERDISCIPLINARY APPLICATIONS

5

MACIEJ ROMANIUK
PIOTR NOWAK

MONTE CARLO METHODS: THEORY, ALGORITHMS AND APPLICATIONS TO SELECTED FINANCIAL PROBLEMS

INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES

Warsaw, 2015

Publication issued as a part of the project:
'Information technologies: research and their interdisciplinary applications',
Objective 4.1 of Human Capital Operational Programme.
Agreement number UDA-POKL.04.01.01-00-051/10-01.

Publication is co-financed by European Union from resources of European Social Fund.

Project leader: Institute of Computer Science, Polish Academy of Sciences

Project partners: System Research Institute, Polish Academy of Sciences, Nałęcz Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences

Editors-in-chief: Olgierd Hryniewicz
Jan Mielniczuk
Wojciech Penczek
Jacek Waniewski

Reviewer: Dariusz Gątarek

Maciej Romaniuk
Systems Research Institute, Polish Academy of Sciences,
Newelska 6, 01-447 Warsaw, Poland and The John Paul II
Catholic University of Lublin
E-mail: mroman@ibspan.waw.pl
WWW: www.researchgate.net/profile/Maciej_Romaniuk

Piotr Nowak
Systems Research Institute, Polish Academy of Sciences,
Newelska 6, 01-447 Warsaw, Poland
E-mail: pnowak@ibspan.waw.pl
WWW: www.researchgate.net/profile/Piotr_Nowak6

Publication is distributed free of charge

ISBN 978-83-63159-14-6
e-ISBN 978-83-63159-15-3

Layout: Piotr Borkowski
Cover design: Waldemar Słonina

Contents

List of Figures

List of Tables

List of Algorithms

Foreword

Maciej Romaniuk, Piotr Nowak

There are many mathematical problems in statistics, economy, finance, biology, physics etc. which can be solved only in an approximate way because of their complexity. Even such "simple" problem as calculation of value of the one-dimensional integral is in many cases too difficult to find the relevant solution in purely analytical form.

Two groups of general approaches to such challenges are commonly used. The first group consists of strictly *deterministic* algorithms, like the widely known Newton's method, which finds roots of a real-valued function. These methods have many advantages, especially their fast convergence in many cases should be emphasized. However, they share also one common disadvantage – the necessity of fulfilling various assumptions before any such method could be even applied. For example, in the case of Newton's method, derivative for the considered function should exist.

The second group of methods consists of algorithms based on numerical simulations, and this group is related to *random* approach. Simulations are especially useful nowadays in our computer era, when relatively cheap and fast hardware and commonly accessible, special-tailored software can be easily used to solve many important practical and theoretical problems in minutes or even in seconds. In Chapter 1 we discuss some advantages and disadvantages of such numerical methods.

Chapter 2 contains some basic definitions, facts and theorems of probability theory and stochastic analysis in continuous time, needed in the chapters that follow. These elements of theory are used for description of simulations algorithms. They are also very useful in Chapter 7, where methods of stochastic analysis are applied to derive and prove the valuation formula of the there described financial instrument. In that chapter there are also many references, which can help the interested readers to familiarize themselves with further details concerning defined notions.

Of course, in order to undertake any random simulations, the efficient way to obtain abundant quantity of *random values* is necessary. These values could be acquired as an effect of some *physical random event* or derived directly as an output from special *computer algorithm*. The physical events mentioned are related to the so called *hardware random number generators*. A simple example

of such a generator is given by tossing a coin. However, practitioners rather rely on specially devised computer algorithms, because of the previously mentioned availability of hardware and software. But, as stated by John von Neumann, "Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin." It is, namely, so that, such special computer algorithm, known as *(pseudo)random number generator*, gives purely deterministic series, which only resembles (in some statistical sense) *independent, identically distributed* sample of random variables. In Chapter 1, the difference between the hardware and the software random number generators are considered in a more detailed way. We also discuss important problems concerning treating deterministic values as random variables. Especially, the statistical approach, based on the sets of various tests, is emphasized.

There are many kinds of simulation algorithms. Some of them are related to transformations of variables between various random distributions known in statistics, like the method to change some variable, which is distributed according to the uniform distribution into the related value from normal distribution. There are also more sophisticated approaches, which are intended to solve real, complex problems, like the previously mentioned issue of the evaluation of integrals. In this case, especially Monte Carlo (MC) or Markov Chain Monte Carlo (MCMC) methods should be emphasized. The first one is based on *iid* sample, the second one is related to mathematical concept of Markov chains. The necessary theoretical background concerning Markov chains is introduced in Chapter 3. In Chapter 4 useful algorithms, various examples and necessary theorems, concerning both the generation of random variables from different distributions and the sampling of trajectories of some stochastic processes are provided. Also the so called *curse of dimensionality* problem, which arises in multidimensional settings, is discussed in a more detailed way. In turn, Chapter 5 is devoted to detailed considerations of MC methods. As previously, practical algorithms, introductory examples and theoretical background for this simulation approach are provided. The same concerns Chapter 6, where MCMC methods are discussed. Moreover, special attention is paid to the significant problem of convergence diagnosis for MCMC methods. Some possible answers to the question when the simulations should be finished are considered there.

Simulation methods are used in order to solve important problems from various scientific and practical areas, like e.g. statistics, physics, biology etc. In this book, though, we focus on applying Monte Carlo methods in financial mathematics.

In Chapter 7, we consider pricing of a new kind of financial instrument, which is known as *catastrophe bond* (or *cat bond*). Nowadays, catastrophic events like hurricanes, floods and earthquakes are serious problems for the reserves of many insurers. The losses from such catastrophes can even lead to bankruptcy of insurance enterprises. Catastrophe bond is one of the solutions to deal with these problems, as this derivative is used to transfer risk from the insurer onto the financial markets. We derive and prove a generalized version of the catastro-

phe bond pricing formula. Applying stochastic processes in continuous time, we propose theoretical models of catastrophe losses and risk-free interest rates. Moreover, we assume a general form of cat bond payoff function. Using methods of stochastic analysis and financial mathematics, we derive the instrument's valuation expression, which can be applied for various types of cat bond payoff functions and affine interest rate models.

Because of complex pricing formulas developed in Chapter 7, simulations are necessary to estimate the relevant expected values. Numerical analysis of cat bond prices, based on Monte Carlo methods, is presented in Chapter 8. Moreover, Monte Carlo methods are also used to solve the important problem of probability of bankruptcy of the insurer in this part of the book. In order to do this, the portfolio of the insurer, which consists of a few layers is modelled there. As these layers we apply the classical risk process based on the insurance premiums, the reinsurance contract and the payments related to the specially tailored catastrophe bond. Additionally, one-factor Vasicek model of the interest rate is embedded in the structure of such portfolio. Then, by applying numerical simulations of catastrophic events calibrated to real-life data, the stability of the constructed portfolio is evaluated. Using methods of descriptive statistics, outputs of various scenarios based on simulations are analysed. In particular, probabilities of bankruptcy for some interesting cases are estimated, as the key factors for the insurer during the process of construction of the related portfolio.

1

Introduction to (pseudo)random number generation

Maciej Romaniuk

1.1 Basic concepts of (pseudo)random number generators

We start from introducing some basic concepts and definitions concerning so called (pseudo)random number generators – special algorithms, which are used to imitate behaviour of random samples, but are based only on strictly deterministic sequence of numbers.

1.1.1 John von Neumann's Middle-Square Method

Historically, one of the first (pseudo)random number generators is the algorithm known as *John von Neumann's Middle-Square Method*. It was introduced in 1949 (see (von Neumann 1951)).

A sequence of numbers with m digits is created as an output of this method. The parameter m is always even. This algorithm can be described by the following steps:

Algorithm 1.1 (John von Neumann's Middle-Square Method)**.**

1. Take the previously generated number X_{n-1}.
2. Calculate $Y_n = X_{n-1}^2$.
3. If it is necessary, add additional zeroes at the beginning of the Y_n, so that Y_n has $2m$ digits.
4. Take middle m digits of Y_n as the new output X_n.

It can be easily seen that von Neumann's algorithm may be described by the general formula

$$X_n = f(X_{n-1}) \ , \tag{1.1}$$

where $f(.)$ is some fixed mathematical function. Therefore the output of this algorithm is a strictly deterministic sequence of numbers. But, because of their "irregular" behaviour, these values could be regarded as "random sample". It is easy to see that, in order to start such method, it is necessary to choose some starting value X_0, which is known as *seed* or *initialization value*.

It should be noted that there are some problems with treating the output of the Algorithm 1.1 as "random values". The generated numbers very frequently decrease to the sequence of zeroes. This algorithm can also get stuck on a number other than zero (e.g. for $m = 2, X_0 = 10$). Therefore, nowadays, this method has a rather historical significance and more complex and better algorithms are applied.

Additionally, in the case of all methods described by the general formula (1.1), there is a phenomenon known as *period* of the algorithm. For the output given be the sequence

$$X_0, X_1, X_2, \ldots, X_n, X_{n+1}, \ldots, X_{n+p}, X_{n+p+1}, \ldots \tag{1.2}$$

there are always parameters n and p, so that $X_i = X_{i+jp}$ for $i \geq n$ and $j = 1, 2, \ldots$. It means that the generated numbers are continually repeated in the same sequence of p values, which is called the period of the generator. Therefore, only aperiodic part of the whole output $X_0, X_1, \ldots, X_{n+p}$ (or more precisely – output without the starting value X_0) may be treated as approximation of random numbers.

1.1.2 Hardware and software generators

The methods which are used to generate "random samples" can be divided into two groups: hardware (or physical) generators and software generators (computer algorithms).

A *hardware random number generator* is a device that generates random numbers based on some kind of physical process (e.g. thermal noise or the photoelectric effect or other quantum phenomena, see e.g. (Davies 2000)). This physical process creates "noise" which is converted afterwards into sequence of numbers (e.g. binary digits 0 or 1). These numbers may be treated as the statistically random output. In theory, such output is "truly" random because it is completely unpredictable. Also more "macroscopic" processes like coin flipping or dice may be used as the source of the random output.

However, hardware random number generators are often relatively slow, i.e. they create limited size of random sequence per unit time. Additionally, there may be some problems with stability of the output of such devices. Even a small change in physical parameters of environment of the device may influence the generated sequence. Therefore, usually hardware generators are rather used to initialize (i.e. to create the seed) for software (pseudo)random number generators. A value taken from system clock, which may be found in every computer, is an example of such approach.

Usually, instead of hardware random number generators, *(pseudo)random number generator* (abbreviated as PRNG), also known as a deterministic random bit generator (DRBG), is used to conduct simulations. PRNG is an algorithm which generates a sequence of numbers that seems to "be random" (see

(Law 2007, Wieczorkowski & Zieliński 1997)). The "randomness" may be identified with satisfying some statistical goodness-of-fit and independence tests (see Definition 1.3 for a more detailed approach).

Generally speaking, software generator is started from an arbitrary starting state (*seed*) and afterwards some mathematical function determined by the algorithm is used to obtain subsequent values as in the case of the Algorithm 1.1 and the general formula (1.1). If we know this seed and the applied method, the corresponding output is completely anticipated.

PRNGs are usually much faster than hardware generators and because only computer is necessary (instead of some special device), usually the software generators are used for modern simulations.

Software generators have additional advantage in that the generated output could be reproduced using the same seed for each run. It is often necessary to test what happens when some parameters of the simulations are altered but exactly the same sequence of generated values is used in many different simulation runs. In other cases, it must be possible to stop a simulation and then re-continue the same analysis. Therefore, the current state of the PRNG should be stored in memory to use this value during the next run.

1.1.3 Stages of generation

We introduce first an important definition of uniform distribution.

Definition 1.2 (Uniform distribution)**.** *A distribution of the random variable X is a uniform distribution on the interval $[a, b]$ (further we use notation $X \sim U\,[a, b]$) if the density of this distribution on $[a, b]$ is given by*

$$f(t) = \frac{1}{b-a}\ .$$

If $X \sim U\,[a, b]$, then $\mathbb{E}\,X = \frac{a+b}{2}$, $\operatorname{Var} X = \frac{(b-a)^2}{12}$.

The whole simulation process may be divided into the following steps.

During the first step, usually, the numbers generated by PRNG are values from some subset of $\mathbb{N}$, e.g. $0, 1, 2, \ldots, m$ (as in the case of linear congruential generator, see Section 1.2.1). But in many applications and more sophisticated algorithms (e.g, see Chapter 4, Chapter 7 and Chapter 8) we need random variables distributed uniformly on the interval $[0, 1]$. In such a case, during the second step, the output from PRNG, denoted by $X_1, X_2, \ldots$, is transformed into a new sequence $U_1, U_2, \ldots$ using straightforward formula

$$U_i = \frac{X_i}{m} \tag{1.3}$$

or a similar one (e.g. $U_i = \frac{X_i}{m+1}$). It should be noted that contrary to the random sample from the "true" uniform distribution, in the case of the numbers

obtained via PRNG the probabilities of a single value (e.g. like 0 or 1) may be not uniformly equal to zero.

In order that the new points $U_1, U_2, \ldots$ be sufficiently "densely" distributed on the interval $[0, 1]$, the value of the parameter m should be large. As it will be described in Section 1.2.1, the larger values of m are sometimes also necessary to a obtain longer period of some PRNGs.

During the third step, the sequence $U_1, U_2, \ldots$ is transformed into the values, which may be identified with samples from other random distributions, e.g. normal distribution or exponential distribution. In this case, some special algorithms are applied which are based on mathematical theorems. The examples of the relevant methods are described in Chapter 4 in a more detailed way.

Then, finally, the obtained output is used for some numerical simulations or procedures, e.g. as in the case of Monte Carlo or Markov Chain Monte Carlo methods (see Chapter 5 and Chapter 6).

1.1.4 Problems with "randomness"

Von Neumann said that "Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin." As previously stated, the output generated by PRNG is entirely deterministic, because it is based on some mathematical function (in other words – some deterministic algorithm). Therefore, there are significant problems with "randomness" of such sequences (see e.g. (Law 2007, Thompson 2000, Wieczorkowski & Zieliński 1997)), which could be related to:

- length of the period,
- existence of spatial structure (i.e. problem with multidimensional distributions),
- accordance with *iid* requirement (i.e. the output may be considered as independent, identically distributed random variables).

Now we consider these issues in a more detailed way.

In order to approximate a random sample, the period of the considered algorithm should be longer than the required number of simulations n in each numerical experiment. Many authors insist that the length of the period p should be much greater than the value n. Some of them state that $n < p^{2/3}$ (see (Maclaren 1992)) or even $n < p^{1/2}$ (see (Ripley 1987)), especially for linear generators (see Section 1.2.1 for the description of this type of software generators). We will discuss some theorems, concerning the length of the period in Sections 1.2.1 and 1.2.2. However, we should keep in mind that apart from the considered method, the length of the period also depends on the applied seed. It is known that the quality of some of the algorithms considered as "good" PRNGs can be lowered if the seed is not appropriately chosen. Therefore, for more complex methods the practical examination of the generated output is necessary.

Another issue is related to multidimensional distribution of the obtained output. As indicated in Section 1.1.3, we are usually interested in the sequence

$U_1, U_2, \ldots$ from the interval $[0, 1]$. For the fixed parameter $d \geq 1$ we may consider sequence of points in the d dimensional space described by coordinates

$$(U_1, U_2, \ldots, U_d), (U_2, U_3, \ldots, U_{d+1}), \ldots \tag{1.4}$$

If the sequence $U_1, U_2, \ldots$ is "sufficiently" random, then the points (1.4) should be "uniformly" and "randomly" distributed in the d dimensional unit cube (as in the case of Figure 1.1). But for some PRNGs, these points have often more regular and not "chaotic" spatial structure. Then we could see some "crosses", "lines" and other geometrical shapes instead of the irregular "cloud" of points (as in the case of Figure 1.2). Apart from the coordinates described by (1.4), we may also analyse the spatial behaviour of other sets, e.g.

$$(U_1, U_2, \ldots, U_d), (U_{d+1}, U_3, \ldots, U_{2d}), \ldots$$

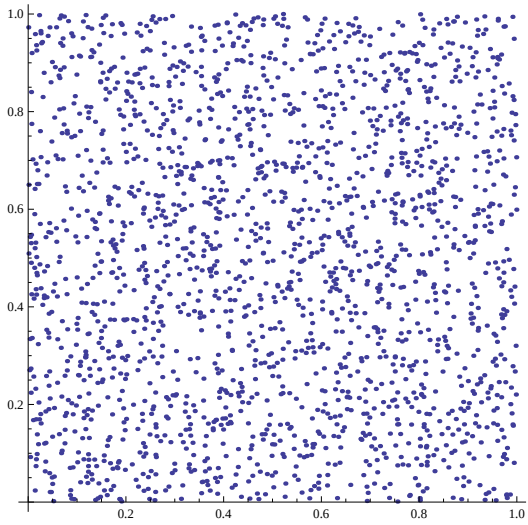

Fig. 1.1. Example of "random" spatial structure

The third problem is related to the statistical notion of "random sample". Let us suppose that we need a generator which generates random values from some fixed random distribution. It means that the output from this PRNG has to satisfy two important conditions, known from mathematical statistics and probability theory:

- the values $(X_i)_{i=1}$ constitute *independent* (from probabilistic point of view) sequence,
- the values $(X_i)_{i=1}$ are consistent with a fixed random distribution, i.e. they are identically distributed.

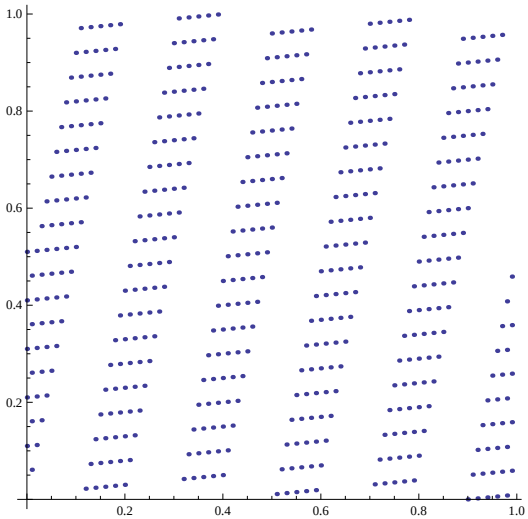

Fig. 1.2. Example of "lines" produced by PRNG

In such case we may say that $(X_i)_{i=1}$ is *iid* (i.e. independent, identically distributed) sequence.

There are many statistical methods for checking the above mentioned conditions. We can briefly mention some of them (see also (Law 2007, Wieczorkowski & Zieliński 1997)): histogram and other graphs, descriptive statistics, correlation coefficient, time series analysis, statistical tests etc.

Let us suppose that our sequence should be distributed according to $U\ [0,1]$. Then the histogram for such data should be accordingly "flat". In case of other distributions, we could also check the shapes of the corresponding histograms. Other useful graph in this setting is a box-and-whisker plot which could be used to check skewness of the empirical distribution or to find possible outliers.

If the random distribution is fixed, then we could compare values of theoretical moments like average $\mathbb{E}\,X$ and variance $\mathrm{Var}\,X$ with corresponding empirical estimators like mean $\bar{X}$ and empirical variance S_X^2 of the generated sequence.

Another tool is Pearson's r (Pearson's correlation coefficient) which may be used to check if there is a linear correlation between two sets of output from PRNG. In such a case the sequence $(X_i)_{i=1}$ can be divided into two parts to find the relevant value

$$\rho = \frac{\frac{1}{n/2}\sum_{i=1}^{n/2}\left(X_i - \bar{X}_1\right)\left(X_{n/2+i} - \bar{X}_2\right)}{\sqrt{\frac{1}{n/2}\sum_{i=1}^{n/2}\left(X_i - \bar{X}_1\right)^2 \frac{1}{n/2}\sum_{i=n/2+1}^{n}\left(X_i - \bar{X}_2\right)^2}}\ ,$$

where $\bar{X}_1$ is average for the first part of $(X_i)_{i=1}$ and $\bar{X}_2$ is average for the second one. If ρ is close to zero, then these two sets are uncorrelated. It means that there

is no linear correlation between these two sequences, but there may be other kind of probabilistic dependency between them.

Time series analysis could be also applied. For example, we may analyse the simple model given by the equation

$$X_n = aX_{n-1} + b$$

and then check the corresponding autocorrelation coefficients.

There are also many statistical tests for checking if the sequence $(X_i)_{i=1}$ is distributed according to the specified distribution and / or if the values are independent. Such tests as χ^2 goodness-of-fit test or Kolmogorov–Smirnov test should be mentioned here.

The tests and methods referred to may be also classified otherwise:

- Theoretical tests which are based on theoretical (i.e. mathematical) properties of the algorithms. These tests are mathematically strict but they are often very difficult to perform and they give only asymptotic conclusions.
- Empirical tests are based on testing the algorithms and their implementations by producing the relevant output. These tests are suitable for all algorithms, but it is often difficult to say how extensive testing is sufficient (i.e. how many simulations should be generated) and whether variation of some parameters of the algorithm has any influence on the results of these tests.
- Visual tests can be used to directly locate some "deviations" from the "randomness" on graphs.

1.1.5 Definition of a "good" PRNG

Taking into account the above mentioned problems with "randomness", we introduce the "working definition" following the one presented by Wieczorkowski & Zieliński (1997).

Definition 1.3. *We say that the given algorithm generates (pseudo)random values, if none of the selected statistical tests rejects the hypothesis that the generated sequence* $(X_i)_{i=1}$ *is an* iid *sample from the specified random distribution.*

This definition is very useful from the practical point of view, because we focus only on the methods of checking the "randomness", instead of philosophical considerations. However, there are also some problems caused by such approach. The first one is related to the selection of the set of statistical tests. It is possible that all the tests applied earlier did not reject the *iid* hypothesis, but the new, additional test rejects it. The second one is the issue of the selection of the sequence. As previously, it is possible that all the sequences analysed earlier did not lead to the rejection of the *iid* hypothesis, but the next sequence (or sequence generated for the new set of parameters) gives rise to rejection of this statement.

Therefore, it is necessary to select the relevant, wide set of statistical tests (e.g. like "Diehard", "Crypt-XS" or "DieHarder") and to check the whole collection

of possible sequences and parameters of the considered generator. For example, the "Diehard" suite (see (Marsaglia 1995)) consists of the tests: birthday spacings, overlapping permutations, ranks of 31x31 and 32x32 matrices, ranks of 6x8 matrices, monkey tests on 20-bit words, monkey tests OPSO, OQSO, DNA, count the 1's in a stream of bytes, count the 1's in specific bytes, parking lot, minimum distance, random spheres, squeeze, overlapping sums, runs, the craps.

The birthday spacings test is conducted as follows (see, e.g. (L'Ecuyer & Simard 2001)):

1. Partition the interval $[0, 1)$ into d equal segments. This determines a partition of $[0, 1)^t$ into $k = d^t$ cubic boxes of equal size.
2. Number the boxes from 0 to $k - 1$ in the natural order: the box with lower corner at $(i_0/d, i_1/d, \ldots, i_{t-1}/d)$ has number $i_{t-1} + di_{t-2} + \ldots + d^{t-1}i_0$.
3. Generate nt values $U_0, \ldots, U_{nt-1}$
4. Place the points $V_i = (U_{ti}, \ldots, U_{ti+t-1})$ for $i = 0, 1, \ldots, n - 1$ in the corresponding boxes.
5. Let $I_1 \leq I_2 \leq \ldots \leq I_n$ be the numbers of the boxes where these points have fallen (sorted by increasing order).
6. Compute the *spacings* $S_j = I_{j+1} - I_j$.
7. Let Y be the number of collisions between these spacings, i.e. the value of events such that $S_{(j+1)} = S_{(j)}$ where $S_{(1)}, S_{(2)}, \ldots, S_{(n-1)}$ are the sorted spacings.

This experiment is similar to the situation, when we have birthdays of n people and the year has k days, therefore this procedure is called *birthday spacings* or *discrete spacings* test. If n is large and $\lambda = \frac{n^3}{4k}$ is small, Y should follow approximately the Poisson distribution with mean λ.

It should be mentioned that the most important PRNGs are constantly being checked using various tests and different outputs. It is known that some of "good" algorithms were found to lack "randomness" in the case of e.g. some seeds (like Mersenne-Twister algorithm, see Section 1.2.6).

As mentioned earlier, the other approach requires the mathematical reasoning leading to the conclusion that the algorithm generates random numbers. However, in many cases this more exact and certain way is not possible, because the relevant theorems are not (or even could not be) proved.

1.2 Examples of software generators

As indicated in Section 1.1.3, the whole simulation procedure may be divided into a few stages: firstly the output given by values from some subset of $\mathbb{N}$, then the output from the uniform distribution, next the transformation to the specified probability distribution and then application of the obtained sequence in the numerical procedure.

In further considerations (especially in case of more complicated generators – see Chapter 4, and MC and MCMC methods – see Chapters 5 and 6) we

will assume that we dispose of an efficient PRNG for the uniform distribution. Therefore, in this section we discuss various simulation algorithms and present some examples of PRNGs, keeping in mind that the output from the considered method should be transformed into the sequence $U_1, U_2, \ldots, U_n$ of variables from the interval $[0, 1]$ (e.g. via formula (1.3)).

1.2.1 Linear congruential generators

PRNG described by the function

$$X_{n+1} = (a_1 X_n + a_2 X_{n-1} + \ldots + a_k X_{n-k+1} + c) \mod m \tag{1.5}$$

is the generalized form of the *linear congruential generator (LCG)* , where constants $a_1, a_2, \ldots, a_k, c, m \in \mathbb{Z}$ are parameters of this generator.

The simplest generator of this form is given by the formula

$$X_{n+1} = (aX_n + c) \mod m \tag{1.6}$$

and it was proposed by Lehmer in 1949 (see (Lehmer 1951)). This very simple PRNG is often used in standard compilers and programming libraries even nowadays. If $c = 0$, then such generator is called *multiplicative*, and for $c \neq 0$ it is known as *mixed* generator.

In the case of the generator given by the formula (1.6), the length of the period p is equal to $\min\{i > 0 : X_i = X_0\}$. We always have $p \leq m$, because for the operator $\mod m$ there are only m remainders given by $0, 1, \ldots, m-1$, so p could not be greater than m. Therefore, in practical applications m should be a "really big" number. However, in many cases the period is actually significantly smaller than its upper limit m.

There are various lemmas and theorems concerning the length of the period for LCG (see (Jansson 1966, Knuth 1966, Ripley 1987, Wieczorkowski & Zieliński 1997)). We mention only some of them.

Lemma 1.4. *The period of the multiplicative generator*

$$X_{n+1} = (aX_n) \mod m$$

is maximal for $m = 2^L$ (if $L \geq 4$) and is equal to 2^{L-2}, if and only if, X_0 is odd and $a = 3 \mod 8$ (i.e. a is divided by 8 with remainder 3) or $a = 5 \mod 8$.

The proof of this lemma may be found in (Hull & Dobel 1962).

Example 1.5. *RANDU generator: $a = 2^{16} + 3, m = 2^{31}$.*

Lemma 1.6. *The period of the mixed generator* (1.6) *is maximal, if and only if, all of the following conditions are fulfilled:*

1. *c and m do not have common divisors,*
2. *$a = 1 \mod p$ for each prime factor p of m (i.e. for each prime number that divides m exactly),*

3. $a = 1 \mod 4$, *if 4 divides* m.

Example 1.7. *Let* $a = 69069, c = 1, m = 2^{32}$.

It can be easily seen that in the case of the more general formula (1.5), the seed of such a generator is given by the whole sequence $X_0, X_1, \ldots, X_{k-1}$, instead of only one value, as for (1.6).

There are many ways to generalize the linear congruential generators. For example, we may consider a multivariate function given by

$$\mathbf{X}_{n+1} = \mathbf{A}\mathbf{X}_n \mod m \ , \tag{1.7}$$

where $\mathbf{X}_0, \mathbf{X}_1, \ldots$ are some vectors in $\mathbb{R}^k$ and $\mathbf{A}$ is $k \times k$ matrix. In this case, the operation $\mod m$ is done for each term of the output vector $\mathbf{A}\mathbf{X}_n$. Generators given by (1.7) are especially useful in the case of multivariate uniform distribution on k-dimensional cube.

1.2.2 Fibonacci generators

The famous Fibonacci sequence is described by the formula

$$X_n = X_{n-1} + X_{n-2} \text{ for } n \geq 2 \ ,$$

if $X_0 = X_1 = 1$. A similar function (see (Taussky & Todd 1956))

$$X_n = (X_{n-1} + X_{n-2}) \mod m \text{ for } n \geq 2 \tag{1.8}$$

is used to generate numbers, which are approximately "random" values (in the sense of Definition 1.3). The statistical tests do not reject the hypothesis that the output sequence $(X_i)_{i=1}$ is derived from the uniform distribution, but some tests reject the hypothesis that the generated variables are independent.

Therefore, the function (1.8) was generalized into the form

$$X_n = (X_{n-s} + X_{n-r}) \mod m \text{ for } n \geq r, r > s \geq 1 \ , \tag{1.9}$$

which is known as *Additive Lagged Fibonacci Generator (ALFG)*. In past, ALFG were rarely used, because they are slower than simple LCG given by (1.6) and they need more computer memory – more of previous values should be stored to generate the subsequent output. Nowadays, these generators are applied more frequently, because the formula (1.9) introduces more "randomness" than the simple algorithm (1.6).

ALFG can have longer period than LCG (e.g., see (Wieczorkowski & Zieliński 1997)). For $m = 2^L$ the maximal period of ALFG is equal to $(2^r - 1)2^{L-1}$.

The generators described by (1.9) may be generalized to the form

$$X_n = (X_{n-s} \diamond X_{n-r}) \mod m \ , \tag{1.10}$$

where $\diamond$ is some arithmetical operation (e.g. subtraction, multiplication). In case of multiplication, this leads to *Multiplicative Lagged Fibonacci Generator (MLFG)* given by the formula

$$X_n = (X_{n-s} \cdot X_{n-r}) \mod m \ . \tag{1.11}$$

Maximal period for MLFG is equal to $(2^r - 1)2^{L-3}$. For subtraction, the maximal period is the same as for ALFG. Proofs of the results mentioned may be found in Marsaglia (1984) and Marsaglia & Tsay (1985).

1.2.3 Generalised Feedback Shift Register generators

Let us consider the sequence of bits described by the formula

$$b_n = (a_1 b_{n-1} + a_2 b_{n-2} + \ldots + a_k b_{n-k}) \mod 2 \ , \tag{1.12}$$

where $a_1, a_2, \ldots, a_k$ are some binary constants (i.e. the parameters which are equal to 0 or 1) and $b_0, b_1, \ldots, b_{k-1}$ is the binary seed for this generator. Instead of operation modulo 2 for bits, logical xor (exclusive-or) in the formula (1.12) can be used. The xor operator is more straightforward to implement in the case of compilers. If only $a_{j_1} = a_{j_2} = \ldots = a_{j_l} = 1$ and all other a_j's are equal to 0, then (1.12) may be written down as

$$b_n = b_{n-j_1} \operatorname{xor} b_{n-j_2} \operatorname{xor} \ldots \operatorname{xor} b_{n-j_l} \ .$$

Obviously, there are only 2^k various arrangements of the sequence $(b_1, b_2, \ldots, b_k)$. Therefore, the maximal period of the generator given by the formula (1.12) is equal to $2^k - 1$ (not 2^k, because if there are only zeros in the output, all of the consecutive values are also equal to 0). There are many algebraic methods for finding the parameters, which then give the maximal period (see e.g. (Golomb 1967)).

The simplest case of (1.12) is described by the formula

$$b_n = b_{n-p} \operatorname{xor} b_{n-q} \ , \tag{1.13}$$

where only two parameters a_j are not equal to zero. There are also some strict results concerning the maximal period in such a case.

Example 1.8. *If $p = 89$ and $q = 38$, then period of* (1.13) *is maximal (see (Berdnikov, Compagner & Turtia 1996)).*

Next step is the transformation of the bit sequence $b_1, b_2, \ldots$ into the sequence of real numbers from the interval $[0, 1]$. Tausworthe (1965) proposed the formula (known as *Tausworthe generator*) given by

$$U_n = \sum_{j=1}^{L} 2^{-j} b_{ns+j} = 0.b_{ns+1} \ldots b_{ns+L} \tag{1.14}$$

for $n = 0, 1, \ldots$, where s and L are some constants fulfilling the inequality $s \leq L$. If $s < L$, then the consecutive values U_n are partially based on the same bit sequences and if $s = L$, there is no such overlapping. This method is easy to implement in a computer algorithm using shift registers and xor operator.

The generated bis sequence can be also transformed into integer values. Lewis & Payne (1973) proposed the formula

$$Y_n = b_n b_{n-l_1} \ldots b_{n-l_{L-1}}$$

to obtain L-bit words (which may be treated as integer values), where $l_1, \ldots, l_{L-1}$ are some fixed delays. If we apply the simplest formula (1.13) for this case, we get the *Generalised Feedback Shift Register (GFSR)* generator, described by

$$X_n = X_{n-p} \operatorname{xor} X_{n-q} \ . \tag{1.15}$$

To obtain the maximal period of this generator the best choices for p and q are Mersenne primes (i.e. the prime numbers of the form $2^p - 1$) which satisfy the condition that $p^2 + q^2 + 1$ is also prime.

Example 1.9. *In the case of R250 generator, we have $p = 250$ and $q = 103$ (see (Kirkpatrick & Stoll 1981)). This is one of the most commonly used generators of this class, which generates 31-bit integers. To initialize the algorithm, 250 uncorrelated seeds (random integers) are necessary. The length of its period is equal to $2^{250} - 1$.*

1.2.4 Nonlinear generators

The generators described in the previous sections were related to linear recurrence formulas or they were based on generalizations of such functions. This approach has some important advantages and disadvantages. We should mention, in particular, that the linear generators are easy to implement, which was especially important in the case of the first, slow computers, but there are problems with the spatial structure of the output. It means that generated values are concentrated only on some hyperplanes of multidimensional cube (as discussed in Section 1.1.4). Therefore, new kinds of generators are also considered, where instead of the linear function – some nonlinear one is applied.

Eichenauer & Lehn (1986) introduced the generator given by the formula

$$X_n = \left(a\check{X}_{n-1}^{-1} + b\right) \mod m \ , \tag{1.16}$$

where m is a prime number and $\check{X}^{-1}$ denotes *modular multiplicative inverse*. This operator is defined as follows:

Definition 1.10. *For $x = 0$ we have $\check{x}^{-1} \mod m = 0$, and for $x \neq 0$ the value $\check{x}^{-1}$ is given by the condition $(x \cdot \check{x}^{-1}) \mod m = 1$.*

Lemma 1.11. *The length of the period of the generator* (1.16) *is maximal and is equal to m if $m^2 - 1$ is the minimal integral, for which $z^{m^2-1} = 1 \mod \left(z^2 - bz - a\right)$.*

Modular multiplicative inverse is also used in the generator, proposed by Eichenauer-Hermann (1993), given by the formula

$$X_n = \left(a(n + n_0) + b\right)^{-1} \mod m \ , \tag{1.17}$$

where a, b, n_0, m are parameters of this algorithm. It can be easily seen, in the case of (1.17) the new value X_n may be found regardless of the previous values $X_1, \ldots, X_{n-1}$. Therefore, this generator may be used for parallel simulations.

Another type of nonlinear generator was considered by Blum, Blum & Shub (1986). In this case, the quadratic function given by the formula

$$X_n = X_{n-1}^2 \mod m$$

is applied.

1.2.5 Combination generators

Instead of applying more complex functions, the other approach to obtain the generator with "good quality" is to combine two or more simpler PRNGs into a new algorithm. In such a case, the output generated by the combined generators can be "more independent", "more uniform" and it usually has a longer period. For example, if sequence $(X_i)_{i=1}$ has period p_1, sequence $(Y_i)_{i=1}$ has period p_2, and p_1 and p_2 are relatively prime, then the new sequence

$$Z_i = X_i \diamond Y_i \ ,$$

has period $p_1 p_2$ (see, e.g. (Graham, Knuth & Patashnik 1994)). The operator $\diamond$ may be, e.g., addition, modulo operator, xor operator, etc.

1.2.6 Mersenne Twister generator

Matsumoto & Nishimura (1998) developed a very fast and reliable PRNG called *Mersenne Twister generator (MT)*. As authors stressed, this generator has many advantages. The length of the period is equal to $2^{19937} - 1$. The generated sequence is 623-distributed to 32 bits accuracy. This second feature is described by the following definition:

Definition 1.12. *A (pseudo)random sequence x_i of w-bit integers of period p satisfying the following condition is said to be k-distributed to v-bit accuracy: let* $\mathrm{trunc}_v(x)$ *denote the number formed by the leading v bits of x, and consider p of the kv-bit vectors*

$$\left(\mathrm{trunc}_v\left(x_i\right), \mathrm{trunc}_v\left(x_{x+1}\right), \ldots, \mathrm{trunc}_v\left(x_{x+k-i}\right)\right) \ ,$$

where $0 \leq i \leq p$. Then, each of the 2^{kv} possible combinations of bits occurs the same number of times in a period, except for the all-zero combination that occurs once less often. For each $v = 1, 2, \ldots, w$, let $k(v)$ denote the maximum number such that the sequence is $k(v)$-distributed to v-bit accuracy.

The equidistribution property has geometrical interpretation as the same number of points in the cubes which are given by partitioning of the unit cube (see e.g. (Matsumoto & Nishimura 1998)) except for the cube located next to the origin of the axes. The MT generator has passed diehard tests, Load Tests and Ultimate Load Tests, but there are problems with passing some of the tests from the library TestU01 (see (L'Ecuyer & Simard 2007)) Additionally, it is known that MT is sensitive to poor initialization and can take a long time to recover from a zero-excess initial state.

The generator is based on idea of twisted GFSR, incomplete array and matrix linear recurrence over a finite binary field. Now we briefly described some further details.

The MT algorithm is using *word vectors*, i.e. w-dimensional row vectors of bits. The most important part of the generator is described by the following recurrence formula

$$x_{n+k} = x_{m+k} \oplus \left(x_k^u | x_{k+1}^l\right) \mathbf{A} \ , \tag{1.18}$$

where $k = 0, 1, \ldots$, the integers n and $1 \leq m \leq n$ are parameters of the generator, and $\mathbf{A}$ is a constant $w \times w$ binary matrix. The initial seed is given by the word vectors of the size w, namely $x_0, x_1, \ldots, x_{n-1}$. To obtain the new values $x_n, x_{n+1}, \ldots$, the recurrence formula (1.18) is used for $k = 0, 1, \ldots$. For example, in the case of $k = 0$ we have

$$x_n = x_m \oplus \left(x_0^u | x_1^l\right) \mathbf{A} \ .$$

The operator x_k^u in (1.18) denotes "the most significant (the upper) $w-r$ bits" of x_k and x_{k+1}^l states for "the least significant (the lower) r bits" of x_{k+1}, therefore, for the row vector $x = \left(x^{(w)}, x^{(w-1)}, \ldots, x^{(1)}\right)$ we get

$$x^u = \left(x^{(w)}, x^{(w-1)}, \ldots, x^{(r+1)}\right) \ , \ x^l = \left(x^{(r)}, x^{(r-1)}, \ldots, x^{(1)}\right) \ .$$

The constant r is the parameter of the generator. The operator $\left(x_k^u | x_{k+1}^l\right)$ in (1.18) is the concatenation of the vectors x_k^u and x_{k+1}^l, so

$$\left(x_k^u | x_{k+1}^l\right) = \left(x_k^{(w)}, x_k^{(w-1)}, \ldots, x_k^{(r+1)}, x_{k+1}^{(r)}, x_{k+1}^{(r-1)}, \ldots, x_{k+1}^{(1)}\right).$$

In the next step of the algorithm, the matrix $\mathbf{A}$ of the form

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & \ldots & 1 \\ a_w & a_{w-1} & a_{w-2} & \ldots & a_1 \end{pmatrix} \ ,$$

where $a_1, a_2, \ldots, a_w$ are binary constants, is multiplied from the right by the vector $\left(x_k^u | x_{k+1}^l\right)$. Because of the special form of $\mathbf{A}$, this multiplication reduces

to shiftright and bitwise addition modulo two, denoted by operator $\oplus$ in (1.18). Then, the vector x_{m+k} is added (also via the bitwise addition modulo two) to the previously obtained value.

The next important step in the generator, after evaluation of the output vector $x = x_{n+k}$, is known as *tempering*. This transformation is done by the following formulas

$$\begin{aligned} y &= x \oplus (x >> u) \\ z &= y \oplus ((y << s) \text{ AND } b) \\ q &= z \oplus ((z << t) \text{ AND } c) \\ v &= q \oplus (q >> l) \ , \end{aligned}$$

where l, s, t, u are integers, b, c are suitable bitmasks of word size, the operator $x >> u$ denotes the u-bit shiftright of the vector x and, similarly, $x << u$ denotes the u-bit shiftleft.

Because of the previously mentioned problems with the MT algorithm, the generator was improved by Saito & Matsumoto (2008) and it is known as SIMD-oriented Fast Mersenne Twister (abbreviated as SFMT). SFMT is a Linear Feedbacked Shift Register (LFSR) generator that generates a 128-bit pseudorandom integer at one step. This algorithm has also many advantages: it is roughly twice as fast as the Mersenne Twister, it has a better equidistribution property of v-bit accuracy than MT, it has quicker recovery from zero-excess initial state than MT and it supports various periods from $2^{607} - 1$ to $2^{216091} - 1$.

2

Short introduction to probability theory and stochastic analysis in continuous time

Piotr Nowak

The methods and results, presented in this book, require application of probability theory and stochastic analysis. Especially, continuous time stochastic processes and theory concerning equivalent change of probability measure are applied to the catastrophe bonds pricing approach, presented in Chapter 7. The first section of the present chapter is an exposition of basic notions used in probability theory, including probability space, random variables, their distributions and independence. In Section 2.2 stochastic processes in continuous time are defined. Section 2.3 is devoted to Lévy processes. Its two subsections are concerned with Wiener and Poisson processes. We also describe the space of processes integrable with respect to Wiener process and define Itô processes. In the last section of this chapter we present some facts, connected with equivalent probability measures, in particular the Girsanov theorem, which plays an important role in financial mathematics.

2.1 Basic notions of probability theory

2.1.1 Probability space

We denote by the symbols $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{N}$ the sets of real numbers, non-negative real numbers and positive integers, respectively. For each subset A of a set Ω the symbol I_A denotes the function defined by the equality

$$I_A(\omega) = \begin{cases} 1 \text{ for } \omega \in A \\ 0 \text{ for } \omega \in A^c \end{cases},$$

where $A^c = \Omega \setminus A$.
At the beginning, we define the notion of a *σ-field.*

Definition 2.1. *Let X be a non-empty set. A class $\mathcal{X}$ of subsets of X is called a* σ-field *on X if it contains X itself and is closed under the formation of complements and countable unions, i.e.,*

1. $X \in \mathcal{X}$;
2. $A \in \mathcal{X}$ *implies* $A^c \in \mathcal{X}$;

3. $A_1, A_2, ... \in \mathcal{X}$ *implies* $A_1 \cup A_2 \cup ... \in \mathcal{X}$.

For a given collection $\mathcal{K}$ of subsets of X one can consider the *σ-field generated by* $\mathcal{K}$, which is the intersection of all the *σ-field*s on X containing $\mathcal{K}$.

Definition 2.2. *Let X and Y be two non-empty sets and let $\mathcal{X}$ and $\mathcal{Y}$ be σ-fields on X and Y, respectively. The product σ-field $\mathcal{X} \otimes \mathcal{Y}$ on $X \times Y$ is the σ-field generated by measurable rectangles $A \times B$, where $A \in \mathcal{X}$ and $B \in \mathcal{Y}$.*

We introduce the notions of measurable space and measurable mapping.

Definition 2.3. *A* measurable space *is a pair $(X, \mathcal{X})$ consisting of a non-empty set X and a σ-field $\mathcal{X}$ on X.*

Definition 2.4. *Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be two measurable spaces. A mapping $T : X \to Y$ is* measurable $\mathcal{X}/\mathcal{Y}$ *if $T^{-1}(G) \in \mathcal{X}$ for each $G \in \mathcal{Y}$.*

The subsequent important notions are measure and measure space.

Definition 2.5. *We call a set function $\mu : \mathcal{X} \to [0, \infty]$ a* measure *and a triple $(X, \mathcal{X}, \mu)$ a* measure space *if $(X, \mathcal{X})$ is a measurable space and*

(1) $\mu(\varnothing) = 0$;
(2) if $A_1, A_2, ... \in \mathcal{X}$ and $A_i \cap A_j = \varnothing$ for $i \neq j$, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Definition 2.6. *Let $(X, \mathcal{X}, \mu)$ be a measure space. The measure $\mu : \mathcal{X} \to [0, \infty]$ is* finite *if $\mu(X) < \infty$, infinite if $\mu(X) = \infty$ and σ-finite if there exists a sequence $\{A_n\}_{n=1}^{\infty} \subset \mathcal{X}$, such that $\mu(A_n) < \infty$ for each $n \in \mathbb{N}$ and $X = \bigcup_{n=1}^{\infty} A_n$.*

We say that a property that holds for $x \in X$ outside a set of measure 0 holds *almost everywhere* (*a.e.*, for short).

For measure spaces $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ one can define the *product measure space* denoted by $(X \times Y, \mathcal{X} \otimes \mathcal{Y}, \mu \otimes \nu)$, where $\mu \otimes \nu$ is the *product measure* on $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ (see (Billingsley 1995) for definition).

A *random experiment* is an experiment, whose outcome cannot be told in advance. A *probability space* is a mathematical model of such an experiment.

Definition 2.7. *A measure space $(\Omega, \mathcal{F}, \mathrm{P})$ is a* probability space *if* P *is a measure on $\mathcal{F}$ such that* $\mathrm{P}(\Omega) = 1$.

In the above definition Ω stands for all possible outcomes of the random experiment. We say that a set $A \subset \Omega$ occurs if the outcome of the random experiment happens to belong to A. The *σ-field* $\mathcal{F}$ on Ω consists of such subsets of Ω, called *events*, whose occurrence is decidable and noteworthy. Finally, for each $A \in \mathcal{F}$ the value $\mathrm{P}(A)$ models the chance of the occurrence of the event A. It is called the probability that the event A occurs.

Definition 2.8. *A subset N of Ω is called* negligible *if there exists $A \in \mathcal{F}$ such that $N \subset A$ and $\mathrm{P}(A) = 0$. A probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is called* complete *if $\mathcal{F}$ contains all negligible subsets of Ω.*

From the theoretical point of view, completeness of the probability space is often required. The next important notion to introduce is the notion of independence.

Definition 2.9. *σ-fields $\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_n \subset \mathcal{F}$ are independent if for any sequence of events $A_1 \in \mathcal{G}_1, A_2 \in \mathcal{G}_2, ..., A_n \in \mathcal{G}_n$*

$$\mathrm{P}(A_1 \cap A_2 \cap ... \cap A_n) = \mathrm{P}(A_1)\,\mathrm{P}(A_2) ... \mathrm{P}(A_n).$$

Definition 2.10. *Events A_1,A_2, ...,$A_n \in \mathcal{F}$ are* independent *if for each subsequence A_{i_1},A_{i_2}, ...,A_{i_k}*

$$\mathrm{P}(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}) = \mathrm{P}(A_{i_1})\,\mathrm{P}(A_{i_2}) ... \mathrm{P}(A_{i_k}).$$

Let us assume that $A \in \mathcal{F}$ and $\mathrm{P}(A) = 1$. Then we say that the event A occurs *almost surely* (*a.s.*, for short)

2.1.2 Random variables and their distributions

Let $n \in \mathbb{N}$ and $\mathcal{R}^n$ be the family of bounded rectangles of the form

$$(a_1, b_1] \times \cdots \times (a_n, b_n] : a_i, b_i \in \mathbb{R},\ a_i < b_i,\ i = 1, 2, ..., n.$$

The σ-field $\mathcal{B}(\mathbb{R}^n)$ of Borel subsets of $\mathbb{R}^n$ is the σ-field generated by $\mathcal{R}^n$ (i.e., the smallest σ-field of subsets of $\mathbb{R}^n$ containing $\mathcal{R}^n$).

In this book we will consider *random variables* taking values in the measurable space $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Definition 2.11. *A function $X : \Omega \to E$, measurable $\mathcal{F}/\mathcal{E}$, is called a* random variable *(or an E-*valued random variable*).*

Definition 2.12. *A (probability)* distribution *of a random variable $X : \Omega \to E$ is a probability measure described by formula*

$$\mu_X(B) = \mathrm{P}\left(X^{-1}(B)\right) = \mathrm{P}(X \in B),\ B \in \mathcal{E}.$$

For each random variable $X : \Omega \to \mathbb{R}$, apart from its probability distribution, one can define its *cumulative distribution function* F^X.

Definition 2.13. *Let $X : \Omega \to \mathbb{R}$ be a random variable with a probability distribution μ_X. The cumulative distribution function $F^X : \mathbb{R} \to [0, 1]$ of X is described by the formula*

$$F^X(t) = \mathrm{P}(X \leq t) = \mu_X((-\infty, t]),\ t \in \mathbb{R}.$$

An important property of the cumulative distribution function (CDF, for short) is that if two random variables have the same CDF, then they have the same distribution.

A set $S \in \mathcal{B}(\mathbb{R})$ is a *support* of the distribution μ_X if $\mu_X(S) = 1$.

Definition 2.14. *Let $X : \Omega \to \mathbb{R}$ be a random variable. If a discrete set $S = \{x_1, x_2, ...\}$ is a support of μ_X, then X and its distribution μ_X are called discrete. In this case, μ_X is determined by the sequence*

$$\mu_X(x_1), \mu_X(x_2), ...$$

Definition 2.15. *A random variable $X : \Omega \to \mathbb{R}$ and its distribution μ_X are called continuous if μ_X has a density f_X with respect to the Lebesgue measure (i.e., a non-negative function on $\mathbb{R}$, which integrates to 1 on $\mathbb{R}$). Then, for each $A \in \mathcal{B}(\mathbb{R})$ the following equality holds:*

$$\mathrm{P}(X \in A) = \mu_X(A) = \int_A f_X(x)\, dx.$$

The above introduced density function f_X is determined uniquely only to within a set of Lebesgue measure zero.

Definition 2.16. *The* expected value *of a random variable $X : \Omega \to \mathbb{R}$ is defined as the integral with respect to the probability measure* P

$$\mathbb{E}^{\mathrm{P}} X = \int_\Omega X d\,\mathrm{P},$$

if such an integral exists, i.e. $\mathbb{E}^{\mathrm{P}} |X| < \infty$.

For the definition of the integral with respect to a probability measure we refer the reader to (Billingsley 1995) or (Jakubowski & Sztencel 2000). If $\mathbb{E}^{\mathrm{P}} |X| < \infty$, then we say that the random variable X is integrable.

Definition 2.17. *We identify the integrable random variables that differ only on a set of probability* 0 *and denote their space by $L^1(\Omega, \mathcal{F}, \mathrm{P})$. More generally, for each $k \in \mathbb{N}$ we denote by $L^k(\Omega, \mathcal{F}, \mathrm{P})$ the spaces of random variables $X : \Omega \to \mathbb{R}$ identified as above, such that*

$$\mathbb{E}^{\mathrm{P}} |X|^k = \int_\Omega |X|^k d\,\mathrm{P} < \infty.$$

For $X \in L^k(\Omega, \mathcal{F}, \mathrm{P})$, the expected values $\mathbb{E}^{\mathrm{P}} |X|^k$ and $\mathbb{E}^{\mathrm{P}} X^k$ are called *absolute k-th moment* and *k-th moment* of X, respectively. Moreover, *variance* of each random variable $X \in L^2(\Omega, \mathcal{F}, \mathrm{P})$ is given by

$$\mathrm{Var}^{\mathrm{P}} X = \mathbb{E}^{\mathrm{P}} \left(X - \mathbb{E}^{\mathrm{P}} X \right)^2 = \mathbb{E}^{\mathrm{P}} X^2 - \left(\mathbb{E}^{\mathrm{P}} X \right)^2.$$

For a random variable $X \in L^1(\Omega, \mathcal{F}, \mathrm{P})$ and a σ-field $\mathcal{G} \subset \mathcal{F}$ one can define the *conditional expected value.*

Definition 2.18. *Let $X : \Omega \to \mathbb{R}$ be an integrable random variable on $(\Omega, \mathcal{F}, \mathrm{P})$ and let $\mathcal{G} \subset \mathcal{F}$ be a σ-field. The* conditional expected value *of the random variable X with respect to the σ-field $\mathcal{G}$ is a random variable $\mathbb{E}^{\mathrm{P}}(X|\mathcal{G})$, satisfying the following assumptions:*

1. $\mathbb{E}^{\mathrm{P}}(X|\mathcal{G})$ *is $\mathcal{G}$-measurable.*
2. $\forall A \in \mathcal{G} \quad \int_A X d\,\mathrm{P} = \int_A \mathbb{E}^{\mathrm{P}}(X|\mathcal{G})\, d\,\mathrm{P}$.

A very important notion is the notion of characteristic function.

Definition 2.19. *For a probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ the* characteristic function φ *is defined by the equality:*

$$\varphi(t) = \int_{\mathbb{R}} e^{itx} \mu(dx) = \int_{\mathbb{R}} \cos(tx)\, \mu(dx) + i \int_{\mathbb{R}} \sin(tx)\, \mu(dx)$$

for $t \in \mathbb{R}$.

Definition 2.20. *The characteristic function of a random variable $X : \Omega \to \mathbb{R}$ having distribution μ_X has the form:*

$$\varphi_X(t) = \mathbb{E}^{\mathrm{P}} e^{itX} = \int_{\mathbb{R}} e^{itx} \mu_X(dx)$$

for $t \in \mathbb{R}$.

Similarly as the cumulative distribution function F^X, the characteristic function φ_X uniquely determines the distribution of X.

For $n > 1$ the random variable $X = (X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^n$ is also called an *n-dimensional random variable* or a *random vector*. The (joint) distribution of the random vector $X : \Omega \to \mathbb{R}^n$ is uniquely determined by the (joint) cumulative distribution function $F^X : \mathbb{R}^n \to [0, 1]$, defined for any $t = (t_1, t_2, ..., t_n) \in \mathbb{R}^n$ by the equality

$$\begin{aligned} F^X(t) &= \mathrm{P}(X_1 \le t_1, X_2 \le t_2, ..., X_n \le t_n) \\ &= \mu_X((-\infty, t_1] \times (-\infty, t_2] \times ... \times (-\infty, t_n]). \end{aligned}$$

Similarly as in the one-dimensional case, the distribution μ_X of X may be discrete (i.e., have a countable support) or continuous with a non-negative density function $f_X : \mathbb{R}^n \to \mathbb{R}$ with respect to n-dimensional Lebesgue measure, i.e.

$$\mathrm{P}(X \in A) = \mu_X(A) = \int_A f_X(x)\, dx, \;\; A \in \mathcal{B}(\mathbb{R}^n).$$

Definition 2.21. *The expected value of a random vector $X = (X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^n$ is defined by the equality*

$$\mathbb{E}^{\mathrm{P}} X = \left(\mathbb{E}^{\mathrm{P}} X_1, \mathbb{E}^{\mathrm{P}} X_2, ..., \mathbb{E}^{\mathrm{P}} X_n\right),$$

under the condition that $\mathbb{E}^{\mathrm{P}} |X_i| < \infty$ for each $1 \le i \le n$.

For a finite or infinite collection $(X_i)_{i\in I}$ of random vectors the σ-field $\sigma(X_i,\ i\in I)$ generated by $(X_i)_{i\in I}$ is the smallest σ-field with respect to which each of the vectors is measurable.

Definition 2.22. *Random vectors $X_1, X_2, ..., X_k$ are independent if the σ-fields $\sigma(X_1), \sigma(X_2), ..., \sigma(X_k)$, which they generate, are independent. An infinite collection of random vectors is independent if each of its finite subcollections is.*

If random variables $X_1, X_2, ..., X_k$ are independent and integrable, then also $X_1X_2...X_k$ is integrable and

$$\mathbb{E}^{\mathrm{P}}(X_1X_2...X_k) = \mathbb{E}^{\mathrm{P}} X_1\, \mathbb{E}^{\mathrm{P}} X_2 ... \mathbb{E}^{\mathrm{P}} X_k.$$

Moreover, variances add for sums of independent random variables belonging to $L^2(\Omega, \mathcal{F}, \mathrm{P})$ (see, e.g., (Billingsley 1995)).

Further basic notions and facts from probability theory, including conditional distributions can be found in (Billingsley 1995) and (Jakubowski & Sztencel 2000). To shorten notation, in cases where the probability measure is known from the context, we will write $\mathbb{E}$ and Var in place of the symbols $\mathbb{E}^{\mathrm{P}}$ and $\mathrm{Var}^{\mathrm{P}}$, respectively.

2.1.3 Examples of one-dimensional distributions

We focus on several examples of known discrete and continuous one-dimensional distributions. Some other probability distributions will be defined in the subsequent chapters.

Bernoulli distribution

Bernoulli distribution with parameter p, $0 < p < 1$, has a simple discrete form. For a random variable X having this distribution

$$\mathrm{P}(X = 0) = 1 - p,\ \mathrm{P}(X = 1) = p.$$

Moreover,

$$\mathbb{E}X = p,\ \mathrm{Var}\, X = p(1-p).$$

Binomial distribution

Binomial distribution (denoted $\mathrm{Bin}(n,p)$) is a discrete distribution with two parameters: $n \in \mathbb{N}$ and $p \in (0,1)$. For a random variable X with this distribution (we use the notation $X \sim \mathrm{Bin}(n,p)$) μ_X has support S of the form

$$S = \{0, 1, 2, ..., n\}$$

and for each $i \in S$

$$\mathrm{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i}.$$

Furthermore, $\mathbb{E}X = np$ and $\mathrm{Var}\, X = np(1-p)$.

Poisson distribution

Poisson distribution with parameter $\lambda > 0$ is a distribution with the infinite support $S = \{0, 1, 2, ...\}$. For a random variable X having Poisson distribution with parameter λ (we also use notation $X \sim \text{Poiss}(\lambda)$), there is

$$\mathrm{P}(X = i) = e^{-\lambda}\frac{\lambda^i}{i!}.$$

Moreover, $\mathbb{E}\,X = \operatorname{Var} X = \lambda$.

Uniform distribution

Uniform distribution $U[a, b]$, $a < b$, is a continuous distribution with density function described by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b]; \\ 0 & \text{for } x \notin [a, b] \end{cases}.$$

For a random variable X with distribution $U[a, b]$ (further on, we use notation $X \sim U[a, b]$), $\mathbb{E}\,X = \frac{a+b}{2}$, $\operatorname{Var} X = \frac{(b-a)^2}{12}$.

Exponential distribution

Exponential distribution $\text{Exp}(\lambda)$, $\lambda > 0$, is a continuous distribution with density function of the form

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & for\ x \geq 0; \\ 0 & for\ x < 0. \end{cases}$$

If a random variable X has distribution $\text{Exp}(\lambda)$ (we will also write $X \sim \text{Exp}(\lambda)$ for short), then $\mathbb{E}\,X = \frac{1}{\lambda}$, $\operatorname{Var} X = \frac{1}{\lambda^2}$.

Normal distribution

Normal distribution $N(\mu, \sigma^2)$ is a continuous distribution with density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}.$$

If $X \sim N(\mu, \sigma^2)$ (i.e., X is normally distributed with parameters $\mu \in \mathbb{R}$, σ^2, where $\sigma > 0$), then $\mathbb{E}\,X = \mu$ and $\operatorname{Var} X = \sigma^2$. If $X \sim N(0, 1)$, then we say that X is a random variable from the standard normal distribution.

2.2 Basic notions of stochastic analysis in continuous time

Continuous time stochastic processes describe random phenomena that change as time continuously progresses. They are used, inter alia, in economics, finance, mathematical biology and physics.

In this chapter we assume that the time set under consideration, $\mathcal{T}$, has the form $\mathbb{R}_+ = [0, \infty)$ or $[0, T']$ for $T' < \infty$. At the beginning we define a *filtered probability space*.

Definition 2.23. *A filtered probability space (also called a* stochastic basis*) is a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ equipped with a* filtration $\mathbf{F} = (\mathcal{F}_t)_{t\in\mathcal{T}}$*, which is an increasing family of sub-σ-fields of $\mathcal{F}$. In other words, for any $s \leq t$, $\mathcal{F}_s \subset \mathcal{F}_t$. We denote the filtered probability space by $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$. We say that $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$ satisfies the* usual conditions *if, additionally,*

i) the σ-field $\mathcal{F}$ is P-complete,
ii) $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}$,
iii) the filtration $\mathbf{F}$ is right-continuous, i.e., $\mathcal{F}_{t+} = \mathcal{F}_t$ for each $t \in \mathcal{T}$, where

$$\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s.$$

We can now define a *stochastic process* and related notions.

Definition 2.24. *Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. A* stochastic process $X = (X_t)_{t\in\mathcal{T}}$ *on $(\Omega, \mathcal{F}, \mathrm{P})$ is a collection of ($\mathbb{R}$-valued) random variables on $(\Omega, \mathcal{F}, \mathrm{P})$. X is called* $\mathbf{F}$-adapted *(or* adapted to the filtration $\mathbf{F} = (\mathcal{F}_t)_{t\in\mathcal{T}}$ *or, for short,* adapted*) if for each $t \in \mathcal{T}$ X_t is $\mathcal{F}_t$-measurable. The functions $t \longmapsto X_t(\omega)$, where $\omega \in \Omega$, are called* sample paths *(or* trajectories*) of X. The stochastic process X is called* càdlàg [1] *if it has a.s. sample paths, which are right-continuous with left limits. It is called* càglàd[1]*, if its sample paths are left-continuous with right limits.*

Remark 2.25. *Any stochastic process X is adapted to its* natural filtration $\mathbf{F}^X = \left(\mathcal{F}_t^X\right)_{t\in\mathcal{T}}$*, where for each $t \in \mathcal{T}$ the σ-field $\mathcal{F}_t^X$ has the form*

$$\mathcal{F}_t^X = \sigma\left(X_s, s \leq t\right).$$

We will assume that the processes considered in this section are defined on $(\Omega, \mathcal{F}, \mathrm{P})$ and the stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$ satisfies the usual conditions.

For two stochastic processes the notions of indistinguishability and modification are defined.

Definition 2.26. *Two stochastic processes X and Y are* modifications *if for each $t \in \mathcal{T}$ $X_t = Y_t$ a.s. X and Y are* indistinguishable *if the sample paths $t \longmapsto X_t(\omega)$ and $t \longmapsto Y_t(\omega)$ are equal a.s.*

[1] From the French expressions, respectively, "continue à droite, limité à gauche" and "continue à gauche, limité à droite".

It is clear that if the processes X and Y are indistinguishable, then Y is a modification of X. The converse implication holds for right-continuous (in particular, for càdlàg) processes, see (Jacod & Shiryaev 2003)).

Theorem 2.27. *Let X and Y be stochastic processes with right-continuous sample paths. If Y is a modification of X, then they are indistinguishable.*

The notion of *stopping time* plays very important role in stochastic analysis.

Definition 2.28. *A random variable $\tau : \Omega \to \mathcal{T} \cup \{\infty\}$ is a stopping time with respect to filtration $\mathbf{F} = (\mathcal{F}_t)_{t\in\mathcal{T}}$ if $\{\omega \in \Omega : \tau(\omega) \le t\} \in \mathcal{F}_t$ for each $t \in \mathcal{T}$.*

Stochastic processes can be also treated as mappings on $\Omega \times \mathcal{T}$. Optional and predictable σ-fields on $\Omega \times \mathcal{T}$ and stochastic processes connected with them are used in many financial applications.

Definition 2.29. *A σ-field $\mathcal{O}$ on $\Omega \times \mathcal{T}$ is called* optional *if it is generated by all the càdlàg adapted processes (understood in the way we mentioned above). A σ-field $\mathcal{P}$ on $\Omega \times \mathcal{T}$ is called predictable, if it is generated by all the* càg *(i.e., left-continuous) adapted processes.*

Definition 2.30. *A stochastic process X is optional (predictable) if the process, regarded as a function of (t, ω), is $\mathcal{O}$-measurable (resp. $\mathcal{P}$-measurable).*

An important class of stochastic processes is the class of martingales.

Definition 2.31. *A stochastic process $(X_t)_{t\in\mathcal{T}}$, which is* integrable *(i.e., $X_t \in L^1(\Omega, \mathcal{F}, \mathrm{P})$ for $t \in \mathcal{T}$) and adapted to filtration $\mathbf{F} = (\mathcal{F}_t)_{t\in\mathcal{T}}$ is a* martingale *with respect to $\mathbf{F}$, if for any $0 \le s < t$, $t \in \mathcal{T}$, $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$.*

Stochastic processes that are widely used are *semimartingales*. They form a broad class of processes containing *local martingales* and *processes of bounded variation*. A detailed exposition concerning local martingales, processes of bounded variation, semimartingales and their theory can be found in the book of Jacod & Shiryaev (2003).

2.3 Lévy processes

The term "Lévy process" honours the French mathematician Paul Lévy, who played a very important role in the study of processes with stationary independent increments. *Lévy processes* are important examples of Markov processes and semimartingales. Wiener process (Brownian motion) and Poisson process are two commonly known processes belonging to this class.

Let us assume that a filtered probability space $\left(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathrm{P}\right)$ satisfies the usual conditions.

Definition 2.32. *A continuous time $\mathbf{F}$-adapted stochastic process X, defined on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$, is called a* Lévy process *(with respect to the filtration $\mathbf{F}$) if it satisfies the following assumptions:*

i) $X_0 = 0$ *a.s.*
ii) (increments independent of the past) For any $0 \leq s < t$, $t \in \mathcal{T}$, $X_t - X_s$ *is independent of* $\mathcal{F}_s$.
iii) (stationary increments) The distribution of $X_t - X_s$ *for any* $0 \leq s < t$, $t \in \mathcal{T}$, *depends only on the difference* $t - s$.
iv) (stochastic continuity) For any $t \in \mathcal{T}$ *and* $\varepsilon > 0$

$$\lim_{s \to t} \mathrm{P}\left(|X_s - X_t| > \varepsilon\right) = 0.$$

(v) (regularity of sample paths) The sample paths are almost surely càdlàg.

As it was noted by Protter (2005), for the definition of a Lévy process it is not necessary to involve filtration. In such a case we obtain the definition of an *intrinsic Lévy process* presented below, which is a Lévy process with respect to its minimal (completed) filtration.

Definition 2.33. *A continuous time stochastic process* X *defined on the probability space* $(\Omega, \mathcal{F}, \mathrm{P})$ *is called an* intrinsic Lévy process *if it fulfils the following assumptions:*

i) $X_0 = 0$ *a.s.*
ii) (independent increments) For any $t_0 < t_1 < ... < t_n$, $t_0, t_1, ..., t_n \in \mathcal{T}$

$$X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}}$$

are independent.
iii) (stationary increments) The distribution of $X_t - X_s$ *for any* $0 \leq s < t$ *depends only on the difference* $t - s$.
iv) (stochastic continuity) For any $t \in \mathcal{T}$ *and* $\varepsilon > 0$

$$\lim_{s \to t} \mathrm{P}\left(|X_s - X_t| > \varepsilon\right) = 0.$$

(v) (regularity of sample paths) The sample paths are almost surely càdlàg.

The following theorem was proved in (Protter 2005).

Theorem 2.34. *Each adapted stochastic process satisfying conditions i)-iv) has a unique modification, which is a càdlàg Lévy process.*

Moreover, Protter (2005) proved the following useful theorem concerning Lévy processes and filtrations.

Theorem 2.35. *Let* X *be a Lévy process with natural filtration* $\mathbf{F}^0 = \left(\mathcal{F}_t^0\right)_{t \in \mathcal{T}}$ *and let* $\mathcal{N}$ *be the set of the* P*-null sets of* $\mathcal{F}$. *Then the filtration* $\mathbf{G}^0 = (\mathcal{G}_t)_{t \in \mathcal{T}}$, *where* $\mathcal{G}_t = \sigma\left(\mathcal{F}_t^0 \cup \mathcal{N}\right)$, $t \in \mathcal{T}$, *is right-continuous.*

Jacod–Grigelionis characteristic triple (B, C, ν) for semimartingales is widely applied in stochastic analysis. The semimartingale characteristics (also called

local characteristics) have, in addition, a great importance for mathematical finance. For the general definition of the stochastic processes B, C and ν for real-valued semimartingales the interested reader is referred, e.g., to (Shiryaev 1999, Jacod & Shiryaev 2003), and to (Nowak 2002), where Jacod–Grigelionis characteristics for quasi-left continuous Hilbert space-valued semimartingales were introduced and their existence was proved.

For a Lévy process X, the local characteristics have a simple form. Let us assume that a truncation function h_d is defined by the formula $h_d(x) = xI_{|x|\leq d}$ for a constant $d > 0$. Let us denote by $\mathcal{M}(\mathbb{R})$ the space of non-negative measures on $\mathbb{R}$. For X, the characteristics (B, C, ν) are described by the following version of the Lévy–Khintchine formula

$$\varphi_t(\theta) = \mathbb{E}^{\mathrm{P}}\, e^{i\theta X_t} = \exp\left\{ i\theta B_t - \frac{1}{2}\theta^2 C_t + \int_{\mathbb{R}} \left(e^{i\theta x} - 1 - i\theta h_d(x)\right) \nu_t(dx) \right\},$$

where $\varphi_t(\theta)$ is the characteristic function of X_t,

$$B_t : [0,T] \to \mathbb{R},\ B_t = bt, \tag{2.1}$$

$$C_t : [0,T] \to \mathbb{R}, \quad C_t = ct, \tag{2.2}$$

$$\nu_t : [0,T] \to \mathcal{M}(\mathbb{R}), \quad \nu_t(dx) = \nu(dx)\, t, \tag{2.3}$$

$$\nu(\{0\}) = 0,\ \int_{\mathbb{R}} \left(|x|^2 \wedge 1\right) \nu(dx) < \infty\ ,$$

$b \in \mathbb{R}$, $c = \sigma^2 \geq 0$ and $\nu \in \mathcal{M}(\mathbb{R})$. The triple $\left(b, \sigma^2, \nu\right)$ is often called the characteristic triple of a Lévy process X (or *Lévy characteristics*) and it uniquely represents (B, C, ν). In particular, ν is called Lévy measure of a Lévy process X. Moreover, only the constant b depends on the value of the parameter d, describing the truncation function h_d.

Using measure ν one can distinguish between two types of Lévy processes, finitely active and infinitely active. For the processes of the first type $\nu(\mathbb{R}) < \infty$, while the equality $\nu(\mathbb{R}) = \infty$ characterizes the processes of the second type.

We focus our attention on two important examples of Lévy processes, i.e. the Wiener process and the Poisson process, together with certain notions related to them.

2.3.1 The Wiener process

Let $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$ be a stochastic basis. Let $T = T'$ for $\mathcal{T} = [0, T']$ and $T = \infty$ in the case of $\mathcal{T} = [0, \infty)$. We denote by $\mathcal{F}_\infty$ the σ-field $\mathcal{F}_\infty = \sigma\left(\bigcup_{t\in[0,\infty)} \mathcal{F}_t\right)$.

Definition 2.36. *A* standard Wiener process *(or a* Wiener process*, for short) on $(\Omega, \mathcal{F}, \mathrm{P})$ is a Lévy process (with respect to $\mathbf{F}$) $W = (W_t)_{t\in\mathcal{T}}$ such that*

i) W is almost surely continuous.

ii) For each $t \in \mathcal{T}$ W_t *is normally distributed with mean* 0 *and variance* t.

A Wiener process, also called a Brownian motion, is a continuous martingale. The sample paths of W are almost surely nowhere differentiable and they have infinite variation on any interval. The theory of Wiener processes is presented, e.g., in (Revuz & Yor 1999).

For a standard Wiener process the Itô integral and Itô processes are defined.

Definition 2.37. *A stochastic process* $X = (X_t)_{t\in\mathcal{T}}$ *on* $(\Omega, \mathcal{F}, \mathrm{P})$ *is called* measurable *if it is measurable with respect to the product* σ*-field* $\mathcal{B}(\mathcal{T}) \otimes \mathcal{F}_T$, *where* $\mathcal{B}(\mathcal{T})$ *is the Borel* σ*-field on* $\mathcal{T}$.

We denote by J_1 the class of measurable and $\mathbf{F}$-adapted processes X such that

$$\mathrm{P}\left(\int_0^T X_t^2 dt < \infty\right) = 1.$$

Stochastic processes from the class J_1 are *integrable with respect to* W and for each $X \in J_1$ the integral

$$\int_0^T X_t dW_t$$

is called *Itô integral*. The Itô integral $\left(\int_0^t X_s dW_s\right)_{t\in\mathcal{T}}$ for $X \in J_1$ is a local martingale with respect to $\mathbf{F}$ (which has a continuous modification) and is a linear mapping.

Let us assume that two $\mathbf{F}$-adapted and measurable processes $Y = (Y_t)_{t\in\mathcal{T}}$ and $Z = (Z_t)_{t\in\mathcal{T}}$ fulfil the following two conditions:

$$\mathrm{P}\left(\int_0^T |Y_t|\, dt < \infty\right) = 1$$

and

$$\mathrm{P}\left(\int_0^T Z_t^2 dt < \infty\right) = 1.$$

Then the following process is well defined:

$$X_t = X_0 + \int_0^t Y_s ds + \int_0^t Z_s dW_s,\ t \in \mathcal{T}. \tag{2.4}$$

One can write (2.4) in the differential form:

$$dX_t = Y_t dt + Z_t dW_t.$$

We call a process $X = (X_t)_{t\in\mathcal{T}}$ of the above form an Itô processes.

More details concerning Itô integrals, Itô processes and their properties can be found, e.g., in (Shiryaev 1999). A general approach to stochastic integration

with respect to semimartingales is presented in (Jacod & Shiryaev 2003). Similarly as in the case of Itô integrals, it is possible to describe analytically the space of processes, which are integrable with respect to a semimartingale (see (Kwapień & Woyczyński 1991, Nowak 2003)).

2.3.2 The Poisson process and the compound Poisson process

The homogeneous Poisson process is the next important example of a (finitely active) Lévy process.

Let $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$ be a stochastic basis.

Definition 2.38. *A* Poisson process $N = (N_t)_{t\in\mathcal{T}}$ *with a positive constant intensity* λ *on* $(\Omega, \mathcal{F}, \mathrm{P})$ *is a Lévy process (with respect to* $\mathbf{F}$*) such that for each* $t \in \mathcal{T}$ *the random variable* N_t *is Poisson distributed with mean* λt*.*

The Poisson process with a positive constant intensity λ is also called *homogeneous Poisson process* (HPP). If $\lambda = 1$, the process N is called *standard homogeneous Poisson process*.

We denote by $(T_n)_{n\in\mathbb{N}}$ the sequence of arrival times (i.e., jump times) of N, where

$$T_n = \inf\{t \in \mathcal{T} : N_t \geq n\} = \inf\{t \in \mathcal{T} : N_t = n\}.$$

Let $L_1, L_2, ...$ be a sequence of independent exponentially distributed random variables with a constant parameter $\lambda > 0$.
Let $(S_n)_{n\in\mathbb{N}}$ be a sequence, described for each positive integer n by the equality

$$S_n = L_1 + L_2 + ... + L_n. \tag{2.5}$$

Finally, let for $t \in \mathcal{T}$

$$N(t) = \#\{n \geq 1 : S_n \leq t\}. \tag{2.6}$$

We present a theorem proved in (Mikosch 2009).

Theorem 2.39. *The following statements hold:*

1. *The process* $(N(t))_{t\in\mathcal{T}}$*, given by formula* (2.6)*, is a homogeneous Poisson process with intensity* λ*.*
2. *Let* $N = (N_t)_{t\in\mathcal{T}}$ *be a homogeneous Poisson process with intensity* $\lambda > 0$ *and arrival times* $0 \leq S_1 \leq S_2 \leq ....$ *Then* N *has representation (2.6) and* $(S_n)_{n\in\mathbb{N}}$ *has representation (2.5) for an* iid *(exponentially distributed with parameter* λ*) sequence* $(L_n)_{n\in\mathbb{N}}$*.*

Definition 2.40. *Let* $N = (N_t)_{t\in\mathcal{T}}$ *be a Poisson process on* $(\Omega, \mathcal{F}, \mathrm{P})$ *with a constant intensity* $\lambda > 0$*. Let* $\{U_i\}_{i=1}^{\infty}$ *be a sequence of* iid *random variables independent of* N*, having a distribution* ρ*. A* compound Poisson process *on* $(\Omega, \mathcal{F}, \mathrm{P})$ *is a stochastic process* $\tilde{N} = \left(\tilde{N}_t\right)_{t\in\mathcal{T}}$*, described by the formula*

$$\tilde{N}_t = \sum_{i=0}^{N_t} U_i, \ t \in \mathcal{T},$$

where we use the convention $\sum_{i=0}^{1} U_i = 0$.

From the fact that N is a Poisson process, it easily follows that the compound process $\tilde{N}$ is a Lévy process.

Poisson processes have their generalizations, called *non-homogeneous Poisson processes* (NHPP), which are stochastic processes with independent increments.

Let $\Lambda : \mathcal{T} \to [0, \infty)$ be a right-continuous non-decreasing function with $\Lambda(0) = 0$.

Definition 2.41. *A càdlàg* $\mathbf{F}$*-adapted stochastic process* $N(\Lambda) = (N(\Lambda)_t)_{t \in \mathcal{T}}$ *with independent increments on* $(\Omega, \mathcal{F}, \mathrm{P})$ *is called a* non-homogeneous Poisson process *with the mean value function* Λ *if the following conditions are fulfilled:*

1. $N_0 = 0$ *a.s.,*
2. *for each* $0 \leq s < t$, $t \in \mathcal{T}$, *we have* $N_t - N_s \sim \mathrm{Poiss}(\Lambda(t) - \Lambda(s))$.

We say that $N(\Lambda)$ *has the* intensity function $\lambda : \mathcal{T} \to [0, \infty)$ *(also called the* rate function*) if* Λ *is absolutely continuous and has the representation*

$$\Lambda(s,t) = \Lambda(t) - \Lambda(s) = \int_s^t \lambda(u) du$$

for each $0 \leq s < t$, $t \in \mathcal{T}$. *The intensity function* λ *is measurable and non-negative.*

It is worth noting that in many applications also generalizations of the compound Poisson processes with jumps (arrival times) modelled by NHPP are considered.

2.4 Equivalent change of probability measure and the Girsanov theorem

The notion of an equivalent probability measure plays a very important role in mathematical finance.

Definition 2.42. *A probability measure* P *is equivalent to a probability measure* Q *on a measurable space* $(\Omega, \mathcal{F})$ *if for each* $A \in \mathcal{F}$ $\mathrm{P}(A) = 0$ *if and only if* $\mathrm{Q}(A) = 0$.

Similarly as in Subsection 2.3.1, we assume that $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$ is a stochastic basis, $T = T'$ for $\mathcal{T} = [0, T']$ and $T = \infty$ in the case of $\mathcal{T} = [0, \infty)$. Moreover, $\mathcal{F}_\infty = \sigma(\bigcup_{t \in [0,\infty)} \mathcal{F}_t)$.

Let $W = (W_t)_{t\in\mathcal{T}}$ be a Wiener process on $(\Omega, \mathcal{F}, \mathrm{P})$ with respect to $\mathbf{F}$.

We present theorems and facts concerning the equivalent change of probability measure, basing on (Jakubowski, Palczewski, Rutkowski & Stettner 2004).

Theorem 2.43. *Assume that two probability measures* P *and* Q *are equivalent on* $(\Omega, \mathcal{F}_T)$*. Then there exists the Radon-Nikodym derivative* $g = \frac{d\,\mathrm{Q}}{d\,\mathrm{P}}$*, i.e.* $\mathrm{Q}(A) = \int_A g\, d\,\mathrm{P}$ *for* $A \in \mathcal{F}_T$ *and* g *is an* $\mathcal{F}_T$*-measurable positive random variable.*

Let us denote by $\mathbf{F}^W = \left(\mathcal{F}_t^W\right)_{t\in\mathcal{T}}$ the natural filtration of the Wiener process $W = (W_t)_{t\in\mathcal{T}}$.

Theorem 2.44. *If* P *is equivalent to* Q *on* $\left(\Omega, \mathcal{F}_T^W\right)$*, then there exists a stochastic process* $\gamma \in J_1$*, such that* $\frac{d\,\mathrm{Q}}{d\,\mathrm{P}} = \mathcal{E}_T$*, where*

$$\mathcal{E}_t = \exp\left(\int_0^t \gamma_s dW_s - \frac{1}{2}\int_0^t \gamma_s^2 ds\right)$$

for each $t \in \mathcal{T}$*.*

In turn, let us assume that

$$\frac{d\,\mathrm{Q}}{d\,\mathrm{P}} = \mathcal{E}_T$$

for a stochastic process $\gamma \in J_1$. Then, Q is equivalent to P if and only if $\mathbb{E}^{\mathrm{P}}\,\mathcal{E}_T = 1$ (i.e., $\mathcal{E}$ is a martingale).

A sufficient condition for the equality $\mathbb{E}^{\mathrm{P}}\,\mathcal{E}_T = 1$ is provided by the Novikov criterion, given below.

Corollary 2.45. *If process* γ *satisfies the inequality*

$$\mathbb{E}^{\mathrm{P}} \exp\left(\frac{1}{2}\int_0^T \gamma_s^2 ds\right) < \infty,$$

then $\mathbb{E}^{\mathrm{P}}\,\mathcal{E}_T = 1$*.*

The following version of the Girsanov theorem (see (Jacod & Shiryaev 2003) for its generalization and proof) describes the behaviour of the Wiener process W after an equivalent change of probability measure.

Theorem 2.46. *Let* Q *be a probability measure equivalent to* P *on* $(\Omega, \mathcal{F}_T)$*, such that*

$$\frac{d\,\mathrm{Q}}{d\,\mathrm{P}} = \mathcal{E}_T$$

for a stochastic process $\gamma \in J_1$*. If* $W = (W_t)_{t\in\mathcal{T}}$ *is a Wiener process with respect to* P *and filtration* $\mathbf{F}$*, then*

$$\tilde{W}_t = W_t - \int_0^t \gamma_s ds,\ t \in \mathcal{T},$$

is a Wiener process with respect to Q *and filtration* $\mathbf{F}$*.*

3

Foundations of theory of Markov chains

Maciej Romaniuk

In this chapter we consider the most important properties and theorems, concerning theory of Markov chains, which are necessary for introducing the MCMC methods (see Chapter 6). Additional details concerning these stochastic processes can be found in, e.g. (Brémaud 1999, Robert & Casella 2004, Doyle 2006, Jakubowski & Sztencel 2000).

3.1 Countable state space

A stochastic process $(X_i)_{i\in T}$ is called a *discrete-time stochastic process* if T is some subset of $\mathbb{N}$. In this chapter we assume that $T = \{n \in \mathbb{N} : n \geq 0\}$.

Definition 3.1. *Let $(X_i)_{i=0}$ be a discrete-time stochastic process with countable state space $\mathcal{S}$. If for all $k > 0$ and all $\mathcal{A} \subset \mathcal{S}$*

$$\mathrm{P}\left(X_k \in \mathcal{A} | X_{k-1}, \ldots, X_0\right) = \mathrm{P}\left(X_k \in \mathcal{A} | X_{k-1}\right) \ , \tag{3.1}$$

whenever both sides are well-defined, this process is called a Markov chain *(abbreviated further as MC).*

Condition (3.1) is known as the *Markov property*. It means that the probabilistic dependence of the new state X_k on the past behaviour of the whole random sequence takes place only through the value of the preceding state X_{k-1} and the complete information about previous history is therefore unnecessary.

The value x_0 is referred to as the *starting value* (or the *initial state*) of the chain. If this value is not set deterministically, but $X_0 \sim \pi_{X_0}$ for some distribution π_{X_0}, then such distribution is called the *initial distribution*.

Let s be the power of the state space $\mathcal{S}$. Because in Definition 3.1 it is assumed that $\mathcal{S}$ is a countable set, then the relevant Markov chain for such a space is called *countable-state MC*.

Example 3.2. *The simplest case of a countable-state MC is given by a sequence of* iid *random variables defined on the same countable state space $\mathcal{S}$. A more sophisticated example is an MC with two states, which are denoted by $\{1, 2\}$. The transition probabilities between these states are given by*

$$\mathrm{P}(X_k = 1|X_{k-1} = 1) = p_{11} \ , \ \mathrm{P}(X_k = 2|X_{k-1} = 1) = p_{12} = 1 - p_{11} \ , \quad (3.2)$$

$$\mathrm{P}(X_k = 1|X_{k-1} = 2) = p_{21} \ , \ \mathrm{P}(X_k = 2|X_{k-1} = 2) = p_{22} = 1 - p_{21} \ , \quad (3.3)$$

where $p_{11}, p_{21} \in [0, 1]$.

With this example in mind, an important definition for the theory of Markov chains may be introduced:

Definition 3.3. *We say that a Markov chain is a* homogeneous *Markov chain (HMC) if the right-hand side of the Markov property* (3.1) *does not depend on* k. *In this case, there exists a matrix*

$$\mathbb{P}_X = \left(\mathrm{P}\left(X_{k+1} = j|X_k = i\right)\right)_{i,j=1}^{s} = (p_{ij})_{i,j=1}^{s} \ ,$$

which is independent of k. *This matrix is a* stochastic matrix, *i.e. it fulfils the conditions*

$$p_{ij} \geq 0 \ , \sum_{l=1}^{s} p_{il} = 1$$

for all states i, j. *Such matrix* $\mathbb{P}_X$ *is called a* transition matrix.

In the case of HMC, the transition probabilities among the states are independent of the step k, meaning that they are independent of time. Because the transition probabilities remain constant in time, they are explicitly described by the entries of the transition matrix $\mathbb{P}_X$. From now on, all the Markov chains considered in this chapter will be homogeneous chains.

Example 3.4. *For the MC described by the transition probabilities* (3.2) *and* (3.3)*, the transition matrix is given by*

$$\mathbb{P}_X = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \ .$$

The following theorem has an important significance for the evaluation of random behaviour of Markov chains:

Theorem 3.5. *The distribution of MC at the step* $k > 0$ *is determined by the initial distribution* π_{X_0} *of this chain and its transition matrix* $\mathbb{P}_X$.

Proof of this theorem can be found in (Jakubowski & Sztencel 2000).

This theorem leads to the formula for the random distribution of the MC at the step k as a vector π_{X_k}, where

$$\pi_{X_k}(i) = \mathrm{P}(X_k = i) \ . \quad (3.4)$$

From Theorem 3.5 and the Bayes' rule, using this probability vector π_{X_k}, we get the equality

$$\pi_{X_k}^T = \pi_{X_0}^T \mathbb{P}_X^k \ , \quad (3.5)$$

where $\mathbb{P}_X^k$ is the k-step transition matrix with the i, j-th term given by

$$p_{ij}(k) = \mathrm{P}\left(X_{k+m} = j|X_m = i\right) \ .$$

Definition 3.6. *A probability distribution π_X satisfying*

$$\pi_X^T = \pi_X^T \mathbb{P}_X \ , \tag{3.6}$$

is called a stationary distribution *of the transition matrix $\mathbb{P}_X$, or (equivalently) of the corresponding MC.*

If π_X is the stationary distribution, then $\pi_X^T = \pi_X^T \mathbb{P}_X^m$ for all $m \geq 1$. Therefore, if the initial distribution is the same as the stationary distribution i.e. $\pi_{X_0} = \pi_X$, then $\pi_{X_m} = \pi_X$ for all steps $m \geq 1$. It means that if the chain is started from the stationary distribution, the same distribution describes the entire behaviour of such a MC.

The following definitions concern the so called communication features of the MC.

Definition 3.7. *State $b \in \mathcal{S}$ is said to be* accessible *from state $a \in \mathcal{S}$ if there exists $n \in \mathbb{N}$ such that $\mathrm{P}(X_n = b | X_0 = a) > 0$. States a and b are said to* communicate *if a is accessible from b and b is accessible from a. Such feature is denoted by $a \leftrightarrow b$.*

Clearly, the communication relation $\leftrightarrow$ is an equivalence relation, because it fulfils reflexivity, symmetry and transivity conditions. Therefore, it generates a partition of the state space $\mathcal{S}$ into disjoint equivalence classes, called *communication classes*.

Definition 3.8. *If for the given MC there exists only one communication class, then this chain and its transition matrix are said to be* irreducible.

Example 3.9. *Let us consider the chain with states $\{1, 2, 3\}$, determined by the transition matrix*

$$\mathbb{P}_X = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.4 & 0.3 & 0.3 \\ 0 & 0 & 1 \end{pmatrix} \ . \tag{3.7}$$

In this case, the states $\{1, 2\}$ communicate. The state $\{3\}$ is accessible from the states $\{1, 2\}$, but $\{1, 2\}$ are not accessible from $\{3\}$, therefore, this chain is not irreducible.

There is yet another classification of the states of MC, which is based on their periodicity feature, that should be considered.

Definition 3.10. *The* period *$o(i)$ of the state $i \in \mathcal{S}$ is given by*

$$o(i) = gcd \left\{ n \geq 1 : p_{ii}(n) > 0 \right\} \ , \tag{3.8}$$

with the convention $o(i) = +\infty$ if there is no $n \geq 1$ with $p_{ii}(n) > 0$. If $o(i) = 1$, then the state i is called aperiodic.

Example 3.11. *Let us consider a "jumping" two-state MC given by the transition matrix*

$$\mathbb{P}_X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ .$$

As it can be easily seen, the period of each of the states of this chain is equal to 2. For example

$$p_{11}(1) = 0, p_{11}(2) = 1, p_{11}(3) = 0, p_{11}(4) = 1, \dots \ ,$$

which leads to $o(1) = 2$.

Theorem 3.12. *If states a and b communicate, they have the same period. If MC is irreducible, then all the states of this chain have the same period or they are aperiodic.*

Proof of this theorem may be found in, e.g. (Brémaud 1999).

Therefore, we can speak of the period (or aperiodicity) of a communication class, or of the whole chain if this chain is irreducible.

Another classification of states is based on the probability of returns to a specified state.

Definition 3.13. *Let $a \in \mathcal{X}$. The state a is called* recurrent *if*

$$\mathrm{P}\left(\bigcup_{k=1}^{\infty} \{X_k = a\} \middle| X_0 = a \right) = 1 \ , \tag{3.9}$$

and transient *if*

$$\mathrm{P}\left(\bigcup_{k=1}^{\infty} \{X_k = a\} \middle| X_0 = a \right) < 1 \ . \tag{3.10}$$

It should be noted that every state in a Markov chain must be either transient or recurrent. If the state a is recurrent, it means that starting from a, the chain will almost surely visit this state an infinite number of times. These features can be also stated using an equivalent notation. Let

$$\eta_a = \sum_{k=0}^{\infty} \mathbb{1}(X_k = a) \tag{3.11}$$

be a function, which counts the number of visits in the state a. Then we have the following theorem:

Theorem 3.14. *The state $a \in \mathcal{S}$ is recurrent if and only if*

$$\mathrm{P}\left(\eta_a = \infty | X_0 = a\right) = 1 \Leftrightarrow \mathbb{E}\left(\eta_a | X_0 = a\right) = \infty \ . \tag{3.12}$$

The state $a \in \mathcal{X}$ is transient if and only if

$$\mathrm{P}\left(\eta_a < \infty | X_0 = a\right) = 1 \Leftrightarrow \mathrm{P}\left(\eta_a = \infty | X_0 = a\right) = 0 \Leftrightarrow \mathbb{E}\left(\eta_a | X_0 = a\right) < \infty \ .$$

Proof of this theorem may be found in, e.g. (Jakubowski & Sztencel 2000).

Theorem 3.15. *If MC is irreducible, then all of the states have the same type – if one of them is recurrent, all of the states are also recurrent, if one of them is transient, the rest of them is also transient.*

Proof of this theorem may be found in, e.g. (Jakubowski & Sztencel 2000).

Taking into account Theorems 3.12 and 3.15, the terms "aperiodic" and "recurrent" may be used as a description of the whole Markov chain if such chain is irreducible.

3.2 Uncountable state space

The definitions and theorems considered in Section 3.1 for the case of the countable state space $\mathcal{S}$ will be discussed further for uncountable (continuous) state space setting. Some important differences between these two instances will be also highlighted.

Definition 3.16. *Let* $(X_i)_{i=0} = (X_0 = x_0, X_1, \ldots, X_n, \ldots)$ *be a discrete-time stochastic process, where the state space* $\mathcal{S}$ *is some subset of* $\mathbb{R}^p$ *for some given* $p \in \mathbb{N}$. *If for all* $k > 0$ *and all* $\mathcal{A} \in \mathbb{B}(\mathcal{S})$

$$\mathrm{P}\left(X_k \in \mathcal{A} | X_{k-1}, \ldots, X_0\right) = \mathrm{P}\left(X_k \in \mathcal{A} | X_{k-1}\right) \ ,$$

whenever both sides are well-defined, this process is called a Markov chain *(MC).*

Because in Definition 3.16 of the Markov chain the state space $\mathcal{S}$ is some uncountable set, then for simplicity we say that such Markov chain is an *uncountable-state MC*.

Instead of the transition matrix, in the case of the continuous state space $\mathcal{S}$ a *transition kernel* is considered.

Definition 3.17. *A* transition kernel *is a function* $\mathcal{K}_X : \mathcal{S} \times \mathbb{B}(\mathcal{S}) \times \mathbb{N} \to \mathbb{R}$ *such that*

1. *for all* $x \in \mathcal{S}$, $\mathcal{K}_X(x, ., .)$ *is probability measure,*
2. *for all* $\mathcal{A} \in \mathbb{B}(\mathcal{S})$, $\mathcal{K}_X(., \mathcal{A}, .)$ *is measurable.*

The transition kernel $\mathcal{K}_X(x, y, k)$ may be identified with the set of densities describing distributions for the transitions between the state x and the state y at the step k. Then, we can define the relation between the transition kernel $\mathcal{K}_X(., ., .)$ and the transition probability for the considered MC as

$$\mathrm{P}(X_{k+1} \in \mathcal{B} | X_k = x) = \int_{\mathcal{B}} \mathcal{K}_X(x, y, k)\, dy \tag{3.13}$$

for all $\mathcal{B} \in \mathbb{B}(\mathcal{S})$.

Definition 3.18. *The Markov Chain* $(X_i)_{i=0}$ *is called* homogeneous *(or* time homogeneous*) if the transition kernel does not depend on the step* k. *In this case the transition kernel* $\mathcal{K}_X(x, y, k)$ *is simply denoted by* $\mathcal{K}_X(x, y)$, *where* $\mathcal{K}_X : \mathcal{S} \times \mathbb{B}(\mathcal{S}) \to \mathbb{R}$.

From now on only homogeneous MCs (HMCs) will be considered.

Example 3.19 (Wiener process with discrete time)**.** *A Wiener process with discrete time* $(W_i)_{i=0} = (W_0 = 0, W_1, \ldots, W_n, \ldots)$ *is an example of a MC with uncountable state space. In such a case, for all* $x, y \in \mathbb{R}$ *the transition kernel is given by*

$$\mathcal{K}_X(x, y) = f_{N(x,1)}(y) \ , \tag{3.14}$$

where $f_{N(x,1)}(y)$ *is the density of* $N(x, 1)$*. Then from* (3.13) *we obtain the transition probability*

$$\mathrm{P}(X_{k+1} \in \mathcal{B} | X_k = x) = \int_{\mathcal{B}} f_{N(x,1)}(y) dy \ .$$

In the case of the discrete state space, MC is irreducible if all of its states communicate. In the uncountable state space setting, it is necessary to introduce an auxiliary measure ρ in order to correctly define the notion of irreducibility.

Definition 3.20. *A Markov chain* $(X_i)_{i=0}$ *with the transition kernel* $\mathcal{K}_X(x, y)$ *is* ρ-irreducible *if there exists a measure* ρ *such that for every* $\mathcal{A} \in \mathbb{B}(\mathcal{S})$ *with* $\rho(\mathcal{A}) > 0$ *there exists an* n *such that*

$$\mathcal{K}_X^n(x_0, \mathcal{A}) = \mathrm{P}(X_n \in \mathcal{A} | x_0) > 0 \tag{3.15}$$

for all $x_0 \in \mathcal{S}$*. The chain is* strongly ρ-irreducible *if* $n = 1$ *for all measurable sets* $\mathcal{A}$*.*

It should be noted that a specific form of the measure ρ in Definition 3.20 of ρ-irreducibility plays no crucial role – the irreducibility is an intrinsic property of the considered MC and does not rely on the choice of this measure (see (Robert & Casella 2004) for additional details).

Example 3.21. *The chain considered in Example 3.19 is irreducible.*

Definition 3.22. *A* σ*-finite measure* π_X *is* invariant *for the transition kernel* $\mathcal{K}_X(.,.)$ *and the associated MC, if*

$$\pi_X(\mathcal{A}) = \int_{\mathcal{A}} \int_{\mathcal{S}} \mathcal{K}_X(x, y) \, d\pi_X(x) \, dy \tag{3.16}$$

for all $\mathcal{A} \in \mathbb{B}(\mathcal{S})$*. If this measure* π_X *is a probability measure, then such distribution is called* stationary distribution*. When there exists a stationary measure for a* ρ*-irreducible MC, such chain is* positive*.*

In order to properly define the periodicity in the case of the continuous state space $\mathcal{S}$, it is necessary to introduce an auxiliary definition of a *small set*.

Definition 3.23. *A set* $\mathcal{C}$ *is small if there exists* $m \in \mathbb{N}$ *and a measure* ν_m *such that*

$$\mathrm{P}(X_m \in \mathcal{A} | X_0 = x) \geq \nu_m(\mathcal{A}) > 0 \tag{3.17}$$

for all $x \in \mathcal{C}$ *and* $\mathcal{A} \in \mathbb{B}(\mathcal{S})$*.*

The condition (3.17) in the above definition means that for the small set $\mathcal{C}$ there exists some fixed number of steps m and the measure ν_m, for which the probability of transition between any state $x \in \mathcal{C}$ and the event $\mathcal{A}$ has some lower limit given by $\nu_m > 0$. It should be noted that this value m is independent of the selected set $\mathcal{A}$. The small sets are a common feature of Markov chains. For example, if MC is irreducible, then the state space $\mathcal{S}$ may be decomposed into a denumerable partition of small sets, as it is discussed in (Robert & Casella 2004).

Definition 3.24. *A ρ-irreducible MC has a* cycle *of length d if there exists a small set $\mathcal{C}$, $M \in \mathbb{N}$ and a probability distribution ν_M such that d is gcd of*

$$\{m \geq 1 : \exists\ \delta_m > 0 \text{ such that } \mathcal{C} \text{ is small for } \nu_m \geq \delta_m \nu_M\}\ . \tag{3.18}$$

It should be noted that the value d in the above definition is independent of the small set $\mathcal{C}$ and it intrinsically characterizes the considered Markov chain. Then, this leads to the following definition:

Definition 3.25. *The* period *of MC is the largest integer d satisfying Definition 3.24. If $d = 1$ then MC is* aperiodic.

Let $\mathcal{A} \in \mathbb{B}(\mathcal{S})$. Then, in an analogy to the function described by the formula (3.11), let $\eta_{\mathcal{A}}$ be

$$\eta_{\mathcal{A}} = \sum_{k=0}^{\infty} \mathbb{1}(X_k \in \mathcal{A})\ ,$$

i.e. $\eta_{\mathcal{A}}$ is the function counting the number of visits in the set $\mathcal{A}$. The notions of transience and recurrence are more complicated in the case of the uncountable state space than in the discrete setting, discussed in Section 3.1.

Definition 3.26. *A set $\mathcal{A}$ is called* recurrent *if*

$$\mathbb{E}_x \eta_{\mathcal{A}} = \mathbb{E}\left(\eta_{\mathcal{A}} | X_0 = x\right) = \infty \tag{3.19}$$

for every $x \in \mathcal{A}$. The set $\mathcal{A}$ is uniformly transient *if there exists a constant M such that*

$$\mathbb{E}_x \eta_{\mathcal{A}} < M$$

for every $x \in \mathcal{A}$. This set is transient *if there exists a countable collection of uniformly transient sets $\mathcal{B}_i$ such that*

$$\mathcal{A} = \bigcup_i \mathcal{B}_i\ . \tag{3.20}$$

A Markov chain is recurrent *if this chain is ρ-irreducible and for every $\mathcal{A} \in \mathbb{B}(\mathcal{S})$ such that $\rho(\mathcal{A}) > 0$, this set $\mathcal{A}$ is recurrent. The chain is* transient *if it is ρ-irreducible and if $\mathcal{S}$ is transient.*

Theorem 3.27. *A ρ-irreducible MC is either recurrent or transient.*

This theorem is proved in (Robert & Casella 2004).

Theorem 3.28. *A ρ-irreducible MC is recurrent if there exists a small set $\mathcal{C}$ such that $\rho(\mathcal{C}) > 0$ and*

$$\mathrm{P}_x\left(\inf_k\{X_k \in \mathcal{C}\} < \infty\right) = \mathrm{P}\left(\inf_k\{X_k \in \mathcal{C}\} < \infty \middle| X_0 = x\right) = 1 \ , \tag{3.21}$$

for every $x \in \mathcal{C}$.

This theorem is proved in (Robert & Casella 2004).

A more important property in the application of MCMC methods for the case of the continuous state space is known as *Harris recurrence*:

Definition 3.29. *A set $\mathcal{A}$ is* Harris recurrent *if*

$$\mathrm{P}(\eta_{\mathcal{A}} = \infty | X_0 = x) = 1 \tag{3.22}$$

for all $x \in \mathcal{A}$. The chain is Harris recurrent *if there exists a measure ρ such that MC is ρ-irreducible and for every set $\mathcal{A}$, such that $\rho(\mathcal{A}) > 0$, the set $\mathcal{A}$ is Harris recurrent.*

Theorem 3.30. *If for every $\mathcal{A} \in \mathbb{B}(\mathcal{X})$*

$$\mathrm{P}_x\left(\inf_k\{X_k \in \mathcal{A}\} < \infty\right) = 1 \tag{3.23}$$

for every $x \in \mathcal{A}$, then $\mathrm{P}_x(\eta_{\mathcal{A}} = \infty) = 1$ for all $x \in \mathcal{S}$ and this chain is Harris recurrent.

This theorem is proved in (Robert & Casella 2004).

There is an important difference between recurrence and Harris recurrence of MC for the uncountable state space. In the case of recurrence, an infinite average number of visits only for a small set is required (see Theorem 3.28). But in the case of Harris recurrence, this condition is related to infinite number of visits for every path of MC, as it can be seen from Definition 3.29 and Theorem 3.30.

3.3 Ergodicity theorems

When applying Markov Chain Monte Carlo methods we are interested in conditions which guarantee that the empirical average

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \ ,$$

where the generated sequence $(X_i)_{i=0}$ forms a Markov chain, converges to the relevant expected value in a similar way as in the *iid* case, i.e. for the Monte Carlo methods described in Section 5.2.3.

The following theorems are related to this issue:

Theorem 3.31 (Countable state space)**.** *Let $(X_i)_{i=0}$ be an irreducible, recurrent countable-state MC with the stationary distribution π_X, and let $h : \mathcal{S} \to \mathbb{R}$ be such that*

$$\mathbb{E}_{\pi_X}|h(X)| = \sum_{x\in\mathcal{S}} |h(x)|\pi_X(x) < \infty \ .$$

Then for any initial distribution π_{X_0}

$$\frac{1}{n}\sum_{k=1}^{n} h(X_k) \xrightarrow[n\to\infty]{a.s.} \sum_{x\in\mathcal{S}} h(x)\pi_X(x) \ .$$

Proof of this theorem can be found in (Jakubowski & Sztencel 2000).

Theorem 3.32 (Uncountable state space)**.** *Let $(X_i)_{i=0}$ be a positive, ρ-irreducible, Harris recurrent uncountable-state MC with the stationary distribution π_X, and let $h : \mathcal{S} \to \mathbb{R}$ be such that*

$$\mathbb{E}_{\pi_X}|h(X)| = \int_{\mathcal{S}} |h(x)|d\pi_X(x) < \infty \ .$$

Then for any initial distribution π_{X_0}

$$\frac{1}{n}\sum_{k=1}^{n} h(X_k) \xrightarrow[n\to\infty]{a.s.} \int_{\mathcal{S}} h(x)\ d\pi_X(x) \ .$$

If an MC fulfils the assumptions required in the above theorems, then such chain is called *ergodic MC*. Therefore the outcomes mentioned are known as *ergodicity theorems*. A more detailed discussion and the proofs can be found in (Robert & Casella 2004).

Apart from the convergence of the mean to the expected value, which is guaranteed by the ergodicity theorems, the distance between the stationary distribution and the distribution of MC after n steps is measured. In order to do this, the *total variation norm* in the form of

$$\|\mu_1 - \mu_2\|_{\mathrm{TV}} = \sup_{\mathcal{A}} |\mu_1(\mathcal{A}) - \mu_2(\mathcal{A})|$$

is sometimes applied, where supremum is taken for the considered probability measures μ_1 and μ_2 and for all sets $\mathcal{A} \in \mathbb{B}(\mathcal{S})$. Such measure is especially important in the case of convergence analysis of MCMC output, which will be discussed in Section 6.7 in a more detailed way.

4

Generation of random variables and stochastic processes

Maciej Romaniuk

In this chapter we discuss methods, which are necessary for transforming the output from (pseudo)random number generator (i.e. the sequence, which may be treated as the *iid* random variables from the uniform distribution) into the random variables from various, more complex probability distributions, like normal distribution.

The relevant methods may be dived into two groups: the general methods and the more specific, particular algorithms. The set of the general methods is related to algorithms and theorems, which may be used for various probability distributions. Usually only some general assumptions should be satisfied in order be able to apply these algorithms. The second group of generation methods consists of theorems and lemmas, which are applicable only for some, certain distributions, e.g. beta distribution. They are related, as a rule, to very specific transformations of random variables based on probability theory.

In this chapter we also consider some algorithms which are useful in obtaining the output (i.e. the trajectories) of the widely used stochastic processes, like the Poisson process.

4.1 General methods of (pseudo)random number generation

As it was noted in Section 1.1.3, generation of (pseudo)random numbers usually consists of two phases:

1. During the first stage, the PRNG is used to generate *iid* output from the $U[0,1]$ distribution.
2. Then, the obtained sequence is transformed into another one which constitutes a sample from some other probability distribution (e.g. normal distribution).

In the case of the first step, various examples of the relevant algorithms were discussed in Section 1.1.3. Therefore, now we assume that the function (i.e. the algorithm) `GenerateU`, which generates *iid* random variables from $U[0,1]$ distribution can be directly used in our considerations. Keeping this in mind,

we focus only on the second phase. In this section, we describe some general methods, which are used to obtain random variables from various probability distributions. For additional details about the relevant algorithms, see, e.g., (Law 2007, Ripley 1987, Robert & Casella 2004, Wieczorkowski & Zieliński 1997).

4.1.1 The inversion method

One of the most important algorithms, the *inversion method* (also known as the *inverse-transform method*) is based on the following theorem:

Theorem 4.1 (Inversion method)**.** *Let U be a random variable distributed according to $U[0,1]$ distribution and $F_X(.)$ be a continuous and increasing cdf (cumulative distribution function) of some probability distribution. Then, the variable X, given by the formula*

$$X = F_X^{-1}(U) \tag{4.1}$$

is a sample from $F_X(.)$ (i.e. $X \sim F_X(.)$).

Proof. As assumed, $F_X(.)$ is a continuous and increasing cdf. Therefore, the inverse function $F_X^{-1}(.)$ exists and from (4.1) we have

$$\mathrm{P}(X \leq x) = \mathrm{P}\left(F_X^{-1}(U) \leq x\right) = \mathrm{P}\left(U \leq F_X(x)\right) = F_X(x)$$

which concludes the proof. □

If we apply Theorem 4.1, then we obtain the following algorithm:

Algorithm 4.2 (Inversion method)**.**

```
U = GenerateU;
X = F^(-1) (U);
return X.
```

In the above pseudocode `F^(-1) (.)` denotes $F_X^{-1}(.)$.

It can be easily seen that the inversion method is a rather universal approach, because if $F_X^{-1}(.)$ exists, the algorithm can be straightforwardly applied. However, this method may be too slow or numerically unstable for certain distributions. We now present two important examples of applications of Algorithm 4.2, which are widely used in practice and frequently mentioned in literature.

Example 4.3. *Let us suppose that $X \sim U[a,b]$. Then, the cdf of X is given by*

$$F(x) = \frac{x-a}{b-a} \ .$$

This function is continuous and increasing, therefore the inverse function exists and is given by

$$F^{-1}(y) = y(b-a) + a \ .$$

Then, by applying Theorem 4.1 we conclude that

$$X = U(b-a) + a$$

is a sample from $U[a,b]$. The corresponding algorithm is given by the source code

Algorithm 4.4 (Inversion method for the uniform distribution)**.**

```
U = GenerateU;
X = U * (b-a) + a;
return X.
```

Example 4.5. *Let us suppose that $X \sim \mathrm{Exp}(\lambda)$. Then the inverse function of cdf of the exponential distribution is given by*

$$F^{-1}(y) = -\frac{1}{\lambda}\ln(1-y) \ , \tag{4.2}$$

therefore the variable

$$X = -\frac{1}{\lambda}\ln U \tag{4.3}$$

is a sample from $\mathrm{Exp}(\lambda)$.

Theorem 4.1 may be generalized to the case of a non-continuous and non-decreasing cdf. To obtain this result, we introduce the *generalized definition of the inverse function*, given by the formula

$$F_X^-(t) = \inf\{x : t \le F(x)\}$$

meaning that $F_X^-(t)$ is a minimum value of $F_X(x)$ which is higher than or equal to t.

Theorem 4.6 (Generalization of the inversion method)**.** *Let U be a random variable distributed according to $U[0,1]$ distribution and $F_X(.)$ be a cdf of some probability distribution. Then, the variable given by the formula*

$$X = F_X^-(U) \tag{4.4}$$

is a sample from $F_X(.)$ (i.e. $X \sim F_X(.)$).

Proof of this theorem may be found, e.g., in (Ripley 1987).

The generalized inversion method implied by Theorem 4.6, may be easily applied also for discrete distributions. Such distribution is considered in the following example:

Example 4.7. *If we want to generate X according to some discrete distribution given by the probabilities p_i, where $\mathrm{P}(X = x_i) = p_i$ for $i = 1, 2, \ldots$, then the transformation*

$$X = \min\left\{k : U \leq \sum_{i=1}^{k} p_i\right\}$$

should be applied.

Even the generalized inverse method may be not adequate for some distributions, because of numerical problems with calculation of $F_X^{-1}(.)$ or $F_X^{-}(.)$. In many cases, finding of the relevant inverse function is not a trivial problem. There are other methods, which may be easier to implement for such distributions. We present some of them in our further considerations and examples.

The normal distribution is an illustration of problems which may be encountered if we apply the inverse method. Obviously, there is no inverse function given by an analytical formula for the cdf of the normal distribution. Therefore, we should use some approximation. One of such approaches is described by the following example:

Example 4.8. *For the inverse function of the cdf of the standard normal distribution the approximation*

$$F_X^{-1}(t) = \begin{cases} g(t) & \text{if } 10^{-20} < t < 0.5 \\ -g(1-t) & \text{if } 0.5 \leq t < 1 - 10^{20} \end{cases}$$

could be used, where

$$g(t) = \sqrt{-2\ln t} - \frac{L(\sqrt{-2\ln t})}{M(\sqrt{-2\ln t})}$$

and $L(.)$ and $M(.)$ are some predetermined fourth degree polynomials (for additional details see (Odeh & Evans 1974)).

It should be emphasized that the inverse method has some important advantages when compare to other general methods. Firstly, as assured by the probability theory, if $U_1, U_2, \ldots \overset{iid}{\sim} U[0,1]$, then also $X_1, X_2, \ldots$ are *iid* random variables. Moreover, in order to obtain one output (i.e. one value X_i) we should generate only one variable U_i. Therefore, this method is numerically very fast, without the unnecessary burden of additional invocation of the function `GenerateU`.

4.1.2 The rejection method

The *rejection method*, sometimes called the *acceptance method* or the *acceptance-rejection method*, was proposed by von Neumann (1951). We discuss three cases of this approach.

We start from the particular, specific case, which is simpler to introduce than the more general approach. Let us suppose that we need a sample from variable X described by pdf (probability density function) $f_X(t)$ which is defined only on the interval $[0,1]$ and is equal to zero elsewhere. Let $f_X(t) \leq M$ for some fixed constant M, if $t \in [0,1]$ (compare with Figure 4.1). Then the following algorithm

Algorithm 4.9 (Rejection method – a particular case)**.**

```
repeat
{
 U1 = GenerateU;
 U2 = GenerateU;
}
until M * U2 <= f_X (U1);
X = U1;
return X.
```

generates the values of random variable X from the density $f_X(.)$.

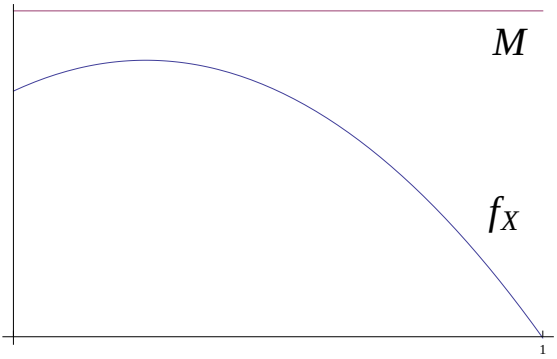

Fig. 4.1. Rejection method – a particular case

This approach could be described by the following steps:

1. Generate the point `(U1,MU2)`. This point is sampled uniformly from the rectangle $[0,1]\times[0,M]$ and the two coordinates of this point are independent.
2. If the point lies below the graph of the function $f_X(.)$, then it is accepted and its coordinate `U1` becomes the output X.
3. Otherwise, the point is rejected and the algorithm returns to the step 1.

In order to consider a more general approach, let us suppose that we could generate variable Y from some fixed pdf $g_Y(t)$, called the *majorizing density* or

the *envelope*, and

$$f_X(t) \leq M g_Y(t) \tag{4.5}$$

for some given constant M for every t for which $f_X(t) > 0$ (see Figure 4.2 for illustration of this assumption).

As in the previous case, described by Algorithm 4.9, we are interested in sampling X from pdf $f_X(t)$. Then we could apply the following algorithm:

Algorithm 4.10 (Rejection method – the general case)**.**

```
repeat
{
 U = GenerateU;
 Y = GenerateG;
}
until M * U * g (Y) <= f (Y);
X = Y;
return X.
```

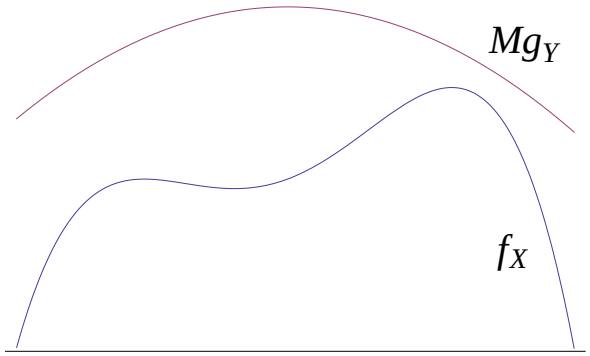

Fig. 4.2. Rejection method – the general case

And this approach could be described by the following steps:

1. Generate $U \sim U[0, 1]$.
2. Use some already available algorithm `GenerateG` to obtain $Y \sim g_Y(.)$.
3. These two variables establish the coordinates of the point `(Y,M*U*g (Y))`. If this point lies below the graph of the function $f_X(.)$, then it is accepted and its coordinate `Y` becomes the output X.
4. Otherwise, the point is rejected and the algorithm returns to the step 1.

The general case of the rejection method is based on the following lemma:

Lemma 4.11. *If the previously mentioned assumptions for the rejection method in the general case are fulfilled, then we have*

$$\mathrm{P}(Y \le x | Y \textit{ is accepted}) = \frac{\int_{-\infty}^{x} f_X(t)dt}{\int_{-\infty}^{\infty} f_X(t)dt} \ . \tag{4.6}$$

Proof. We have

$$\mathrm{P}(Y \le x \wedge Y \text{ is accepted}) = \int_{-\infty}^{x} \frac{f_X(t)}{M g_Y(t)} g_Y(t)\ dt = \int_{-\infty}^{x} \frac{f_X(t)}{M}\ dt \ .$$

Then

$$\mathrm{P}(Y \text{ is accepted}) = \int_{-\infty}^{\infty} \frac{f_X(t)}{M}\ dt$$

and from the definition of the conditional probability

$$\mathrm{P}(Y \le x | Y \text{ is accepted}) = \frac{\mathrm{P}(Y \le x \wedge Y \text{ is accepted})}{\mathrm{P}(Y \text{ is accepted})} \ ,$$

we obtain (4.6). □

From the right hand side of (4.6) we can see that the *normalizing constant* of the density $f_X(t)$, which is given by

$$\frac{1}{\int_{-\infty}^{\infty} f_X(t)dt}$$

is not necessarily taken into account in the rejection method. Therefore, the described algorithm always generates $X \sim f_X(.)$, even if $f_X(t)$ is merely a non-negative function, and not some "real" density. The appropriate normalizing constant is always "automatically calculated and taken into account" by the rejection method.

The constant M plays an important role in the rejection method. Easily seen, if the condition $f_X(t) \le M g_Y(t)$ is fulfilled for some M, then for any $M_1 \ge M$ the condition $f_X(t) \le M_1 g_Y(t)$ is also satisfied. The probability that the generated point is accepted, depends on M. For higher values of M, the probability that the acceptance condition

$$MU g_Y(Y) \le f_X(Y) \tag{4.7}$$

is fulfilled for some Y, is lower, because

$$\mathrm{P}(MU g(Y) \le f(Y)) = \int_{\mathcal{Y}} \int_{0}^{f(y)/Mg(y)} g(y)\ du\ dy = \frac{1}{M} \ .$$

Then the optimal value of M is given by

$$M^* = \min\{M : f_X(.) \le M g_Y(.)\} \ .$$

Our considerations concerning the normalizing constant and finding the appropriate value of the constant M can be illustrated by the following example:

Example 4.12 (Normal distribution)**.** *We will sample X from the standard normal distribution in two stages:*

1. *We focus only on the right half of the distribution $N(0,1)$. Therefore, we generate variable X_1 from pdf $f_{X_1}(t) = \sqrt{2/\pi}\exp(-t^2/2)$ for $t \geq 0$. In this step we apply the rejection method for the exponential distribution* Exp(1) *as the majorizing density to generate the variable X_1.*
2. *We transform the output to variable X_2, by adding "+" or "-" to X_1 with equal probabilities 0.5.*

The algorithm for sampling from the exponential distribution was described in Example 4.5. Then, for $g_Y(t) = e^{-t}$ and $f_{X_1}(t) = \sqrt{2/\pi}\exp(-t^2/2)$, the optimal value of the constant M is equal to $\sqrt{2e/\pi}$. Therefore, the algorithm for the first step is as follows:

Algorithm 4.13 (Rejection method for the normal distribution (part 1))**.**

```
repeat
{
 U = GenerateU;
 Y = GenerateExponential;
}
until Sqrt(2e /pi) * U * exp(-Y)
 <= Sqrt(2/pi) * exp(-Y^2 / 2);
X_1 = Y;
return X_1.
```

In the second step some additional, independent random variable from the Bernoulli distribution (see Example 4.7) is used. The relevant transformation of X_1 into the final output X_2 is done using the following method:

Algorithm 4.14 (Rejection method for the normal distribution (part 2))**.**

```
U = GenerateU;
if U <= 0.5 then X_2 = - X_1
 else X_2 = X_1;
return X_2.
```

As it is easily seen, the acceptance condition in Algorithm 4.13 (which is responsible for the first step in Example 4.12) has the form

$$\sqrt{2e/\pi}U\exp(-Y) \leq \sqrt{2/\pi}\exp(-Y^2/2)\ ,$$

which could be simplified to

$$\sqrt{e}U \exp(-Y) \leq \exp(-Y^2/2) \ . \tag{4.8}$$

In (4.8) the function $f_{X_1}(t)$ is without its normalizing constant, as indicated by Lemma 4.11.

It should be emphasized that the rejection method has some important advantages. First, as stressed by Lemma 4.11, the normalizing constant of $f_X(t)$ is unnecessary for the calculations. Additionally, we do not need to know the cdf of $f_X(t)$. It may be helpful in the cases, which are not suitable for the inversion method. There is also some flexibility in the selection of the pdf as the majorizing density and in the application of the relevant algorithm for sampling from this density.

However, compared with the inversion method, the rejection algorithm displays also some disadvantages of the rejection algorithm. The most important one is that in order to obtain one output (i.e. one value X_i) we should generate at least two (or even more) random variables. Of course, the previously mentioned assumptions of this approach should be also satisfied. Then, the relevant algorithm for sampling from the majorizing density should be known and it should be sufficiently numerically fast. But there may appear problems with finding the appropriate envelope for some classes of densities $f_X(t)$, especially for tails of $f_X(t)$, i.e. when $t \to +\infty$ or $t \to -\infty$.

The rejection method can be generalized to more complex approaches if we are interested in the speed of the algorithm. In Algorithm 4.10, in order to check the acceptance condition

$$MUg_Y(Y) \leq f_X(Y) \tag{4.9}$$

the functions $f_X(.)$ and $g_Y(.)$ have to be evaluated for some given value Y. This could be numerically inconvenient. Therefore, it may be easier to find some numerically simpler functions which fulfil the inequality

$$\alpha_1(t) \leq \frac{f_X(t)}{Mg_Y(t)} \leq \beta_1(t) \tag{4.10}$$

for all t. In this case, if the generated random variables $U \sim U[0,1]$ and $Y \sim g_Y(.)$ satisfy the condition (known as the *fast acceptance condition*)

$$U \leq \alpha_1(Y) \ ,$$

then also the "standard" acceptance condition (4.9) is fulfilled and the value Y should be accepted. Otherwise, if U and Y satisfy the condition (the *fast rejection condition*)

$$U \geq \beta_1(Y) \ ,$$

then the "usual" acceptance condition (4.9) is not fulfilled and the variable Y should be rejected.

It can be easily seen that for checking the fast acceptance condition and the fast rejection condition we do not need to evaluate the exact value of $f_X(.)/g_Y(.)$.

The approach introduced is known as *pretesting* or *squeezing* (see (Marsaglia 1977)) and it may be described by the following algorithm:

Algorithm 4.15 (Pretesting)**.**

```
flag = 0;
repeat
{
 U = GenerateU;
 Y = GenerateG;
 if U <= alpha (Y) then
 {
  flag = 1:
 }
 else
 {
  if U <= beta (Y) then
  {
   if M * U * g (Y) <= f (Y) then
   {
    flag = 1;
   }
  }
 }
}
until flag = 1;
X = Y;
return X.
```

The if-else conditions are nested in the above algorithm in order that the evaluation of $f_X(.)/g_Y(.)$ is done as rarely as possible.

The squeezing method can be also straightforwardly generalized. Instead of the only one fast acceptance / rejection condition, the whole set of limits and the relevant functions can be considered. Let us suppose that

$$\alpha_k(t) \leq \ldots \leq \alpha_1(t) \leq \frac{f_X(t)}{Mg_Y(t)} \leq \beta_1(t) \leq \ldots \beta_l(t) \tag{4.11}$$

for all t. Then the conditions are checked in the natural order

$$U \leq \alpha_k(Y), \;\; U \geq \beta_l(Y), \;\; U \leq \alpha_{k-1}(Y), \ldots, \;\; U \leq \frac{f_X(Y)}{Mg_Y(Y)}$$

until one of them is fulfilled.

4.1.3 The series method

The generalization of the rejection method described by the conditions (4.11) leads to the algorithm known as the *series method*. In this case, instead of the set of conditions (4.11), other methods of approximation of the target density $f_X(.)$ are considered. Let us suppose that for all t we have

$$\underline{f}_n(t) \leq \underline{f}_{n-1}(t) \leq \ldots \underline{f}_1(t) \leq f_X(t) \leq \overline{f}_1(t) \leq \ldots \overline{f}_{n-1}(t) \leq \overline{f}_n(t) \ , \qquad (4.12)$$

so that the relevant density $f_X(.)$ is approximated by the whole sequence of functions $\underline{f}_i(.)$ and $\overline{f}_i(.)$. Then the following algorithm can be applied:

Algorithm 4.16 (Series method)**.**

```
repeat
{
 U = GenerateU;
 Y = GenerateG;
 i = n + 1;
 repeat
 {
  i = i - 1;
  if M * U * g (Y) <= f_i (Y) then return Y;
 }
 until M * U * g (Y) > f^i (Y);
}
until false.
```

It can be easily seen that in the consecutive steps of the above algorithm the pdf $f_X(x)$ is approximated by $\underline{f}_n(x), \overline{f}_n(x), \underline{f}_{n-1}(x), \ldots$.

This approach can be also applied in the case of the convergent series. Let us suppose that the density $f_X(.)$ is given as a limit of the convergent series

$$f_X(t) = \sum_{i=1}^{\infty} S_i(t) \qquad (4.13)$$

and that the remainder of this series could be numerically approximated by the condition

$$\left| \sum_{i=n}^{\infty} S_i(t) \right| \leq R_n(t)$$

for all t. Then the following algorithm can be applied:

Algorithm 4.17 (Convergent series method)**.**

```
repeat
{
 U = GenerateU;
 Y = GenerateG;
 S = 0;
 n = 0;
 repeat
 {
  n = n + 1;
  S = S + S_n (Y);
 }
 until | S - M * U * g (Y) | > R_{n+1} (Y);
}
until M * U * g (Y) <= S;
X = Y;
return X.
```

The internal loop "repeat... until..." evaluates relevant approximations of $f_X(.)$ by summing up functions $S_1(.), S_2(.), \ldots, S_n(.)$, as in (4.13). This operation is finished when the difference between the obtained approximation and the value $MUg_Y(.)$ is lower than the remainder $R_{n+1}(.)$. In such a case the value of $S_n(.)$ plus the additional error $R_{n+1}(.)$ is lower or greater than $MUg_Y(.)$. Then the acceptance condition is checked.

In some cases a particular functional form of the target density $f_X(.)$ may be considered. Such instance is described by the following lemma:

Lemma 4.18. *Let us suppose that*

$$f_X(t) = Mg_Y(t)\,(1 - a_1(t) + a_2(t) - \ldots) \tag{4.14}$$

and

$$a_1(t) \geq a_2(t) \geq \ldots \geq 0$$

for all t*, and* $a_n(t) \to 0$ *for* $n \to \infty$*. Then the acceptance condition* (4.7) *in the rejection method is given by*

$$U \geq a_1(Y) - a_2(Y) + \ldots \tag{4.15}$$

Proof. The acceptance condition in the general case of the rejection method is given by

$$MUg_Y(t) \leq f_X(t)\ ,$$

which is equivalent to

$$(1-U)Mg_Y(t) \le f_X(t)$$

or

$$UMg_Y(t) \ge Mg_Y(t) - f_X(t) \ . \tag{4.16}$$

From the assumption (4.14) we have

$$Mg_Y(t) - f_X(t) = Mg_Y(t)\,(a_1(t) - a_2(t) + \ldots) \ . \tag{4.17}$$

Combining (4.16) and (4.17), we get the condition (4.15). □

4.1.4 The ratio-of-uniforms method

When applying the *ratio-of-uniforms* method (abbreviated as *ROU*, see, e.g. (Kinderman & Monahan 1977)) we assume that a random point is uniformly distributed over some multidimensional set. This statement is more strictly explained by the following definition:

Definition 4.19. *A random point X is uniformly distributed over the set $\mathcal{A} \subset \mathbb{R}^p$, if for any subset $\mathcal{B} \subset \mathcal{A}$ we have*

$$\mathrm{P}(X \in \mathcal{B}) = \frac{l_p(\mathcal{B})}{l_p(\mathcal{A})} \ ,$$

where $l_p(.)$ is a p-dimensional measure of this set.

The ROU method is based on the following theorem:

Theorem 4.20. *Let $f_X(.)$ be a non-negative and finite integrable function and let*

$$\mathcal{C}_f = \left\{(u,v) : 0 \le u \le \sqrt{f_X\left(\frac{v}{u}\right)}\right\} \ . \tag{4.18}$$

If the point (U,V) is uniformly distributed over the set $\mathcal{C}_f$, then variable $X = \frac{V}{U}$ has pdf equal to $\frac{f_X(.)}{\int f_X(t)\,dt}$.

Proof. If (U,V) is the random point uniformly distributed over the set $\mathcal{C}_f$, then the joint pdf of this point is equal to

$$f_{(U,V)}(u,v) = \frac{1}{l_2(\mathcal{C}_f)} \mathbb{1}_{\mathcal{C}_f}(u,v) \ .$$

If the transformation

$$X = \frac{V}{U} \ , \ Y = U$$

is applied, then the Jacobian of this function is given by

$$\left|\det \begin{pmatrix} 0 & 1 \\ y & x \end{pmatrix}\right| = y \ .$$

Therefore, the joint pdf of the new variables (X, Y) is equal to

$$f_{(X,Y)}(x,y) = \frac{1}{l_2(\mathcal{C}_f)} y \mathbb{1}_{\mathcal{C}_f}(y, xy) = \frac{y}{l_2(\mathcal{C}_f)} \mathbb{1}_{[0,\sqrt{f(x)}]}(y) \ ,$$

which leads to

$$f_X(x) = \int f_{(X,Y)}(x,y)\, dy = \frac{1}{l_2(\mathcal{C}_f)} \int_0^{\sqrt{f(x)}} y\, dy = \frac{1}{2l_2(\mathcal{C}_f)} f(x) \ ,$$

where $\frac{1}{2l_2(\mathcal{C}_f)}$ is the relevant normalizing constant. □

Theorem 4.20 leads directly to the following algorithm, which constitutes the ROU method:

Algorithm 4.21 (ROU method)**.**

```
repeat
{
 (U,V) = GenerateCf;
 X = V / U;
}
until U^2 <= f(X);
return X.
```

The function `GenerateCf` in the above algorithm generates the random point (U, V), distributed uniformly over the set $\mathcal{C}_f$. Of course, the form of such function depends on the shape of $\mathcal{C}_f$.

It should be noted that according to Theorem 4.20, we do not have to know the normalizing constant for the function $f_X(.)$. As in the case of the rejection method, this constant is automatically taken into account during sampling.

The Algorithm 4.21 can be described by the following steps:

1. Generate the point (U, V) uniformly distributed over the set $\mathcal{C}_f$.
2. Calculate $X = \frac{V}{U}$.
3. If the point (U, V) is inside the set $\mathcal{C}_f$ (compare the acceptance condition in Algorithm 4.21 with the description of the set $\mathcal{C}_f$ given by (4.18)), return X. Otherwise, the algorithm returns to step 1.

As can be easily seen, the most crucial point for this method is application of the function `GenerateCf` – in other words, finding the procedure of sampling from the set $\mathcal{C}_f$. But the relevant method is straightforward, if $\mathcal{C}_f$ is inside some rectangle and then the rejection sampling could be used to check the condition $U^2 \leq f(X)$ in Algorithm 4.21.

Some properties of $\mathcal{C}_f$ may be directly observed, as stated by the following remark:

Remark 4.22. *The set $\mathcal{C}_f$ can be described by the following conditions:*

- *From the definition of the set* (4.18) *we get $u \geq 0$ (i.e. we should take into account only the right hand part of the coordinate system),*
- *If $f_X(.)$ is symmetrical around zero, then the set $\mathcal{C}_f$ is symmetrical with respect to axis u,*
- *If $f_X(t) \geq 0$ only for $t \geq 0$, then we have*

$$f_X\left(\frac{v}{u}\right) \geq 0 \Rightarrow \frac{v}{u} \geq 0 \Rightarrow v \geq 0 \ ,$$

which leads to $u \geq 0, v \geq 0$ (i.e. only the upper right part of the coordinate system should be taken into account).

The standard approach to describe the set $\mathcal{C}_f$ in a more precise way is to apply

$$z = \frac{v}{u} \tag{4.19}$$

which characterizes the boundaries of $\mathcal{C}_f$ in a parametric way. From the definition of $\mathcal{C}_f$, given by (4.18) and the formula (4.19) we get

$$u(z) = \sqrt{f(z)} \ , \ v(z) = z\sqrt{f(z)} \ ,$$

so that

$$0 \leq u \leq \sup_z \sqrt{f(z)} \ , \ \inf_z z\sqrt{f(z)} \leq v \leq \sup_z z\sqrt{f(z)} \ .$$

The above conditions describe the method for enclosing the set $\mathcal{C}_f$ in some given rectangle. We illustrate this approach in the two following examples:

Example 4.23. *Let $X \sim \mathrm{Exp}(\lambda)$. Because the normalizing constant of pdf is unnecessary in the case of the ROU method, then $f_X(t) = \exp(-\lambda t)$. Therefore*

$$\sup_t \sqrt{e^{-\lambda t}} = 1 \ , \ \inf_t t\sqrt{e^{-\lambda t}} = 0 \ , \ \sup_t t\sqrt{e^{-\lambda t}} = \frac{2}{\lambda e} \ ,$$

and

$$0 \leq u \leq 1 \ , \ 0 \leq v \leq \frac{2}{\lambda e} \ . \tag{4.20}$$

Then, the set $\mathcal{C}_f$ is contained in the rectangle $[0,1] \times [0, \frac{2}{\lambda e}]$ and the pdf of the uniform distribution over this rectangle is given by

$$g_{\mathcal{C}_f}(t) = \frac{\lambda e}{2} \ .$$

Example 4.24. *Let $X \sim N(0,1)$. Then*

$$f_X(t) = e^{-\frac{t^2}{2}} \ ,$$

therefore

$$\sup_t \sqrt{e^{-\frac{t^2}{2}}} = 1 \ , \ \inf_t t\sqrt{e^{-\frac{t^2}{2}}} = -\sqrt{\frac{2}{e}} \ , \ \sup_t t\sqrt{e^{-\frac{t^2}{2}}} = \sqrt{\frac{2}{e}} \ ,$$

which gives

$$0 \le u \le 1 \ , \ -\sqrt{\frac{2}{e}} \le v \le \sqrt{\frac{2}{e}} \ . \tag{4.21}$$

4.1.5 The composition method

We discuss here three cases of the *composition method*, which enables sampling of the random variable.

Let us start from the *series form*, when the considered pdf $f_X(.)$ could be written as

$$f_X(t) = \sum_{i=1}^{\infty} p_i f_i(t) \ , \tag{4.22}$$

where $p_i > 0, \sum_{i=1}^{\infty} p_i = 1$ and $f_1(.), f_2(.), \ldots$ are densities of some distributions. In such a case, $f_X(.)$ is said to be a *mixture* or *composition* of other distributions.

The formula (4.22) and the mentioned assumptions about p_i and $f_i(.)$ lead directly to the following steps of the variable generation:

1. Generate random variable K according to the probabilities $p_1, p_2, \ldots$.
2. Sample from the density $f_K(.)$ specified by the value K from the previous step.

Then the method described above is directly equivalent to the following algorithm:

Algorithm 4.25 (Composition method I)**.**

```
K = GenerateK;
X = GenerateF (K);
return X.
```

In the above description, the function `GenerateK` generates the variable according to the probabilities $p_1, p_2, \ldots$, and `GenerateF (K)` generates a sample from the density $f_K(.)$.

Usually, the form (4.22) is simplified into the case with finite discrete distribution of p_i, i.e.

$$f_X(t) = \sum_{i=1}^{m} p_i f_i(t) \ . \tag{4.23}$$

For some instances, the whole support of $f_X(.)$ may be divided into separate intervals $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$ (or $\mathcal{A}_1, \mathcal{A}_2, \ldots$ in infinite case) in order that sampling from each pdf $f_i(.)$ be more straightforward. A similar method was used in

Example 4.12, where the whole set $\mathbb{R}$ was divided into subsets $\mathbb{R}_-$ and $\mathbb{R}_+$. Then, we applied the rejection sampling for the right tail of the distribution $N(0,1)$ with the exponential distribution as its envelope.

If the intervals $\mathcal{A}_1, \mathcal{A}_2, \ldots$, mentioned before, are used, then from (4.22) we get

$$p_i = \int_{\mathcal{A}_i} f_X(t)\, dt \ , \ f_i(t) = \frac{f_X(t)}{p_i} \mathbb{1}_{\mathcal{A}_i}(t) \ .$$

Therefore, knowledge about the probabilities p_i, determined by pdf $f_X(.)$ for each subset $\mathcal{A}_i$, is necessary to apply this approach.

The composition method is particularly useful if it is mixed with the rejection approach (as in Example 4.12). For intervals smaller than the whole support of $f_X(t)$ it is easier to find the envelopes $g_i(.)$ for the relevant parts $f_i(.)$ of the target density $f_X(.)$.

Instead of (4.22), we could apply a more general, *integral form*, given by

$$f_X(t) = \int_{\mathcal{Z}} f_z(t) h(z)\, dz \ , \tag{4.24}$$

where $f_z(.)$ is some pdf for any value of z and $h(z)$ is a density for some set $\mathcal{Z}$ (e.g. $\mathcal{Z} \subset \mathbb{R}$).

The integral (4.24) has important interpretation, related to the concept of conditional density. The density $f_z(.)$ may be seen as a conditional pdf $f(.|z)$ and $h(z)$ may be then identified with the density describing the distribution of the parameter z. Therefore, $f_X(.)$ is the unconditional density for such setting. This approach is close to the method known as Gibbs sampler, which will be discussed in Section 6.3. Obviously, if we know the appropriate conditional densities, the form (4.24) may be used to sample from the unconditional pdf.

The formula (4.24) leads to the following algorithm:

Algorithm 4.26 (Composition method II)**.**

```
Z = GenerateH;
X = GenerateF (Z);
return X.
```

Another approach to the composition method is related to the *polynomial form* of the target density $f_X(t)$. Let us assume that the relevant pdf may be written as

$$f_X(t) = \sum_{i=0}^{m} c_i t^i \ , \tag{4.25}$$

where $c_i \geq 0$ and $t \in [0,1]$. Then we get

$$\sum_{i=0}^{m} \frac{c_i}{i+1} = 1 \ , \tag{4.26}$$

because from (4.25) we have

$$\int_0^1 \sum_{i=0}^{m} c_i t^i = \sum_{i=0}^{m} \left(\frac{c_i}{i+1} t^{i+1} \right) \Big|_0^1 = \sum_{i=0}^{m} \frac{c_i}{i+1}$$

and $f_X(t)$ is some density.

Using the conclusion (4.26), the following steps of the procedure of sampling from the density $f_X(t)$ of the form (4.25) can be applied:

Algorithm 4.27 (Composition method for the polynomials)**.**

1. Generate variable $I \in \{0, 1, 2, \ldots, m\}$ according to the discrete distribution, given by the probabilities $\mathrm{P}(I = i) = \frac{c_i}{i+1}$.
2. For the fixed value I, generate random variable according to the density $(i+1)t^i$. In order to do this, e.g. the inversion method could be applied.

When not all of $c_i \geq 0$, we may introduce $c_i^+ > 0$ and $c_i^- > 0$ such that $c_i = c_i^+ - c_i^-$. Then we have

$$\sum_{i=1}^{m} c_i t^i \leq \sum_{i=1}^{m} c_i^+ t^i$$

and the function

$$g(t) = \sum_{i=1}^{m} c_i^+ t^i \ ,$$

after adding the necessary normalizing constant, could be used as the majorizing density for the $f_X(t)$ in the rejection sampling.

4.2 Methods for discrete distributions

The discrete distributions, given by the probabilities $\mathrm{P}(X = i) = p_i$ for $i = 1, 2, \ldots$ constitute an important class of distributions appearing in various practical applications. Therefore, we discuss some special sampling methods, which are suitable for this type of distributions.

4.2.1 The generalized inverse method

The most straightforward approach to generate X from a discrete distribution $p_1, p_2, \ldots$ is to apply the generalized inverse method (compare with Example 4.7), which leads to the following algorithm:

Algorithm 4.28 (Generalized inverse method for the discrete distribution)**.**

```
S = 0;
U = GenerateU;
I = 0;
do
{
 I = I + 1;
 S = S + p_I;
}
while (S <= U);
X = I;
return X.
```

This algorithm can be described by the following steps:

1. Divide the whole interval $[0, 1]$ into parts, which are equivalent to the probabilities $p_1, p_2, \ldots$.
2. Generate $U \sim U[0, 1]$.
3. Comparing the cumulative sums $S = p_1 + p_2 + \ldots + p_I$ with U, find the relevant value I, for which the random point U belongs to the interval $[p_{I-1}, p_I]$.

The expected number of comparisons of the cumulated sums S with the sampled value U is equal to $\mathbb{E}\, X$. Therefore, it may be numerically profitable to increase the speed of this algorithm, which can be done in many ways. For example, the whole table of cumulative sums $S_1 = p_1, S_2 = p_1 + p_2, \ldots$ may be prepared during the set-up phase of the algorithm. Another approach is to reorder the probabilities p_i into decreasing order, also during the set-up phase. Additionally, the algorithm can be started in "another place". If the considered discrete distribution is unimodal, we could start the search to the left or to the right of this mode.

4.2.2 The Alias method

Walker (1977) proposed a more efficient method, which is known as the *Alias method*, which is valid for the finite discrete distributions given by the probabilities

$$\mathrm{P}(X = 1) = p_1, \mathrm{P}(X = 2) = p_2, \ldots, \mathrm{P}(X = m) = p_m \ .$$

Let us assume that we have a sequence $q_1, q_2, \ldots, q_m$, $0 \le q_i \le 1$ for $i = 1, 2, \ldots, m$ and $A(1), A(2), \ldots, A(m) \in \{1, 2, \ldots, m\}$. These two sequences are depend on each other via the condition

$$p_i = \left(q_i + \sum_{j:A(j)=i} (1 - q_j) \right) \frac{1}{m}$$

for $i = 1, 2, \ldots, m$.

Then the sampling algorithm may be written down as

Algorithm 4.29 (Alias method)**.**

```
I = GenerateU[m];
U = GenerateU;
if U <= q_I then
{
 X = I;
}
else
{
 X = A (I);
}
return X.
```

The function `GenerateU[m]` in the above algorithm samples the variable $I \in \{1, 2, \ldots, m\}$ from the uniform discrete distribution, given by the probabilities $\mathrm{P}(I = 1) = \mathrm{P}(I = 2) = \ldots = \mathrm{P}(I = m) = \frac{1}{m}$.

The sequence $q_1, q_2, \ldots, q_m$ has the interpretation of the set of "cutting points", for which the final output value is changed from the previously sampled value I to its alias given by $A(I)$. Then, the Algorithm 4.29 may be described by the following steps:

1. Generate I uniformly over the set $\{1, 2, \ldots, m\}$.
2. Generate $U \sim U[0, 1]$.
3. If $U \leq q_I$ (i.e. U is not greater than the relevant "cutting point" for the fixed I), then this value I is returned. Otherwise, $A(I)$ (i.e. the alias of the value I) is returned.

As it can be easily seen, the main problem in the case of the Alias algorithm is to find the sequences q_i and $A(i)$ for the given probabilities p_i. These sequences (which may be also seen as tables or allocations) are not unique (see Figures 4.3 and 4.4 for an example of such allocation). However, they can be found if, e.g., the following algorithm is applied (see (Devroye 1986) for additional details):

Algorithm 4.30 (Evaluation of the sequences q_i and $A(i)$)**.**

```
wTable = Table[m];
n = 0;
p = m + 1;
for i = 1 to m do
{
 q_i = m * p_i;
 if q_i < 1 then
 {
  n = n + 1;
  wTable [n] = i;
 }
 else
 {
  p = p - 1;
  wTable [p] = i;
 }
}
for k = 1 to m - 1 do
{
 i = wTable [k];
 j = wTable [p];
 A (i) = j;
 q_j = q_j + q_i - 1;
 if q_i < 1 then
 {
  p = p + 1;
 }
}
```

In this algorithm, `wTable` is a working table of the length m.

4.3 Algorithms for the normal distribution

As we know, normal distribution is one of the most important distributions in statistics and in statistical modelling. Therefore, there is a need for numerically fast and stable algorithms to sample from this distribution. However, only the algorithms for sampling from the standard normal distribution $N(0,1)$ should be considered, because if $X \sim N(0,1)$, then

$$Y = \mu + \sigma X \sim N(\mu, \sigma^2) \tag{4.27}$$

as stated by the probability theory.

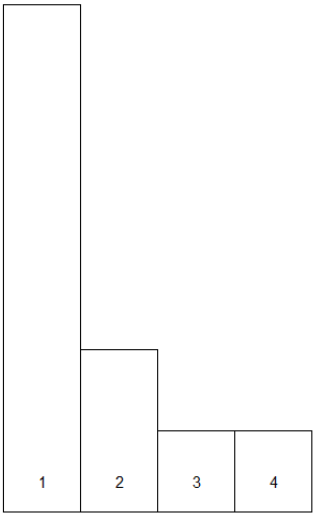

Fig. 4.3. Example of the Alias method – before allocation

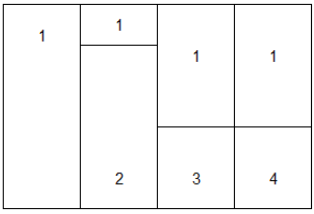

Fig. 4.4. Example of the Alias method – after allocation

4.3.1 A simple algorithm

One of the simplest and most primitive algorithms for sampling from $N(0,1)$ is given by the following pseudocode:

Algorithm 4.31 (Simple algorithm for the normal distribution)**.**

```
for i = 1 to 12 do
{
 X(i) = GenerateU;
}
X = X(1) + X(2) + ... + X(12) - 6;
return X.
```

The approximation, ensured by the Central Limit Theorem is fairly good in this case, but the usefulness of this approach is rather limited because of the burden of sampling of additional random variables. To generate only one output value, twelve values $U_i \sim U[0,1]$ are necessary.

4.3.2 The Box–Muller algorithm

Box & Muller (1958) introduced the following algorithm, which generates two independent standard normal variables:

Algorithm 4.32 (Box–Muller algorithm)**.**

```
U1 = GenerateU;
U2 = GenerateU;
Phi = 2 * Pi * U1
R = Sqrt ( - 2 * ln ( U2 ));
X1 = R * cos ( Phi );
X2 = R * sin ( Phi );
return (X1, X2).
```

This algorithm can be described by the following steps:

1. Generate two independent variables $U_1, U_2 \sim U[0,1]$.
2. Multiply U_1 by 2π, which gives the angle coordinate Φ.
3. Transform U_2 into variable R, where $R^2 \sim \text{Exp}(1/2)$ (as in the case described in Example 4.5), which gives the length of the radius in polar coordinates.
4. Calculate X_1, X_2 based on the polar coordinates Φ and R.

This algorithm is practically very useful, if we have numerically fast procedures for calculation of $\sin(x), \cos(x)$ and $\ln(x)$. Additionally, two *iid* standard normal samples are generated from only two variables $U_i \sim U[0,1]$.

To verify the approach described in Algorithm 4.32, let us consider the joint pdf of the pair (X_1, X_2) of independent standard normal variables

$$f_{(X_1,X_2)}(t_1,t_2) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(t_1^2+t_2^2\right)\right) .$$

If we transform (X_1, X_2) into (R, Φ) in polar coordinates, the relevant joint pdf is given by

$$f_{(R,\Phi)}(t_1,t_2) = \left(\frac{1}{2\pi}\exp\left(-\frac{1}{2}r^2\right)\right)\left|\begin{array}{cc}\cos\phi & \sin\phi \\ -r\sin\phi & r\cos\phi\end{array}\right| = \frac{1}{2\pi}re^{-r^2/2}$$

for $R \in (0,\infty)$ and $\Phi \in [0,2\pi)$, where R and Φ are independent. Additionally, $R^2 = X_1^2 + X_2^2$ has a χ^2 distribution with two degrees of freedom, which is also Exp(1/2) distribution. To obtain Algorithm 4.32, the above transformation to polar coordinates is inverted in the next step.

4.3.3 The Marsaglia's polar method

The method described by Marsaglia & Bray (1964) is very similar to the Box–Muller approach, given by Algorithm 4.32. In this case we have the following algorithm:

Algorithm 4.33 (Marsaglia's polar method)**.**

```
repeat
{
 U1 = GenerateU;
 U2 = GenerateU;
 U1 = 2 * U1 - 1;
 U2 = 2 * U2 - 1;
 W = U1^2 + U2^2;
}
until W < 1;
C = sqrt ( - 2 * W^(-1) * ln ( W ) );
X1 = C * U1;
X2 = C * U2;
return (X1, X2).
```

The above procedure can be described by the following steps:

1. Generate two independent variables $U_1, U_2 \sim U[0,1]$.
2. Transform these variables into $U_1, U_2 \sim U[-1,1]$.
3. Calculate $W = U_1^2 + U_2^2$. If the point (U_1, U_2) is outside of the unit disc, reject this point and return to the step 1. Otherwise, go to the next step.
4. The point (U_1, U_2) specifies the angle uniformly distributed on the interval $[0, 2\pi)$. Calculate C, which gives value of the shift on the straight line, determined by this angle.
5. Find a new point (X_1, X_2), based on the calculated angle and shift C.

Let (R, Φ) denote polar coordinates, as previously. Then

$$R^2 = C^2\left(U_1^2 + U_2^2\right) = C^2 W = \left(-\frac{2}{W}\ln W\right) W = -2\ln W \ ,$$

and

$$X_1 = \sqrt{-2\ln W}\cos\Phi = \sqrt{-2\ln W}\,\frac{U_1}{\sqrt{W}} = CU_1 \ , \ X_2 = \sqrt{-2\ln W}\sin\Phi = CU_2 \ ,$$

which shows the similarities between the Algorithms 4.32 and 4.33.

4.3.4 The Marsaglia–Bray algorithm

Marsaglia & Bray (1964) introduced also a more complex method for sampling from $N(0,1)$ distribution. This approach constitutes also an interesting example, which may be related to a special case of the composition method, discussed in Section 4.1.5.

In this approach, the standard normal distribution is decomposed into separate parts. In the first step, we focus on the tails of this distribution over the intervals $(-\infty,-3]\cup[3,\infty)$. It is possible to sample from these tails by applying other methods of random values generation, e.g. the rejection method with the exponential distribution as the majorizing density. The probability that the $N(0,1)$ variable is sampled from the mentioned intervals, is equal to

$$p_4 = \mathrm{P}(|X|>3) \approx 0.0027 \ .$$

Therefore, we may focus next only on the interval $[-3,3]$.

For the interval $[-3,3]$ the normal pdf may be approximated by the mixture of parabolas, given by

$$f_1(t) = \begin{cases} \frac{3-t^2}{8} & \text{if } |t|<1 \\ \frac{(3-|t|)^2}{16} & \text{if } 1\le|t|\le 3 \\ 0 & \text{otherwise} \end{cases} \ .$$

For this approximation we find the maximal value of p_1, fulfilling the condition

$$f_{N(0,1)}(t) - p_1 f_1(t) \ge 0 \ ,$$

which gives $p_1 \approx 0.86$. Therefore, with this (rather high) probability the output could be sampled from pdf $f_1(.)$, which is also the density of $U_1+U_2+U_3$ for $U_1,U_2,U_3 \overset{iid}{\sim} U[-1,1]$.

The "remnant" of the density, i.e. $f_{N(0,1)}(.) - p_1 f_1(.)$, is decomposed once again. In order to do this, the density given by

$$f_2(t) = \begin{cases} \frac{4}{9}\left(\frac{3}{2}-|t|\right) & \text{if } |t|\le\frac{3}{2} \\ 0 & \text{otherwise} \end{cases}$$

is applied, which is the pdf for

$$\frac{3(U_1+U_2-1)}{2} \ ,$$

where $U_1,U_2 \overset{iid}{\sim} U[0,1]$. As previously, the maximum value of p_2, fulfilling the condition

$$f_{N(0,1)}(t) - p_1 f_1(t) - p_2 f_2(t) \ge 0$$

should be found, which leads to $p_2 \approx 0.111$.

During the last step, the appropriate density, related to the "remaining" probability $p_3 = 1 - p_1 - p_2 - p_4 \approx 0.0226$ is used. It means that we should sample from the density

$$f_3(t) = \frac{1}{p_3}\left(f_{N(0,1)}(t) - p_1 f_1(t) - p_2 f_2(t) - p_4 f_4(t)\right)$$

if $t \in [-3, 3]$, with very low probability p_3. In order to generate random variable according to the pdf $f_3(.)$ other methods can be used, like the rejection sampling, because $f_3(.)$ is bounded on the interval $[-3, 3]$.

4.4 Example of sampling from other distributions

There are also various special algorithms for other random distributions, not only for the normal one, as described in Section 4.3. Usually they are stated in the form of mathematical theorems, which could be then directly transformed into numerical recipes. The example of such an approach is presented for the beta distribution.

Definition 4.34 (Beta distribution)**.** *A distribution of the random variable X is given by a beta distribution (further we use notation $X \sim B(\alpha, \beta)$) if the density of this distribution is equal to*

$$f_X(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1-t)^{\beta-1}$$

for $t \in (0, 1)$. If $X \sim B(\alpha, \beta)$, then $\mathbb{E}\, X = \frac{\alpha}{\alpha+\beta}$, $\operatorname{Var} X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

To sample from beta distribution, the following theorem can be used (see, e.g. (Wieczorkowski & Zieliński 1997)):

Theorem 4.35. *If $U, V \sim U[0, 1]$ are independent, then conditional distribution of the variable*

$$X = \frac{U^{\frac{1}{\alpha}}}{U^{\frac{1}{\alpha}} + V^{\frac{1}{\beta}}} \tag{4.28}$$

under the condition

$$U^{\frac{1}{\alpha}} + V^{\frac{1}{\beta}} \leq 1 \tag{4.29}$$

is given by the beta distribution with parameters (α, β).

As it can be easily seen, the above theorem is equivalent to the following algorithm:

Algorithm 4.36 (Beta distribution)**.**

```
repeat
{
 U = GenerateU;
 V = GenerateU;
}
until U^(1/alpha) + V^(1/beta) <= 1;
X = U^(1/alpha) / (U^(1/alpha) +  V^(1/beta));
return X.
```

4.5 Multivariate generation – general remarks

Before discussion of the methods, which are useful for the multivariate number generation, we introduce the necessary notation. Denote by $\mathbb{X}$ the p-dimensional vector, given by the coordinates $\left(X^{(1)}, X^{(2)}, \ldots X^{(p)}\right)$. In this case, the upper index indicates the running number of the coordinate of this vector, and the lower index – as previously – indicates the number of the variable in a sequence of vectors. Then we have

$$\mathbb{X}_i = \left(X_i^{(1)}, X_i^{(2)}, \ldots, X_i^{(p)}\right) \ .$$

If each of the coordinates of the random vector is independent, then to generate the whole sequence of random vectors $\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \ldots$, p independent random variables should be sampled for each vector. In this case, the methods discussed in the previous sections can be directly applied. It can be easily seen that some of these algorithms (like the Box–Muller algorithm, described in Section 4.3.2) generate multivariate independent variables without any additional modification. In the case of the Box–Muller algorithm we obtain two *iid* variables sampled from the standard normal distribution after one realization of this algorithm.

However, if the relevant coordinates of the multivariate random number are not independent, then to generate such variable the methods described in Section 4.1 could be also applied, especially the rejection method and the composition method. But in the case of the rejection approach, selection of the appropriate envelope might be problematic. This problem is related to the effect known as the *curse of dimensionality*, which will be discussed in Section 4.6.1 in a more detailed way.

4.6 Multivariate uniform distribution on the unit ball

In this section we focus on selected methods which can be applied to sample from the multivariate uniform distribution on the p-dimensional closed unit ball $B^p =$

$\left\{(x_1, \ldots, x_p) : \sum_{i=1}^{p} x_i^2 \leq 1\right\}$. The most naive approach is based on the rejection method (see also (Wieczorkowski & Zieliński 1997) for additional details):

Example 4.37. *Let us assume that the vector* $\mathbb{X}$ *is sampled from multivariate uniform distribution on the p-dimensional cube* $C^p = \{(x_1, \ldots, x_p) : \forall_i \, |x_i| \leq 1\}$. *In such a case*

$$\mathbb{X} = (U_1, U_2, \ldots, U_p)$$

and the relevant coordinates $U_1, \ldots, U_p$ *can be generated independently according to* $U[-1, 1]$ *distribution. If the sampled point* $\mathbb{X}$ *is inside the unit ball* B^p*, i.e. its coordinates fulfil the condition*

$$U_1^2 + U_2^2 + \ldots + U_p^2 \leq 1 \ ,$$

then this point should be accepted. Otherwise, a new variable $\mathbb{X}$ *from* C^p *should be generated again.*

Obviously, the approach described in Example 4.37 is the rejection method. In this case the multivariate uniform distribution on C^p plays the role of the envelope.

The method from Example 4.37 leads to the following algorithm:

Algorithm 4.38 (Rejection method for the uniform distribution on B^p)**.**

```
repeat
{
 for i=1 to p do
 {
  U[i] = 2 * GenerateU - 1;
 }
}
until U[1]^2 + U[2]^2 + ... + U[p]^2 <= 1;
X = (U[1], U[2],..., U[p]);
return X.
```

4.6.1 Curse of dimensionality

As it is stated by its name, during the rejection method some of the generated values are rejected. The problem arises when the number of the rejected points increases faster than linearly in relation to the number of dimensions of the considered multivariate distribution. Such phenomenon is known as the *curse of dimensionality* and it could be observed also in the case of Algorithm 4.38.

The probability $\mathrm{P}(p)$ that the generated point is accepted is equal to the ratio of p-dimensional volume of B^p

$$l_p\left(B^p\right) = \frac{2\pi^{\frac{p}{2}}}{p\Gamma\left(\frac{p}{2}\right)}$$

to p-dimensional volume of C^p, which gives

$$\mathrm{P}(p) = \frac{l_p\left(B^p\right)}{l_p\left(C^p\right)} = \frac{\pi^{\frac{p}{2}}}{p2^{p-1}\Gamma\left(\frac{p}{2}\right)} \ . \tag{4.30}$$

Additionally, we have

$$\lim_{p\to+\infty} \frac{l_p\left(B^p\right)}{l_p\left(C^p\right)} = 0 \ .$$

From the geometrical distribution, the expected number $\mathbb{E}\, N_p$ of the points generated before the first output is accepted, is equal to the inverse of the probability (4.30). Therefore, the numerical effectiveness of Algorithm 4.38 is relatively low for the higher dimensions, as indicated by Table 4.1.

p	$\mathrm{P}(p)$	$\mathbb{E}\, N_p$
2	0.7853981634	1.273239545
5	0.1644934067	6.079271019
10	0.002490394570	401.5427965
20	$2.461136950 \cdot 10^{-8}$	$4.063162758 \cdot 10^{7}$
50	$1.536743396 \cdot 10^{-28}$	$6.507267270 \cdot 10^{27}$

Table 4.1. Probability of acceptance and the expected number of generated points for the naive approach

4.6.2 The method based on polar coordinates

In the case of $p = 2$ (i.e. the two-dimensional ball), the following algorithm, based on polar coordinates, can be applied:

Algorithm 4.39 (Polar coordinates for the uniform distribution on B^2)**.**

```
Phi = 2 * Pi * GenerateU;
X1 = cos (Phi);
X2 = sin (Phi);
U = GenerateU;
Y1 = sqrt (U) * X1;
Y2 = sqrt (U) * X2;
return (Y1, Y2).
```

This algorithm can be described by the following steps:

1. Generate $\Phi \sim U[0,1]$, which gives the angle coordinate after multiplication by 2π.
2. Apply trigonometric functions cos(.) and sin(.) for the sampled value Φ to obtain the point (X_1, X_2), which lies on the unit circle S^2.
3. Transform the output (X_1, X_2) into the new point (Y_1, Y_2), which lies inside the ball B^2.

The transformation mentioned in the third step is done via the formula

$$Y_1 = \sqrt{U}X_1 \ , \ Y_2 = \sqrt{U}X_2 \ , \tag{4.31}$$

where $U \sim U[0,1]$, instead of the (more "intuitive") function

$$Y_1 = UX_1 \ , \ Y_2 = UX_2 \ . \tag{4.32}$$

If the transformation (4.32) is applied, then the generated variables are not uniformly distributed on B^2. On the contrary, the obtained variables are clustered near the point $(0,0)$ (as shown by Figure 4.5).

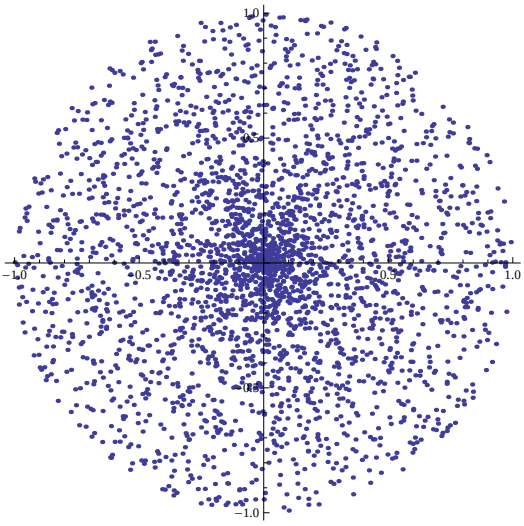

Fig. 4.5. Incorrect sampling from uniform distribution on B^2

Therefore, other kind of transformation, instead of the linear one (given by (4.32)), should be used. In the case of B^p, the length of the radius r is distributed according to pdf given by

$$h(r) = pr^{p-1} \tag{4.33}$$

for $0 \leq r \leq 1$ (see (Wieczorkowski & Zieliński 1997)). Therefore, the relevant cdf is given by $H(r) = r^p$, which leads to the formula (4.31) if the inversion method for $p = 2$ is applied.

But Algorithm 4.39 cannot be generalized for the higher number of dimensions, i.e. $p \geq 3$. Application of a similar transformation leads to the incorrect algorithms for these cases.

4.6.3 Reduction of dimension

Instead of using the previously described methods, in order to sample from the uniform distribution on B^p, we may reduce the dimension of the problem. This may be done via the following lemma (see (Wieczorkowski & Zieliński 1997)):

Lemma 4.40. *If* $\mathbb{Z} = \left(Z^{(1)}, Z^{(2)}, \ldots, Z^{(p)}\right)$ *is sampled uniformly on p-dimensional sphere* $S^p = \left\{(x_1, \ldots, x_p) : \sum_{i=1}^{p} x_i^2 = 1\right\}$, *R is a random variable distributed according to the pdf given by*

$$g(t) = \begin{cases} \frac{ct^{p-1}}{\sqrt{1-t^2}} & \text{if } 0 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases} , \tag{4.34}$$

where c is the relevant normalizing constant, and S is a "random sign" given by the probabilities

$$\mathrm{P}(S = 1) = \mathrm{P}(S = -1) = \frac{1}{2} , \tag{4.35}$$

then the $(p+1)$*-dimensional random variable*

$$\left(RZ^{(1)}, RZ^{(2)}, \ldots, RZ^{(p)}, S\sqrt{1 - R^2}\right) \tag{4.36}$$

is sampled uniformly on S^{p+1}.

Lemma 4.40 leads to the following inductive algorithm:

Algorithm 4.41 (Reduction of dimension for the uniform distribution on B^p)**.**

1. Generate $Z^{(1)}$ uniformly on S^1, where such a sphere is equivalent to two points -1 and 1, i.e. use the probabilities $\mathrm{P}(Z^{(1)} = -1) = \mathrm{P}(Z^{(1)} = 1) = \frac{1}{2}$.
2. Sample R according to the pdf given by (4.34) for $p = 2$ and generate S described by the probabilities (4.35). Then calculate $\left(Z^{(1)}, Z^{(2)}\right)$ given by the formula (4.36).
3. ...

p-1. Sample R according to the pdf given by (4.34) for $p-1$ and generate S described by the probabilities (4.35). Then calculate $\left(Z^{(1)}, Z^{(2)}, \ldots, Z^{(p)}\right)$ given by the formula (4.36).

It should be noted that, in order to sample R from the relevant density (4.34), the variable R is obtained as a square root from the random variable generated according to $B(\frac{p}{2}, \frac{1}{2})$. It means that we should calculate $\sqrt{Y_1/(Y_1+Y_2)}$, where $Y \sim \Gamma(p/2, 1)$ and $Y_2 \sim \Gamma(1/2, 1)$ (see (Wieczorkowski & Zieliński 1997) for additional details).

4.6.4 The method based on normal distribution

A more direct method for sampling from the uniform distribution on B^p is based on the multivariate standard normal distribution, but with the relevant transformation of the obtained output.

Definition 4.42. *We say that p-dimensional random variable*

$$\mathbb{X} = \left(X^{(1)}, X^{(2)}, \ldots, X^{(p)}\right)$$

has spherically contoured *distribution, if its pdf* $g_{\mathbb{X}}(t_1, t_2, \ldots, t_p)$ *depends only on* $\sum_{i=1}^{p} t_i^2$.

The multivariate distribution, given by p independent standard normal variables, is an example of the spherically contoured distribution. It is easy to see that spherically contoured distribution is invariant for each turn in $\mathbb{R}^p$ (see (Wieczorkowski & Zieliński 1997) for additional discussion). Therefore, the relevant multivariate normal distribution can be directly applied:

Algorithm 4.43 (Spherically contoured distribution for the uniform distribution on B^p)**.**

```
for i = 1 to p do
{
 Z[i] = GenerateNormStd;
}
R^2 = Z[1]^2 + ... + Z[p]^2;
for i = 1 to p do
{
 Y[i] = Z[i] / R;
}
U = GenerateU;
R1 = U^(1/p);
for i = 1 to p do
{
 X[i] = Y[i] * R1;
}
return (X[1], ..., X[p]).
```

This algorithm can be described by the following steps:

1. Sample p independent variables Z_i from the standard normal distribution, applying the function `GenerateNormStd`.
2. To obtain a new point $(Y_1, \ldots, Y_p)$, normalize the generated variables Z_i using the distance R. The obtained point is distributed according to the uniform distribution on the sphere S^p.
3. Shift the point $(Y_1, \ldots, Y_p)$ to the new position $(X_1, \ldots, X_p)$ using the transformed variable $U \sim U[0,1]$ (compare with transformation (4.31) and the distribution of the radius given by (4.33)). Then the output $(X_1, \ldots, X_p)$ is sampled from the uniform distribution on B^p.

4.7 Other approaches to multivariate generation

Apart from the general methods, mentioned in Section 4.5, there are more specific approaches which utilize the multivariate structure of the problem in a more direct way. The multivariate joint pdf considered is also a product of the relevant conditional densities

$$f_{\mathbb{X}}\left(x^{(1)}, x^{(2)}, \ldots, x^{(p)}\right) = f_1\left(x^{(1)}\right) f_2\left(x^{(2)} \left| x^{(1)}\right.\right) \ldots f_p\left(x^{(p)} \left| x^{(1)}, \ldots, x^{(p-1)}\right.\right) . \tag{4.37}$$

This formula leads to the following algorithm:

Algorithm 4.44 (Multivariate generation using the conditional densities)**.**

1. Generate $X^{(1)} \sim f_1\left(x^{(1)}\right)$.
2. For the fixed value of $x^{(1)}$, generate $X^{(2)} \sim f_2\left(x^{(2)} \left| x^{(1)}\right.\right)$.
3. ...

p. For the fixed values of $x^{(1)}, \ldots, x^{(p-1)}$, generate $X^{(p)} \sim f_p\left(x^{(p)} \left| x^{(1)}, \ldots, x^{(p-1)}\right.\right)$.

However, calculation of the formula (4.37) or sampling from the conditional distributions, necessary in this approach, may be numerically problematic in some practical cases.

Multivariate random variables can be also generated as some transformation of other random variables for which the relevant sampling algorithms are known. For example, if a multivariate random variable $\mathbb{Y}$ is given by the density $f_{\mathbb{Y}}(.)$ and the function $h : \mathbb{R}^p \to \mathbb{R}^p$ is of C^1 class and its inversion h^{-1} is also of C^1 class, then the variable

$$\mathbb{X} = h(\mathbb{Y})$$

has the density given by

$$g_{\mathbb{X}}(\mathbf{x}) = f_{\mathbb{Y}}\left(h^{-1}(\mathbf{x})\right)\left|\det\left(h^{-1}\right)'(\mathbf{x})\right| \ .$$

Another method is to apply simulations for some relevant *copula* (see also, e.g. (Schmidt 2006)):

Definition 4.45. *A d-dimensional copula $C : [0,1]^d \to [0,1]$ is a function, which is a cumulative distribution function with uniform marginals.*

Then the following useful theorem may be applied:

Theorem 4.46 (Sklar)**.** *Let $H(.,.)$ be a two-dimensional joint cdf and $F_x(.)$, $G_y(.)$ – the relevant marginal cumulative distribution functions. There exists a copula C such that*

$$H(x,y) = C\left(F_x(x), G_y(y)\right) \ .$$

If $F_x(.)$ and $G_y(.)$ are continuous, then C is unique.

As it may be seen from this theorem, the copula is a direct tool of modelling the dependence between the variables. Therefore, it may be also considered as the method to generate bivariate random variables with a fixed kind of dependency between them.

4.8 Multivariate normal distribution

If $\mathbb{X}$ is a multivariate random variable, then we have $\mathbb{E}\,\mathbb{X} = \left(\mathbb{E}\,X^{(1)}, \ldots, \mathbb{E}\,X^{(p)}\right)^T$, i.e. its expected value is a vector of the corresponding expected values for each random component, where T denotes transpose of a vector–column.

Definition 4.47. *Covariance matrix (or dispersion matrix or variance–covariance matrix) of the multivariate variable $\mathbb{X}$ is given by*

$$\mathrm{VAR}\,\mathbb{X} = \begin{pmatrix} \mathrm{Var}\,X^{(1)} & \mathrm{Cov}\left(X^{(1)}, X^{(2)}\right) & \ldots & \mathrm{Cov}\left(X^{(1)}, X^{(k)}\right) \\ \mathrm{Cov}\left(X^{(2)}, X^{(1)}\right) & \mathrm{Var}\,X^{(2)} & \ldots & \mathrm{Cov}\left(X^{(2)}, X^{(k)}\right) \\ \ldots & & & \\ \mathrm{Cov}\left(X^{(k)}, X^{(1)}\right) & \mathrm{Cov}\left(X^{(k)}, X^{(2)}\right) & \ldots & \mathrm{Var}\,X^{(k)} \end{pmatrix} \ ,$$

i.e. the element of this matrix on the i,j-th position is the covariance between the i-th and j-th components of the random vector $\mathbb{X}$.

Additionally, we have

$$\mathrm{VAR}\,\mathbb{X} = \mathbb{E}(\mathbb{X} - \mathbb{E}\,\mathbb{X})(\mathbb{X} - \mathbb{E}\,\mathbb{X})^T \ , \tag{4.38}$$

i.e. the formula (4.38) is a straightforward generalization of the one-dimensional expected value. It is known from the probability theory, that the covariance matrix is symmetric and positive definite.

As in the one-dimensional case, the multi-dimensional normal distribution is one of the most important distributions in practical applications. Therefore, sampling methods for such distribution are much needed. Let us start from the relevant definition.

Definition 4.48 (Multivariate normal distribution)**.** *A distribution of the random variable* $\mathbb{X}$ *is given by a multivariate normal distribution (further on we use notation* $\mathbb{X} \sim N(\mu, \mathbb{W})$*) if the density of this distribution is given by*

$$f_{\mathbb{X}}(t) = (2\pi)^{-p/2} \det\left(\mathbb{W}^{-1}\right) \exp\left[-\frac{1}{2}(t-\mu)^T \mathbb{W}^{-1}(t-\mu)\right] ,$$

where μ *is a* p*-dimensional vector and* $\mathbb{W}$ *is* $p \times p$*-dimensional covariance matrix. If* $\mathbb{X} \sim N(\mu, \mathbb{W})$*, then* $\mathbb{E}\,\mathbb{X} = \mu$ *,* $\operatorname{VAR}\mathbb{X} = \mathbb{W}$*.*

We start our discussion of the methods of sampling from the multivariate normal distribution from the independent, standard normal case (i.e. when $\mathbb{X} \sim N(0, \mathbb{I})$). Then, the more general case (i.e. if $\mathbb{X} \sim N(\mu, \mathbb{W})$) will be considered.

If $\mathbb{X} = \left(X^{(1)}, X^{(2)}, \ldots, X^{(p)}\right)$ and $\mathbb{X} \sim N(0, \mathbb{I})$, where $\mathbb{I}$ is the identity matrix (i.e. $\mathbb{I}$ is a $p \times p$-dimensional matrix with ones on the main diagonal and zeros elsewhere), then from the probability theory we know that

$$X^{(1)}, X^{(2)}, \ldots, X^{(p)} \overset{iid}{\sim} N(0,1) .$$

Therefore, in order to generate $\mathbb{X}$, we should sample p independent values from the standard normal distribution (see Section 4.3 for the discussion of the relevant methods). In the same way we generate the whole *iid* sequence $\mathbb{X}_1, \ldots, \mathbb{X}_n \sim N(0, \mathbb{I})$.

In a more general case, if $\mathbb{X} \sim N(\mu, \mathbb{W})$, then the covariance matrix may be factored uniquely (which is called *Cholesky decomposition*) as

$$\mathbb{W} = \mathbb{C}\mathbb{C}^T , \tag{4.39}$$

where $\mathbb{C}$ is lower triangular $p \times p$ matrix. Then, for $\mathbb{Y} \sim N(0, \mathbb{I})$ we should apply the transformation

$$\mathbb{X} = \mu + \mathbb{C}\mathbb{Y} , \tag{4.40}$$

which gives $\mathbb{X} \sim N(\mu, \mathbb{W})$ (compare with the formula (4.27) in the one-dimensional case).

There are many numerical algorithms to compute $\mathbb{C}$, defined by the condition (4.39) (see, e.g., (Fishman 1973)). The direct calculations based on the formulas

$$c_{ii} = \sqrt{w_{ii} - \sum_{k=1}^{i-1} c_{ik}^2} \tag{4.41}$$

$$c_{ji} = \frac{w_{ji} - \sum_{k=1}^{i-1} c_{jk} c_{ik}}{c_{ii}} , \tag{4.42}$$

where w_{ij} and c_{ij} are the respective elements of $\mathbb{W}$ and $\mathbb{C}$, can be also applied.

When the matrix $\mathbb{C}$ is calculated, then the following algorithm based on transformation (4.40) may be used:

Algorithm 4.49 (Multivariate normal distribution)**.**

1. Generate $Y_1, Y_2, \ldots, Y_p \overset{iid}{\sim} N(0,1)$.
2. For $i = 1, 2, \ldots, p$, let $X_i = \mu_i + \sum_{j=1}^{i} c_{ij} Y_j$.
3. Return $\mathbb{X} = (X_1, X_2, \ldots, X_p)$.

4.9 Generation of selected stochastic processes

We discuss here some methods for generation of trajectories of the most important stochastic processes, like homogeneous and non-homogeneous Poisson process, and Wiener process. The necessary definitions, concerning the theory of stochastic processes may be found in Chapter 2 or, e.g., in (Capasso & Bakstein 2012). Therefore, we refer the reader to these sources for the relevant details and notation.

In our considerations, we set $\mathcal{T} = [0, +\infty)$, so that the parameter t may be identified with time and the state space is some subset of $\mathbb{R}$. Then, the stochastic process may be seen as the evolution of a phenomenon, e.g. sequence of some events, which occur at corresponding time moments, like arrivals of buses at the bus stop or values of share prices for subsequent days or necessary repairs of the water distribution system (see, e.g., (Romaniuk 2015)).

4.9.1 The homogeneous Poisson process

We start from the necessary definitions, which describe the simplest case of the Poisson process, known as the homogeneous Poisson process (see also Section 2.3.2 for additional definitions).

The homogeneous Poisson process is characterized by its *rate* parameter (known also as the *intensity*) λ, which is equal to the expected number of events ("jumps" of trajectory of this process, see example given by Figure 4.6) that occur per unit time. In this part we use notation $N_t(\lambda)$ for this HPP.

The most straightforward method for sampling of the trajectory of homogeneous Poisson process is to use the relation between the exponential distribution $\mathrm{Exp}(\lambda)$ and the homogeneous Poisson process $N_t(\lambda)$. It is known from the probability theory that for a fixed $t \in T$, $N_t \sim \mathrm{Poiss}(\lambda t)$ and the intervals between consecutive events (between the previously mentioned "jumps" of the trajectory) are exponentially distributed with the parameter λ (see Theorem 2.39).

Therefore, for a fixed time t, we have the following algorithm (see, e.g. (Law 2007)):

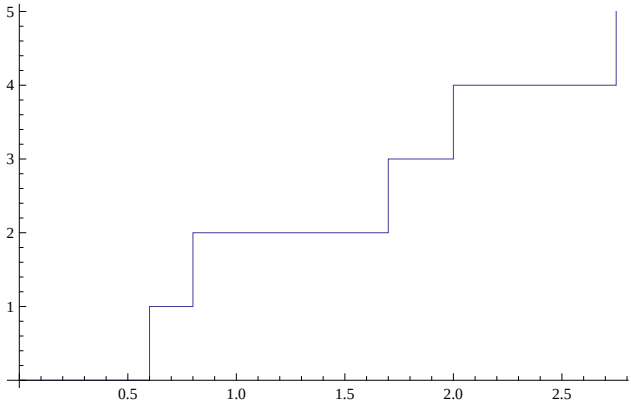

Fig. 4.6. Example of Poisson process trajectory

Algorithm 4.50 (Homogeneous Poisson process I)**.**

```
N = 0;
S = 0;
while S < t do
{
 L = GenerateExp (lambda);
 S = S + L;
 N = N + 1;
}
return N.
```

The above algorithm can be described as follows:

1. At the beginning of the procedure, the value of the process N and the considered time moment S are both equal to 0.
2. Generate the length of the interval $L \sim \text{Exp}(\lambda)$ by applying the function `GenerateExp (lambda)`.
3. Increase the value of the process (number of "jumps" of its trajectory) N by one and add the random variable L to the considered time moment S.
4. If $S < t$, return to step 2, otherwise, return the final output N.

As it can be easily seen, a sequence $L_1, L_2, \ldots$ of *iid* exponentially distributed random variables is generated via this approach. They may be identified with the intervals between the subsequent events, caused by the process N_t. Then, the cumulated sums of lengths of these intervals

$$S_1 = L_1, S_2 = L_1 + L_2, S_3 = L_1 + L_2 + L_3, \ldots$$

are calculated. These sums are also the moments of the subsequent jumps (or *arrival times*) of the process N_t. This algorithm stops, when the value of the last cumulated sum S_k is greater than the fixed time t.

This approach, however, has some important disadvantage – it is numerically inefficient for greater values of λ (i.e. shorter intervals between the events) or of t. In such a case the following theorem can be useful (see, e.g. (Karlin & Taylor 1975)):

Theorem 4.51. *For the fixed value of the process $N_t = n$, the conditional distribution of the vector of the cumulated sums of the intervals $(S_1, S_2, \ldots, S_n)$ is the same as the distribution of vector of order statistics $(X_{1:n}, X_{2:n}, \ldots, X_{n:n})$ obtained for $X_1, \ldots, X_n \overset{iid}{\sim} U[0,t]$.*

This theorem leads directly to the respective algorithm:

Algorithm 4.52 (Homogeneous Poisson process II)**.**

```
n = GeneratePoiss (lambda * t);
for i = 1 to n do
{
 U_i = GenerateU;
}
(X_{1:n}, X_{2:n}, ..., X_{n:n})
 = OrderStatistics (U_1, U_2, ..., U_n);
for i = 1 to n do
{
 S_i = t * X_{i:n};
}
return (S_1, S_2, ..., S_n).
```

The working of this algorithm can by subsumed by the following steps:

1. Generate the value $n \sim \text{Poiss}(\lambda * t)$ of the process $N_t(\lambda)$ for the fixed time t, applying the function `GeneratePoiss (lambda * t)` (as indicated in Definition 2.38).
2. Sample n *iid* random variables $U_1, \ldots, U_n$ from $U[0,1]$.
3. Calculate the order statistics $(X_{1:n}, X_{2:n}, \ldots, X_{n:n})$ based on $U_1, \ldots, U_n$.
4. Multiply these order statistics by t, obtaining the vector of the cumulated sums of the lengths of the intervals $(S_1, S_2, \ldots, S_n)$.

4.9.2 The non-homogeneous Poisson process

In the case of non-homogeneous Poisson process, the rate parameter λ is not a constant value. Instead, we assume that $\lambda(t) \geq 0$ for all $t \geq 0$. As discussed in Section 2.3.2, this function is called the *rate function* (or the *intensity function*).

Non-homogeneous Poisson processes model many real-world systems. For example, road traffic in a central part of the city is heavier during the morning and evening rush hours. Also, the number of arrivals of buses at the bus stop depends on hour.

We discuss two methods of non-homogeneous Poisson process generation, namely the thinning method and the inversion of cumulative rate function.

The *thinning method* was proposed by Lewis & Shedler (1979) and, in some way, is similar to the rejection method (see Section 4.1.2). Let us suppose that

$$\lambda^* = \max_t \{\lambda(t)\} < \infty \ .$$

Then we have the following algorithm:

Algorithm 4.53 (Thinning method I)**.**

1. Generate the cumulative sums $(S_1, S_2, \ldots)$ for the homogeneous Poisson process $N_t(\lambda^*)$.
2. For each S_i, $i = 1, 2, \ldots$, accept this value as a new arrival time $S_i^{'}$ with the probability given by $\lambda(S_i)/\lambda^*$.
3. Return the obtained sequence $\left(S_1^{'}, S_2^{'}, \ldots\right)$.

In this method, after sampling from teh homogeneous Poisson process, some of the obtained arrival times S_i are rejected (with probability $1 - \lambda(S_i)/\lambda^*$) or accepted (with the remaining probability $\lambda(S_i)/\lambda^*$). Therefore, the mentioned similarity to the rejection method is easily seen.

Algorithm 4.53 can be also described in another, more convenient recursive form, for some fixed value T (see, e.g. (Law 2007)):

Algorithm 4.54 (Thinning method II)**.**

1. Set $S = 0$, $N = 0$, $i = 0$.
2. Generate $U \sim U[0, 1]$, then calculate $S = S + (-1/\lambda^*) \ln U$.
3. Generate $U \sim U[0, 1]$. If $U \leq \lambda(S)/\lambda^*$, then accept this value and set $N = N + 1$, $i = i + 1$, $S_i^{'} = S$.
4. If $S > T$, then stop the algorithm and return the obtained sequence $\left(S_1^{'}, S_2^{'}, \ldots\right)$, otherwise return to the step 2.

The thinning method is simple, but it may be numerically inefficient in some cases. If the rate function $\lambda(t)$ has relatively low values, except for a few high and narrow peaks, λ^* is not comparable with $\lambda(t)$ most of the time. In such a case, most of the points generated for the homogeneous Poisson process $N_t(\lambda^*)$

will be rejected. Then, a more general thinning algorithm with nonconstant λ^* may be applied (see, e.g. (Lewis & Shedler 1979) for further details).

The other approach, the *inversion of cumulative rate function*, is similar to the inversion method (see Section 4.1.1). Denote by

$$\Lambda(t) = \Lambda(0,t) = \int_0^t \lambda(y)dy \tag{4.43}$$

the expectation function (or the cumulative rate function). This function is a continuous function of t and it is equal to the expected number of events between time instants 0 and t. Let $\Lambda^{-1}(t)$ be the inverse function of $\Lambda(t)$. Then, we have the following algorithm (see, e.g. (Law 2007)):

Algorithm 4.55 (Inversion of cumulative rate function)**.**

1. Generate the arrival times $(S_1, S_2, \ldots)$ for the homogeneous Poisson process $N_t(1)$, i.e. for $\lambda = 1$.
2. For each S_i, $i = 1, 2, \ldots$, apply $S_i^{'} = \Lambda^{-1}(S_i)$.
3. Return the obtained sequence $\left(S_1^{'}, S_2^{'}, \ldots\right)$.

Contrary to the thinning method, all of the S_i's are used in Algorithm 4.55. But this approach requires calculation and then inversion of $\Lambda(t)$, which can be numerically difficult in some cases.

4.9.3 The Wiener process

Generation of trajectory of the Wiener process is based on its features, described in Section 2.3.1. Let us assume that we are interested only in values of this process for some fixed moments $t_1 < t_2 < \ldots < t_n$, i.e. $W_{t_1}, W_{t_2}, \ldots, W_{t_n}$. Then, the following algorithm can be applied:

Algorithm 4.56 (Wiener process)**.**

```
W_0 = 0;
t_0 = 0;
for i = 1 to n do
{
 N = GenerateNorm (0, t_i - t_{i-1});
 W_{t_i} = W_{t_{i-1}} + N;
};
return (W_{t_1},...,W_{t_n}).
```

This algorithm is described by the following steps:

1. Set $W_0 = 0$ and $t_0 = 0$.
2. For the successive moments $t_1, t_2, \ldots, t_n$, generate $N_1, N_2, \ldots, N_n \overset{iid}{\sim} N(0, t_i - t_{i-1})$ applying the function `GenerateNorm (0, t_i - t_{i-1})`.
3. Calculate the values $W_{t_1}, W_{t_2}, \ldots, W_{t_n}$, based on the successive increases $N_1, N_2, \ldots, N_n$, where

$$W_{t_1} = N_1, W_{t_2} = W_{t_1} + N_2, \ldots, W_{t_n} = W_{t_{n-1}} + N_n \ .$$

5

Foundations of Monte Carlo methods

Maciej Romaniuk

5.1 Historical context

Computational techniques, known as *Monte Carlo* (abbreviated usually as MC) methods, were proposed by Metropolis & Ulam (1949). The name of this approach was suggested by Nicolas Metropolis because of the similarity of MC methods to games of chance, and Monaco is a famous centre of gambling. In their beginning, MC simulations were used during the famous Manhattan project. MC method was applied as a numerical technique for the approximation of integrals which could not be solved by other approaches. And integrals over poorly-behaved functions, as well as multidimensional integrals have been and still are convenient subjects of the MC method. Nowadays, because of advances in computer hardware, especially because of lower costs of computer memory and faster processors, MC methods can be found in various areas of applications, namely integration and optimization problems, generation of random samples from complicated distributions, actuarial sciences, financial mathematics, Bayesian inference etc. Some of these areas are discussed in a more detailed way in Chapters 7 and 8.

Stanisław Ulam described the moment of invention of MC methods (see (Eckhardt 1987)) in this way:

> *The first thoughts and attempts I made to practice [the Monte Carlo Method] were suggested by a question which occurred to me in 1946 as I was convalescing from an illness and playing solitaires. The question was what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays. This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics, and more generally how to change processes described by certain differential equations into an equivalent form interpretable as a succession of random*

operations. Later [in 1946, I] described the idea to John von Neumann, and we began to plan actual calculations.

As it can be seen from the above description, the main idea of Ulam's was to apply random calculations based on statistics and probability theory to approximation of solutions of real-life problems. Earlier, this idea was thought to be rather some statistical "curiosity", not the useful, practical tool. This seems obvious now for modern scientists, but it is a very subtle idea that some entirely physical problem could be approximately solved by a related random process. Nowadays, the name of Monte Carlo is used as a general term for various simulation techniques.

In this chapter, the main ideas of Monte Carlo approach to integration and optimization problems, will be considered. Additional details about MC methods may be found in, e.g. (Robert & Casella 2004), (Law 2007).

5.2 Monte Carlo integration

It is commonly known that many integrals are very difficult to calculate, because of problems with lack of the relevant analytical form of antiderivative. Of course, there are many numerical, strictly deterministic (i.e. without necessity of generation of any random numbers) approaches, like the trapezoidal rule or Simpson's rule (see, e.g. (Atkinson 1989)). But in the case of some integrals, especially multidimensional integrals, these approaches are computationally very expensive and thus less effective than the MC method, based on simulation approach. This advantage of Monte Carlo integration will be explained in a more detailed way in Section 5.2.3.

5.2.1 The simple approach

The simplest case of the MC approach can be considered when the definite integral

$$\int_a^b h^*(x)dx \tag{5.1}$$

for some fixed interval $[a, b]$ is evaluated. This kind of problems is illustrated by a few following examples.

Example 5.1. *As it is known, cdf of the standard normal distribution can not be written in an explicit, analytical form. Therefore, in order to calculate the relevant probabilities, simulations can be directly applied. Let*

$$\Phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}} dx \ . \tag{5.2}$$

Then, for $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(0,1)$*, the integral* (5.2) *can be approximated by the simple estimator*

$$\hat{\Phi}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq y) \ .$$

The correctness of such estimation will be proved in Section 5.2.3. In order to generate the sample $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(0,1)$ *the relevant algorithm for the standard normal distribution should be used. The examples of such algorithms were discussed in Section 4.3.*

Example 5.2. *Let us suppose that we would like to calculate*

$$\int_0^1 \left(\sin(20x) + \cos(10x)\right)^2 dx \ . \tag{5.3}$$

This integral is rather complicated for direct, numerical evaluation, because of high fluctuations of the integrated function, as illustrated by Figure 5.1. Therefore, for the sample $U_1, U_2, \ldots, U_n \overset{iid}{\sim} U[0,1]$*, the integral* (5.3) *can be approximated by*

$$\hat{h^*}(X) = \frac{\sum_{i=1}^{n} \left(\sin(20U_i) + \cos(10U_i)\right)^2}{n} \ .$$

The example of the trajectory of such an estimator, based on various numbers of simulations n *is illustrated with Figure 5.2.*

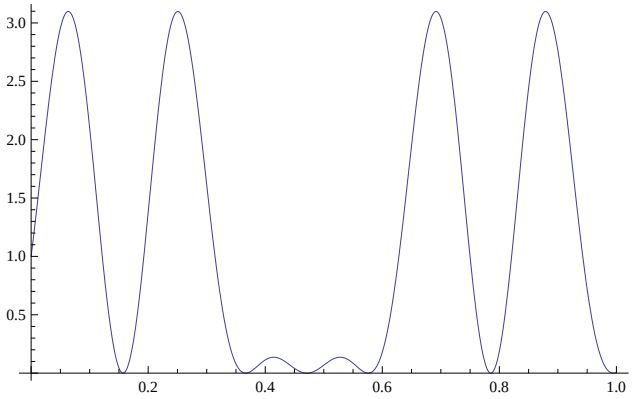

Fig. 5.1. Graph of the function $h^*(x) = \left(\sin(20x) + \cos(10x)\right)^2$ from Example 5.2

5.2.2 Hit-or-miss method

In the so-called *hit-or-miss Monte Carlo integration* (which is also described as *geometrical Monte Carlo*), the definite integral of the form (5.1) is estimated by drawing a rectangle $\mathcal{A}$, which contains the graph of the function $h^*(x)$, i.e.

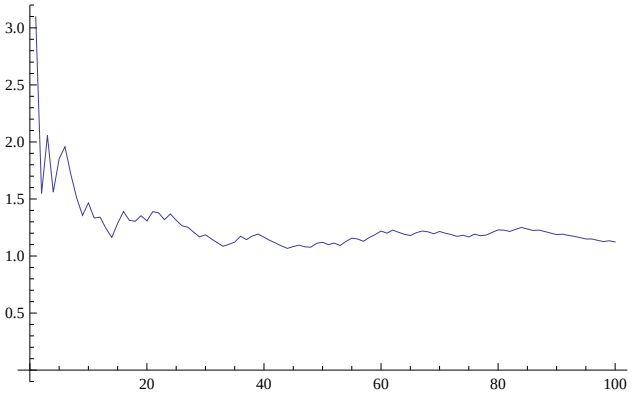

Fig. 5.2. Graph of the estimator $\hat{h^*}(X)$ from Example 5.2

$$\mathcal{A} = \{(x, y) : x \in [a, b], 0 \leq y \leq \max_{t \in [a,b]} h^*(t)\} \ .$$

Then, n points are sampled uniformly over the rectangle $\mathcal{A}$, so the integral (5.1) is approximated via the estimator

$$\hat{h}^*(X) = \frac{n_0}{n} l_2(\mathcal{A}) \ ,$$

where n_0 is the number of points which fall under the graph of the function $h^*(x)$ and $l_2(\mathcal{A})$ is the area of $\mathcal{A}$.

Example 5.3. *In order to approximate the value of π, we uniformly generate n points from the square $C^2 = [-1, 1] \times [-1, 1]$. Then, the number of points, which are inside the 2-dimensional unit ball $B^2 = \{(x, y) : x^2 + y^2 \leq 1\}$, is calculated (as illustrated by Figure 5.3). Because the ratio of B^2 to C^2 is equal to*

$$\frac{l_2(B^2)}{l_2(C^2)} = \frac{\pi}{4} \ ,$$

which could be approximated by $\frac{n_0}{n}$, then the relevant estimator of π is given by

$$\hat{\pi} = \frac{4n_0}{n} \ .$$

5.2.3 Crude Monte Carlo

Instead of the simple formula (5.1), the more general form of the definite integral

$$\int_{\mathcal{X}} h^*(x) dx \tag{5.4}$$

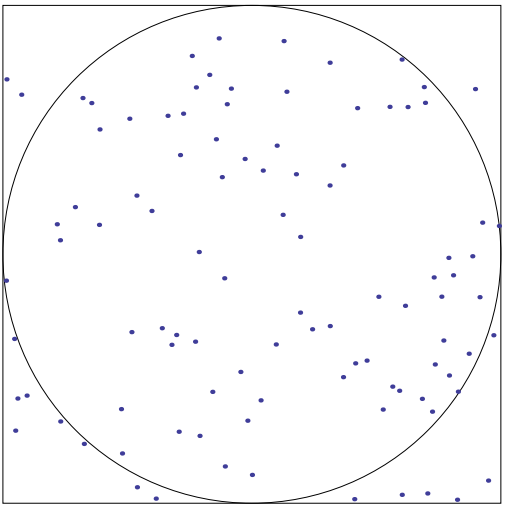

Fig. 5.3. Application of the hit-or-miss method from Example 5.3

may be considered for some fixed set $\mathcal{X}$. Then for the *crude* Monte Carlo method (also known as *sample-mean* method), the integral (5.4) is written in other form, where two specially selected functions $h(x)$ and $f(x)$ are used, i.e.

$$\int_{\mathcal{X}} h^*(x)dx = \int_{\mathcal{X}} h(x)f(x)dx \ . \tag{5.5}$$

If $f(x)$ is pdf of some random distribution with the support $\mathcal{X}$, then this integral is also the expected value of $h(x)$ with the density $f(x)$, i.e.

$$\mathbb{E}_f h(X) = \int_{\mathcal{X}} h(x)f(x)dx \ . \tag{5.6}$$

Because of practical reasons, we assume that $\mathcal{X} \subset \mathbb{R}^p$ for some fixed value p. Additionally, the value of the integral (5.6) should exist in numerical sense and has to be finite, i.e. $\mathbb{E}_f|h(x)| < \infty$.

The derived formula (5.6) is a very general form for the integration problem. For example, if $\mathcal{X}$ is a compact set, then we have

$$\int_{\mathcal{X}} h(x)dx = \int_{\mathcal{X}} l_p(\mathcal{X})h(x)\frac{1}{l_p(\mathcal{X})}dx = l_p(\mathcal{X})\mathbb{E}_{l_p(\mathcal{X})}h(X) \ , \tag{5.7}$$

where $l_p(\mathcal{X})$ is the p-dimensional volume of $\mathcal{X}$. In this case, the density $f(x)$ is given by the uniform distribution over the set $\mathcal{X}$.

In order to calculate the expected value (5.6) via the crude Monte Carlo method, the sample $X_1, X_2, \ldots, X_n$ of *iid* variables according to pdf $f(x)$ should be generated. Then, the empirical average

$$\hat{h}_f(X) = \frac{1}{n}\sum_{i=1}^{n} h(X_i) \tag{5.8}$$

approximates the value of the considered integral (5.6), since $\hat{h}_f(X)$ converges almost surely to $\mathbb{E}_f h(X)$ by the Strong Law of Large Numbers (abbreviated usually as SLLN, see, e.g. (Feller 1968)):

Theorem 5.4 (Strong Law of Large Numbers)**.** *Let* $Y_1, Y_2, \ldots, Y_n$ *be* iid *random variables with* $E|Y_i| < \infty$ *and* $EY_i = \mu$*. Then*

$$\frac{\sum_{i=1}^{n} Y_i}{n} \xrightarrow[n\to\infty]{a.s.} \mu \ .$$

Moreover, if $\mathbb{E}_f h^2(X) < \infty$ (i.e. h^2 has a finite expectation under f), then we may assess the speed of convergence of the estimator $\hat{h}_f(X)$. The variance σ_h^2 of the function $h(X)$ for the density f is approximated by

$$\hat{\sigma}_h^2 = \hat{\mathrm{Var}}\left(h_f(X)\right) = \frac{1}{n}\sum_{i=1}^{n}\Big(h(X_i) - \hat{h}_f(X)\Big)^2 \ .$$

And variance of the estimator $\hat{h}_f(X)$, which is given by

$$\sigma_{\hat{h}}^2 = \mathrm{Var}\left(\hat{h}_f(X)\right) = \frac{1}{n}\int_{\mathcal{X}} \left(h(x) - \mathbb{E}_f h(X)\right)^2 f(x)dx \ ,$$

can be estimated from the sequence $X_1, X_2, \ldots, X_n$ via the sample variance

$$\hat{\sigma}_{\hat{h}}^2 = \hat{\mathrm{Var}}(\hat{h}_f(X)) = \frac{1}{n^2}\sum_{i=1}^{n}(h(X_i) - \hat{h}_f(X))^2 \ .$$

This leads to the equality

$$\hat{\sigma}_{\hat{h}}^2 = \frac{\hat{\sigma}_h^2}{n} \ . \tag{5.9}$$

For large n, from the Central Limit Theorem,

$$\frac{\hat{h}_f(X) - \mathbb{E}_f h(X)}{\hat{\sigma}_{\hat{h}}}$$

is approximately distributed as the $N(0,1)$ variable. Therefore, the relevant convergence tests and confidence bounds for the approximation of $\mathbb{E}_f h(X)$ can be constructed.

From the equality (5.9), the statistical error of the MC methods, measured by the standard deviation $\sigma_{\hat{h}}$, is roughly proportional to $1/\sqrt{n} = n^{-1/2}$ for large n, regardless of the dimension of the evaluated integral. Therefore, teh MC methods become particularly attractive if we consider integration in higher dimensions, when numerical, deterministic approaches have higher errors. For example, in the

case of the trapezoid rule, its error decreases as $n^{-2/d}$, where d is the number of dimensions and n denotes the number of subintervals per dimension. Generally, in "standard", i.e. deterministic approaches, their error is proportional to $n^{-a/d}$ for some constant a depending on the algorithm considered. As it can be easily seen, the MC method for some sufficiently high number of dimensions d has lower error than any deterministic method.

If the assumption that $\mathbb{E}_f h^2(X) < \infty$ is not fulfilled, then some problems with numerical stability of the estimator $\hat{h}_f(X)$ can arise. For example, unexpected fluctuations of values of the estimators calculated for various runs of simulations are observed.

Our considerations, concerning the crude MC approach are illustrated by the following examples:

Example 5.5. *Let us consider the integral*

$$I = \int_0^\infty (1 + \sin(x))\, e^{-\frac{x^2}{2}} dx \ .$$

It leads to

$$I = \int_0^\infty \sqrt{2\pi}\,(1 + \sin(x))\, \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \ .$$

Then, $h(x) = \sqrt{2\pi}\,(1 + \sin(x))$, and $f(x)$ is pdf of the standard normal distribution. Therefore, for $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(0,1)$, the crude MC estimator of the integral I is given by

$$\hat{h}_f(X) = \sqrt{2\pi}\frac{1}{n}\sum_{i=1}^{n} (1 + \sin(X_i))\, \mathbb{1}(X_i \geq 0) \ . \tag{5.10}$$

Example 5.6. *Let us consider the two-dimensional integral*

$$I = \int_0^\infty \int_0^\infty xy^2 e^{-2x-y} dx\, dy \ .$$

Quite obviously,

$$I = \int_0^\infty \int_0^\infty \frac{1}{2} xy^2 2e^{-2x} e^{-y} dx\, dy \ ,$$

therefore $h(x,y) = \frac{1}{2}xy^2$ and the joint two-dimensional pdf $f(x,y) = f(x)f(y) = 2e^{-2x}e^{-y}$ is the product of two densities $f(x)$ and $f(y)$ of the independent random variables $X \sim \text{Exp}(2)$ and $Y \sim \text{Exp}(1)$. For $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Exp}(2)$ and $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \text{Exp}(1)$, the relevant estimator of I is given by

$$\hat{h}_f(X,Y) = \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} X_i Y_i^2 \ .$$

5.2.4 Riemann approximation

One of the deterministic methods to evaluate one-dimensional integrals of the form (5.1) is to apply the widely known analytical definition of the integral, where the integral is given by the limit of Riemann sums. For every sequence $a_{i,n}$ (where $0 \leq i \leq n$, $a_{0,n} = a$, $a_{n,n} = b$, $a_{i,n} \leq a_{i+1,n}$) the Riemann sum

$$\sum_{i=0}^{n-1} h(a_{i,n}) f(a_{i,n}) \left(a_{i+1,n} - a_{i,n}\right)$$

converges to

$$\int_a^b h(x) f(x) dx \tag{5.11}$$

as $n \to \infty$ if $a_{i+1,n} - a_{i,n}$ goes to 0 (in n).

This approach could be combined with Monte Carlo simulations, which leads to Riemann sum with random steps. Such method is known as *weighted Monte Carlo integration* or *simulation by Riemann sums (Riemannian simulation)* (see (Philippe 1997)). In this case the one-dimensional integral (5.11) is approximated by the estimator

$$\hat{h}_f^R(X) = \sum_{i=0}^{n-1} h\left(X_{i:n}\right) f\left(X_{i:n}\right) \left(X_{i+1:n} - X_{i:n}\right) \ , \tag{5.12}$$

where $X_0, X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ and $X_{0:n}, X_{1:n}, \ldots, X_{n:n}$ are order statistics of such a sample.

This method has practical significance, because the variance of the estimator $\hat{h}_f^R(X)$ decreases proportionally to n^{-2} compared to only n^{-1} for the crude MC approach. However, this dominance fails to extend to the case of multidimensional integrals because of the *curse of dimensionality*, considered in Section 4.6.1. Additionally, the simulation by Riemann sums requires storing and ordering of all values of the sample $X_0, X_1, \ldots, X_n$, which increases the numerical burden of this algorithm.

5.3 Variance reduction methods

As emphasized by the equality (5.9), the statistical error of MC methods depends on the number of simulations (changing as $n^{-1/2}$ for the standard deviation) and on the variance σ_h^2 itself. The lower the value of σ_h^2, the lower the number of simulations necessary to assess the fixed level of this error. The described dependency can be illustrated using the following, classical example (see, e.g. (Robert & Casella 2004)):

Example 5.7. *Let us assume that we would like to calculate the integral*

$$I = \int_2^\infty \frac{1}{\pi(1+x^2)} dx \ . \tag{5.13}$$

The general form of pdf for the Cauchy distribution (denoted further by $C(\mu, \gamma)$) is given by

$$f(t) = \frac{1}{\pi\gamma(1+\left(\frac{t-\mu}{\gamma}\right)^2)}$$

for the scale parameter $\gamma > 0$ and the location parameter $\mu \in \mathbb{R}$. Then, the integral (5.13) *is similar to the density of the standard Cauchy (i.e. $C(0,1)$) distribution*

$$f(t) = \frac{1}{\pi(1+t^2)} \ .$$

Therefore, if the algorithm for sampling from $C(0,1)$ is known (it is available e.g. using the inversion method), then the relevant crude MC estimator is equal to

$$\hat{I}_1 = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left(X_i > 2\right) \ , \tag{5.14}$$

where $X_1, X_2, \ldots, X_n \overset{iid}{\sim} C(0,1)$. In this case we have, from the binomial distribution,

$$\operatorname{Var} \hat{I}_1 = \frac{I(1-I)}{n} \approx \frac{0,127}{n} \ . \tag{5.15}$$

Because the density of the Cauchy distribution is symmetrical around zero, then the relevant MC estimator of the integral I is also given by

$$\hat{I}_2 = \frac{1}{2n} \sum_{i=1}^{n} \mathbb{1}\left(|X_i| > 2\right) \ .$$

In this case the error of estimation is lower, because

$$\operatorname{Var} \hat{I}_2 = \frac{I(1-2I)}{2n} \approx \frac{0,052}{n} \ .$$

This result may be further improved – the inefficiency of the previously described approaches is due to the generation of values outside the domain of interest, i.e. only from the tails of the Cauchy distribution. But the "main mass of probability" of this distribution is located around zero. Knowing this, the integral (5.13) *may be written as*

$$I = \frac{1}{2} - \int_0^2 \frac{1}{\pi(1+x^2)} dx = \frac{1}{2} - \int_0^2 \frac{2}{\pi(1+x^2)} \frac{1}{2} dx \ . \tag{5.16}$$

And an alternative method of evaluation of I, based on the formula (5.16)*, is therefore*

$$\hat{I}_3 = \frac{1}{2} - \frac{1}{n}\sum_{i=1}^{n} \frac{2}{\pi(1+U_i^2)} \ ,$$

where $U_1, U_2, \ldots, U_n \overset{iid}{\sim} U[0,2]$. *From integration by parts, we obtain the approximation of variance for this estimator*

$$\mathrm{Var}\, \hat{I}_3 \approx \frac{0,0092}{n} \ .$$

Applying $y = x^{-1}$, *we can write down the considered integral* (5.13) *also as*

$$I = \int_0^{\frac{1}{2}} \frac{1}{\pi(1+\frac{1}{y^2})y^2} dy = \int_0^{\frac{1}{2}} \frac{1}{2\pi(1+y^2)} 2dy \ ,$$

which gives a completely different estimator

$$\hat{I}_4 = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2\pi(1+U_i^2)} \ , \tag{5.17}$$

where $U_1, U_2, \ldots, U_n \overset{iid}{\sim} U[0,\frac{1}{2}]$. *In this case, the same integration by parts shows that*

$$\mathrm{Var}\, \hat{I}_4 \approx \frac{0,00095}{n} \ . \tag{5.18}$$

Comparing the variance of the first estimator, given by (5.15) *with the variance of the last approximation, which is equal to* (5.18), *it can be seen that the reduction in variance is of the order of 1000. This means that the evaluation of the last estimator* $\hat{I}_4$ *requires* $\sqrt{1000} \approx 33$ *times fewer simulations to achieve the same precision as in the case of the first approach* $\hat{I}_1$. *However, to the contrary, more knowledge about the properties of the integral and additional "contemplation" time are also necessary.*

As illustrated by the previous example, it is possible to lower the error of the MC estimator if its variance is reduced. Therefore, special methods of variance reduction (such as importance sampling, antithetic variables, control variates) were developed. Some examples of these approaches and the relevant algorithms are discussed in the sequel.

5.3.1 Importance sampling

The evaluated integral (5.6) has an alternative representation (called *importance sampling fundamental identity*), namely

$$\mathbb{E}_f h(X) = \int_{\mathcal{X}} h(x)\frac{f(x)}{g(x)} g(x)dx \ ,$$

which leads to another formula for the estimator

$$\hat{h}_g(X) = \frac{1}{n}\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)} h(X_i) \ , \tag{5.19}$$

where $X_1, \ldots, X_n$ are *iid* samples from the new density $g(x)$.

The estimator (5.19) converges to $\mathbb{E}_f h(X)$ for the same reason as the crude MC estimator (5.8) does, with the additional assumption that the support of $g(x)$ includes the support of $f(x)$.

This method puts very little restrictions on the choice of the so called *instrumental density* $g(x)$. But some densities give rise to better results than others. The variance of the importance sampling estimator is finite only if

$$\mathbb{E}_g \left(h^2(X) \frac{f^2(X)}{g^2(X)} \right) = \int_{\mathcal{X}} h^2(x) \frac{f^2(x)}{g^2(x)} g(x) dx = \int_{\mathcal{X}} h^2(x) \frac{f^2(x)}{g(x)} dx < \infty \ .$$

It means that the instrumental density $g(x)$ with its tails lighter than those of $f(x)$ (i.e. with unbounded ratio f/g) is not an appropriate choice in the case of this method. In such an instance, the value of the importance sampling estimator (5.19) can vary from one iteration of simulations to the next one.

It is possible, though, to find "the best" instrumental density $g(x)$, which minimizes the variance of the estimator (5.19) and which is the optimal choice for the given function $h(x)$ and the fixed density $f(x)$:

Theorem 5.8. *The variance of the importance sampling estimator* (5.19) *is minimal for the density* $g(x)$ *given by the formula*

$$g^*(x) = \frac{|h(x)| f(x)}{\int_{\mathcal{X}} |h(z)| f(z) dz} \ . \tag{5.20}$$

Proof of this theorem can be found in (Robert & Casella 2004).

This theorem provides for the *constructive approach* in the sense that it gives the exact formula for the optimal density $g^*(x)$. But this optimal choice requires some additional knowledge of the considered integral – there is the factor $\int h(x) f(x) dx$ in the denominator of (5.20). Therefore, from the practical point of view, we should choose the instrumental density $g(x) \propto |h(x)| f(x)$, i.e. the one, for which $|h(x)| f(x)/g(x)$ is almost constant with finite variance.

Definition 5.9 (Student's t distribution)**.** *A distribution of the random variable* X *is given by Student's t distribution (further on we use the notation* $X \sim t(\nu)$*) if the density of this distribution is equal to*

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\ \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

for $\nu > 0$ *degrees of freedom. If* $X \sim t(\nu)$ *and* $\nu > 1$*, then* $\mathbb{E}\, X = 0$ *, if* $\nu > 2$*, then* $\operatorname{Var} X = \frac{\nu}{\nu-2}$.

Example 5.10. *Let us consider the integral* $\mathbb{E}_f h(X)$*, where* $f(x)$ *is equal to pdf of Student's t distribution with* ν *degrees of freedom and*

$$h(x) = \frac{x^5}{1+(x-3)^2}\mathbb{1}\left(x \geq 0\right) \ .$$

In this case there are many possible instrumental densities (e.g. normal distribution, Cauchy distribution – see (Robert & Casella 2004) for a more detailed discussion). An important candidate for the instrumental density $g(x)$ *is the exponential distribution. Then,* $g(x) = e^{-x}$ *for* $x \geq 0$ *and the considered integral can be written as*

$$\mathbb{E}_f h(X) = \int_0^\infty \frac{x^5}{1+(x-3)^2} f(x)dx = \int_0^\infty \frac{x^5 e^x}{1+(x-3)^2} f(x)g(x)dx \ ,$$

which leads to the importance sampling estimator

$$\hat{h}_g(X) = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i^5 e^{X_i}}{1+(X_i-3)^2} f(X_i) \ ,$$

where $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Exp}(1)$.

5.3.2 Antithetic variables

Usually, during simulations, *iid* samples are used to calculate the estimator of the considered integral $\mathbb{E}_f h(X)$. But there may be situations, when it may be preferable to generate correlated variables instead of the independent ones. For example, if we wish to compare two quantities

$$I_1 = \int g_1(x) f_1(x)dx \ , \ I_2 = \int g_2(x) f_2(x)dx \ ,$$

which are close in value, then the estimator $\hat{I}_1 - \hat{I}_2$ has variance $\text{Var}\left(\hat{I}_1\right) + \text{Var}\left(\hat{I}_2\right)$, and so its error can be too high for the appropriate estimation of the difference $I_1 - I_2$. But if both of the estimators, $\hat{I}_1$ and $\hat{I}_2$, are positively correlated, then the variance of $\hat{I}_1 - \hat{I}_2$ is reduced by the factor of $-2\,\text{Cov}\left(\hat{I}_1, \hat{I}_2\right)$.

The *antithetic variables* method is based on a similar idea. In this case, instead of the classical, crude MC estimator (5.8), another approximation,

$$\hat{h}_f(X) = \frac{1}{2n}\sum_{i=1}^{n} \left(h(X_i) + h(Y_i)\right) \ , \tag{5.21}$$

where pairs $(X_1, Y_1), \ldots (X_n, Y_n)$ are *iid* variables, is applied. Both of the random sequences, $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, are sampled from the density $f(x)$. The

antithetic estimator (5.21) is more efficient than the estimator (5.8), based on an *iid* sample of size $2n$, if the variables $h(X_i)$ and $h(Y_i)$ for each pair (X_i, Y_i) are negatively correlated. Then, the variables Y_i are called *antithetic* variables.

It should be noted that the correlation between $h(X_i)$ and $h(Y_i)$ depends both on the pair (X_i, Y_i) and the function $h(.)$ itself. Therefore, there is no general method for the construction of the relevant *dependent* sampling algorithm. Rubinstein (1981) proposed to apply the uniform variables U_i to generate the sequence of X_i and then the transformed variables $1 - U_i$ to generate sequence of Y_i. As it was mentioned in Section 4.1, the uniform variables are the common source of input for the more complex algorithms for sampling from other distributions. But the approach mentioned requires additional assumptions about the function $h(.)$ and special methods of sampling from the uniform random variables.

5.3.3 Control variates

In the case of the *control variates* method, the additional knowledge about some integral of the function $h_0(x)$ for the density $f(x)$ is necessary. Let us assume that two estimators are available, the first one approximates the considered integral

$$\mathbb{E}_f h(X) = \int_{\mathcal{X}} h(x) f(x) dx \ ,$$

and is given by the classical, crude MC formula

$$\hat{h}_f(X) = \frac{1}{n} \sum_{i=1}^{n} h(X_i) \ ,$$

the second estimator approximates another integral

$$\mathbb{E}_f h_0(X) = \int_{\mathcal{X}} h_0(x) f(x) dx$$

and is also given by the crude MC formula

$$\hat{h_0}_f(X) = \frac{1}{n} \sum_{i=1}^{n} h_0(X_i) \ .$$

Because

$$\mathbb{E}_f h(X) = (\mathbb{E}_f h(X) + c\mathbb{E}_f h_0(X)) - c\mathbb{E}_f h_0(X) \ ,$$

for some constant c, then this integral can be also estimated by another formula, which "combines" both $\hat{h}_f(X)$ and $\hat{h_0}_f(X)$, i.e.

$$\hat{h}_f(X)^* = \frac{1}{n} \sum_{i=1}^{n} (h(X_i) + ch_0(X_i)) - c\mathbb{E}_f h_0(X) \ . \tag{5.22}$$

As it is emphasized by the form of the estimator (5.22), in order to use this new approximation $\hat{h}_f(X)^*$, it is necessary to know the exact value of $\mathbb{E}_f h_0(X)$.

The variance of the obtained approximation, known as the *control variate estimator* $\hat{h}_f(X)^*$, is equal to

$$\mathrm{Var}(\hat{h}_f^*(X)) = \mathrm{Var}\left(\hat{h}_f(X)\right) + c^2 \,\mathrm{Var}\left(\hat{h_{0f}}(X)\right) + 2c\,\mathrm{Cov}\left(\hat{h}_f(X), \hat{h_{0f}}(X)\right) . \tag{5.23}$$

For the optimal choice of the parameter c, given by

$$c^* = -\frac{\mathrm{Cov}\left(\hat{h}_f(X), \hat{h_{0f}}(X)\right)}{\mathrm{Var}\left(\hat{h_{0f}}(X)\right)} ,$$

we have

$$\mathrm{Var}(\hat{h}_f^*(X)) = \left(1 - \rho^2\left(\hat{h}_f(X), \hat{h_{0f}}(X)\right)\right) \mathrm{Var}\left(\hat{h}_f(X)\right) ,$$

where $\rho^2\left(\hat{h}_f(X), \hat{h_{0f}}(X)\right)$ is the correlation coefficient between $\hat{h}_f(X)$ and $\hat{h_{0f}}(X)$. In this case, the overall variance is reduced. Once again, as in the case of the importance sampling estimator, there are numerical problems with calculation of c^*, because this optimal value depends on the evaluated integral $\mathbb{E}_f h(X)$. An incorrect choice of c may increase the variance of the control variate estimator.

Example 5.11. *Let us suppose that for some density $f(x)$ we would like to calculate the probability*

$$I = \mathrm{P}(X > a) = \int_a^\infty f(x)dx .$$

In this case the crude MC estimator is given by

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(X_i > a\right) ,$$

where $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x)$. If for some parameter $b < a$ we know the exact value of the probability $\mathrm{P}(X > b)$, then the additional estimator

$$\hat{I}_0 = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(X_i > b\right)$$

is available (for deeper details see (Robert & Casella 2004)). Then, the control variate estimator is given by

$$\hat{I}^* = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(X_i > a\right) + c\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(X_i > b\right) - \mathrm{P}(X > b)\right).$$

Since $\operatorname{Var}\hat{I}^* = \operatorname{Var}\hat{I} + c^2 \operatorname{Var}\hat{I}_0 + 2c \operatorname{Cov}\left(\hat{I}, \hat{I}_0\right)$ *and*

$$\operatorname{Cov}\left(\hat{I}, \hat{I}_0\right) = \frac{1}{n} \operatorname{P}(X > a)\left(1 - \operatorname{P}(X > b)\right) \ ,$$
$$\operatorname{Var}\hat{I}_0 = \frac{1}{n} \operatorname{P}(X > b)\left(1 - \operatorname{P}(X > b)\right) \ ,$$

then the control variate estimator $\hat{I}^*$ *will have lower variance compared to the "standard" approximation* $\hat{I}$*, if* $c < 0$ *and*

$$|c| < 2\frac{\operatorname{Cov}\left(\hat{I}, \hat{I}_0\right)}{\operatorname{Var}\hat{I}_0} = 2\frac{\operatorname{P}(X > a)}{\operatorname{P}(X > b)} \ .$$

5.4 Monte Carlo optimization

Apart from the integration problem, discussed in Section 5.2, also the optimization problem

$$\max_{x \in \mathcal{X}} h(x) \tag{5.24}$$

for some function $h(x)$ and a set $\mathcal{X} \subset \mathbb{R}^p$ is an area of application of Monte Carlo methods. If the deterministic numerical methods are used in this area, then checking the analytical properties of the target function, like convexity, boundedness, smoothness, constitutes usually the most important issue. Therefore, some preliminary steps for verifying the regularity conditions of $h(x)$ and devising a "special" algorithm for such case becomes necessary. Additionally, it is possible that the considered deterministic approach would find only the first, local maximum of $h(x)$, instead of the correct, global maximum of the target function.

Simulation methods do not rely on special analytical properties of the target function $h(x)$. And for some approaches there is probability equal to one that the algorithm will find the correct global maximum. In this section, some of these methods will be considered in a more detailed way. Additional discussion can be also found in, e.g., (Robert & Casella 2004).

The simulation methods for the optimization problem may be divided into two groups. The first group is related to the exploratory approach (e.g. the "naive approach", simulated annealing), where we are interested in fast exploration of the space $\mathcal{X}$. The second one is based on stochastic approximation (e.g. the EM algorithm), which is related to a probabilistic approximation of the target function $h(x)$.

5.4.1 A simple approach

If the set $\mathcal{X}$ is bounded, then the straightforward approach to solve the optimization problem (5.24) is to generate a sample $X_1, X_2, \ldots, X_n$ from the uniform

distribution on $\mathcal{X}$ and then approximate the exact solution via the simple estimator

$$\hat{h}_{\max} = \max_{i=1,\ldots,n} \{h(X_i)\} \ . \tag{5.25}$$

This method converges (for $n \to \infty$), but it may be very slow, since it does not take into account any specific feature of the target function $h(x)$. However, there is no possibility for such algorithm to "get stuck" in some local maximum, as this is the case of the deterministic, numerical method. This "naive" approach will be illustrated with the following example:

Example 5.12. *Let us consider the function*

$$h(x) = (\cos(50x) + \sin(20x))^2 \tag{5.26}$$

for $x \in [0, 1]$. *This function (see Figure 5.4) has many local maxima and minima, therefore a deterministic algorithm (e.g. Newton-Raphson algorithm, see (Robert & Casella 2004)) might go towards the nearest mode. Such behaviour does not allow to escape later from the local maximum to find the global one. Therefore, the possibility of finding the correct answer to the optimization problem depends on the appropriate starting condition in such a case. To the contrary, the "naive" estimator* (5.25) *performs the search in the whole domain* $[0, 1]$ *as it is emphasized by Figure 5.5.*

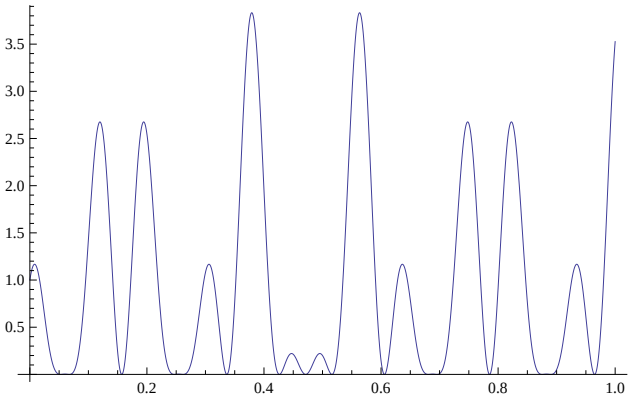

Fig. 5.4. Function $h(x)$ from Example 5.12

If $h(x)$ is positive, and if

$$\int_{\mathcal{X}} h(x)dx < \infty \ ,$$

then such function could be identified with some density without its normalizing constant. In this case, the optimization problem (5.25) may be seen as the one of

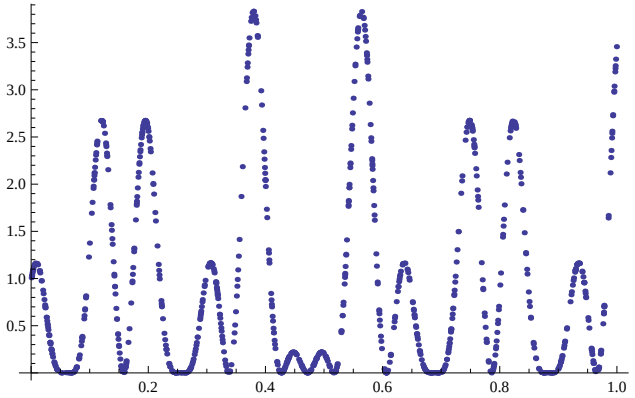

Fig. 5.5. Simulations for the function $h(x)$ from Example 5.12

finding the modes of the density $h(x)$. This approach may be also generalized for the cases when the assumptions mentioned above are not fulfilled (see (Robert & Casella 2004) for additional details).

5.4.2 Simulated annealing

A more sophisticated approach than the simple method, described in Section 5.4.1, is known as the *simulated annealing algorithm*. This method was introduced by Metropolis, Rosenbluth, Rosenbluth, Teller & Teller (1953) and it is based on changing the special parameter, referred to as *temperature* $T > 0$, which allows to partially avoid the traps of local maxima.

For the fixed starting value x_0, the consecutive values X_n are sampled according to the following algorithm:

Algorithm 5.13 (Simulated annealing)**.**

1. Sample the *proposed value* Y_n from some distribution in a neighbourhood of the previous value $X_{n-1} = x_{n-1}$ (e.g. from some uniform distribution). In a more general setting, this sampling could be done from the density given by $g(|x_{n-1} - Y_n|)$.
2. The next value X_n in the algorithm is generated as follows

$$X_n = \begin{cases} Y_n & \text{with probability } p = \min\{\exp(\Delta h/T), 1\} \\ x_{n-1} & \text{with probability } 1 - p \end{cases}, \qquad (5.27)$$

where $\Delta h = h(Y_n) - h(x_{n-1})$.

As it can be seen from the condition (5.27), if $h(Y_n) > h(x_{n-1})$ (i.e. value of the function $h(x)$ is higher for the new, proposed point Y_n than for the previous point x_{n-1}), then this proposed point is always accepted and $X_n = Y_n$. Otherwise, if $h(Y_n) \leq h(x_{n-1})$, then the new, proposed point may still be accepted with probability $p \neq 0$. And in this case $X_n = Y_n$ is also set. This property allows the algorithm to escape the attraction of some local maximum of $h(x)$. Of course, the described behaviour depends on the probability p, which is related to the choice of the temperature T. As T decreases toward 0, the values sampled in the simulated annealing algorithm become concentrated in a narrower neighbourhood of the local maxima of $h(x)$, because for $h(Y_n) \leq h(x_{n-1})$ we get $p \to 0$ if $T \to 0$.

Usually, the temperature T is modified at each iteration of the algorithm, i.e. the whole sequence $T_1, T_2, \ldots$ is set. But the convergence to global maximum depends on the rate of decrease of the temperature. In the case of finite spaces, the logarithmic rate $T_i = \Gamma / \log i$ or the geometric rate $T_i = \alpha^i T_0$ (for some parameter $0 < \alpha < 1$) of decrease are advised by some authors (see, e.g. (Robert & Casella 2004) for additional details).

Because in this algorithm the next value X_n depends on the previous value X_{n-1}, then the simulated annealing method is closer to MCMC (Markov Chain Monte Carlo) methods (see Chapter 6) than to the "standard" MC approach, where independent samples are usually generated.

5.4.3 The EM algorithm

Abbreviation EM originates from the two steps used in this approach: the expectation step (E-step) and the maximization step (M-step).

The EM method was introduced by Dempster, Laird & Rubin (1977) and is usually related to *missing data models*. In these models the target function $h(x)$, which should be optimized, is given by

$$h(x) = \mathbb{E}_Z \left(H(x, Z) \right) \ , \tag{5.28}$$

i.e. the expected value of some additional variable Z. This formulation is observed in practice, e.g. for censored data models (the data is censored by some unobservable variable), mixture models (indicator of the component, generating the observation, cannot be observed) or logistic regression. Additionally, artificial extensions which gives the formula (5.28) are also possible.

From the statistical point of view, in the missing data models the likelihood can be expressed as *demarginalization*

$$g(x|\theta) = \int_{\mathcal{Z}} f\left(x, z|\theta\right) dz \tag{5.29}$$

and the variable Z only simplifies calculations. The way Z is selected in order to satisfy (5.28), should not affect the value of the estimator. In this setting, the target function is expressed as some integral of a more manageable quantity.

For missing data models, the function

$$L^c(\theta|x,z) = f(x,z|\theta) = f_\theta(x,z) \tag{5.30}$$

is called *complete-model* (or *complete-data*) likelihood for observations of the complete data (X,Z). In (5.30), $f(x,z|\theta)$ is the joint density of variables X and Z for the fixed value of parameter of the statistical model θ. Let us suppose that we observe $X_1, X_2, \ldots, X_n$ – the *iid* sample from the density $g(x|\theta) = g_\theta(x)$. Then we are interested in maximum likelihood estimation, i.e. in finding

$$\hat{\theta} = \sup_\theta L(\theta) = \sup_\theta L(\theta|x_1, x_2, \ldots, x_n) = \sup_\theta L(\theta|\mathbf{x}) \ , \tag{5.31}$$

where $\mathbf{x}$ denotes the vector of observations $x_1, x_2, \ldots, x_n$. In a similar way, let $\mathbf{z} = z_1, z_2, \ldots, z_n$ be a vector of observations of additional variables $Z_1, Z_2, \ldots, Z_n$. Let $f(\mathbf{x},\mathbf{z}|\theta) = f_\theta(\mathbf{x},\mathbf{z})$ be a joint density of variables $X_1, \ldots, X_n$ and $Z_1, \ldots, Z_n$ for the fixed value of the parameter θ.

From the definition of conditional density, we get the basic identity for the EM algorithm

$$k(\mathbf{z}|\theta,x) = \frac{f(\mathbf{x},\mathbf{z}|\theta)}{g(\mathbf{x}|\theta)} \ , \tag{5.32}$$

where $k(\mathbf{z}|\theta,\mathbf{x})$ is the conditional density of the missing data $\mathbf{z}$ for the fixed observed data $\mathbf{x}$ and the value θ. This identity leads to the relationship between the complete-data likelihood $L^c(\theta|\mathbf{x},\mathbf{z})$ and the observed-data likelihood $L(\theta|\mathbf{x})$, which is given by

$$\log L(\theta|\mathbf{x}) = \mathbb{E}_{k,\theta_0}\left(\log L^c(\theta|\mathbf{x},\mathbf{z})\right) - \mathbb{E}_{k,\theta_0}\left(\log k(\mathbf{z}|\theta,\mathbf{x})\right) \tag{5.33}$$

for any value θ_0. On the right hand side of (5.33), the expected value is calculated with respect to θ_0 and the relevant density $k(\mathbf{z}|\theta_0,\mathbf{x})$. It should be noted that in the case of the EM algorithm, to maximize $\log L(\theta|\mathbf{x})$, only the first term on the right hand side of (5.33) should be taken into account, as the other term can be ignored (see also (Robert & Casella 2004) for additional details).

Let us denote the expected log-likelihood, as it is done usually in the case of EM algorithm, by

$$Q(\theta|\theta_0,\mathbf{x}) = \mathbb{E}_{k,\theta_0}\left(\log L^c(\theta|\mathbf{x},\mathbf{z})\right) \ .$$

Then, $Q(\theta|\theta_0,\mathbf{x})$ can be maximized, and if θ_1 is the value of θ maximizing $Q(\theta|\theta_0,\mathbf{x})$, then this value θ_1 can replace θ_0 in next step and the whole procedure can be repeated. Therefore, in the subsequent steps of the algorithm we calculate

$$\theta_{n+1} = \max_\theta Q(\theta|\theta_n,\mathbf{x}) \ .$$

This leads to the following method, which is started for some value θ_0:

Algorithm 5.14 (EM algorithm)**.**

1. (the E-step) For the given value θ_n, calculate

$$Q\left(\theta|\theta_n, \mathbf{x}\right) = \mathbb{E}_{k,\theta_n}\left(\log L^c\left(\theta|\mathbf{x}, \mathbf{z}\right)\right) \tag{5.34}$$

2. (the M-step) Find

$$\theta_{n+1} = \arg\max_{\theta} Q\left(\theta|\theta_n, \mathbf{x}\right) \ .$$

The iterations are conducted until a fixed point of Q is obtained.

Convergence of the EM algorithm is based on the following theorem, established by Dempster et al. (1977):

Theorem 5.15. *The sequence $\theta_0, \theta_1, \ldots$, generated by the EM algorithm, satisfies the inequality*

$$L(\theta_{n+1}|\mathbf{x}) \geq L(\theta_n|\mathbf{x}) \ ,$$

with equality holding if and only if

$$Q\left(\theta_{n+1}|\theta_n, \mathbf{x}\right) = Q\left(\theta_n|\theta_n, \mathbf{x}\right) \ .$$

Proof of this theorem can be found in (Robert & Casella 2004).

Theorem 5.15 guarantees that the likelihood will increase at each iteration of the EM algorithm, but it does not guarantee that the obtained sequence $\theta_0, \theta_1, \ldots$ converges to a maximum likelihood estimator. The next theorem gives the conditions for the sequence to converge to a stationary point, which may be a local maximum or saddle-point.

Theorem 5.16. *If the expected complete-data likelihood $Q\left(\theta|\theta_0, \mathbf{x}\right)$ is continuous in both θ and θ_0, then every limit point of the EM sequence $\theta_0, \theta_1, \ldots$ is a stationary point of $L(\theta|\mathbf{x})$, and the sequence $L(\theta_0|\mathbf{x}), L(\theta_1|\mathbf{x}), \ldots$ convergences monotonically to $L(\theta^*|\mathbf{x})$ for some stationary point θ^*.*

Because this convergence is guaranteed only for some stationary point, then additional conditions or techniques are necessary for finding the global maximum, i.e. to obtain the estimator in the main problem (5.31). Then, for example, the EM algorithm can be run a number of times with different (e.g. random) starting points.

Additionally, the MC method may be applied for the EM algorithm, this amounting to MCEM method (*Monte Carlo EM*). Each E-step in Algorithm 5.14 requires the computation of the expected log likelihood $Q\left(\theta|\theta_n, \mathbf{x}\right)$. Because this can be numerically difficult, the sample $Z_1, Z_2, \ldots, Z_m$ from the conditional density $k(\mathbf{z}|\theta_n, \mathbf{x})$ is generated and the approximation of the complete-data log-likelihood

$$\hat{Q}\left(\theta|\theta_n, \mathbf{x}\right) = \frac{1}{m}\sum_{i=1}^{m} \log L^c\left(\theta|\mathbf{x}, Z_1, Z_2, \ldots, Z_m\right)$$

is maximized.

6

Introduction to Markov Chain Monte Carlo methods

Maciej Romaniuk

Markov Chain Monte Carlo methods (abbreviated usually as *MCMC*), introduced by Metropolis et al. (1953) may be considered as a direct extension of teh MC methods. Generally speaking, instead of an independent, identically distributed sample as in the case of the Monte Carlo methods, the sequence of dependent random variables is generated during the MCMC approach. This specially selected sequence forms a stochastic process with discrete time, known as Markov chain. In this chapter the necessary definitions and theorems, concerning theory of MCMC methods, as well as some examples of MCMC algorithms will be considered.

In practice, MCMC methods are applied using various software libraries and special programs (like BUGS, see (Thomas, Spiegelhalter & Gilks 1992)). These simulation methods are used in many areas, e.g., in physics, biology, statistics etc. Some examples of applications will also be further provided.

6.1 MCMC as a generalization of MC

As it was mentioned, MCMC methods may be seen as a generalization of Monte Carlo methods. The main aim of introducing the Markov chain instead of the independent sample is to eliminate one of the most important disadvantages of Monte Carlo approach – the necessity of sampling *iid* random variables directly from the specified density $f(x)$ as this was presented in Chapter 5. Such pdf could be a numerically complex function or its normalizing constant is sometimes unknown or difficult to evaluate. Also the relevant sampling algorithm for $f(x)$ could be numerically inadequate for practical use.

In the case of MCMC methods, instead of sampling *iid* random sequence from the specified density $f(x)$, a Markov chain $(X_i)_{i=0}$ with an appropriate stationary distribution $f(x)$ is generated. If such Markov chain fulfils the assumptions of the ergodicity theorems, considered in Section 3.3 (i.e. we use the ergodic chain), then the empirical average converges to the relevant expected value. Therefore, in this setting there is no necessity for sampling directly from the density $f(x)$, only the sequence approximately distributed from $f(x)$ is generated. This idea leads to the following definition:

Definition 6.1. *A* Markov Chain Monte Carlo method *for simulation of the density $f(x)$ is any method producing an ergodic Markov chain, whose stationary distribution is given by this pdf $f(x)$.*

To apply the MCMC approach and to generate the Markov chain with the fixed stationary distribution in order to solve, e.g., the integration problem stated in (5.6), two general algorithms can be applied. The first one is the Metropolis–Hastings (MH) algorithm. The similarities between the MH algorithm and the simulated annealing approach will be discussed further on. The second method is known as the Gibbs sampler. This algorithm is especially useful in the multi-dimensional setting.

6.2 The Metropolis–Hastings algorithm

In order to generate the sequence $X_1, X_2, \ldots, X_n, \ldots$ via the *Metropolis–Hastings algorithm* (abbreviated further as MH), the conditional density $g(y|x)$, known as the *instrumental* or the *proposal* density, is selected by an experimenter. To implement the MH approach, almost any density $g(y|x)$ can be chosen, but from the practical point of view the following necessary conditions should be fulfilled:

1. It should be numerically easy to simulate from the density $g(y|x)$.
2. It should be possible to evaluate the ratio $f(y)/g(y|x)$ (up to a constant independent of x) or the instrumental density is symmetric, i.e. it fulfils the condition $g(x|y) = g(y|x)$.
3. The set of supports of $g(.|x)$ should include the whole support of the target density $f(.)$.
4. The constructed MC should be ergodic. The necessary conditions for fulfilling this requirement will be considered later on as relevant theorems.

For the fixed starting point x_0, the subsequent steps of the MH algorithm are given by:

Algorithm 6.2 (Metropolis–Hastings algorithm)**.**

1. Sample variable Y_{i-1} from the instrumental density $g(.|x_{i-1})$.
2. Generate new value X_i according to the formula:

$$X_i = \begin{cases} Y_{i-1} & \text{with probability } p(x_{i-1}, Y_{i-1}) \\ x_{i-1} & \text{with probability } 1 - p(x_{i-1}, Y_{i-1}) \end{cases}, \tag{6.1}$$

where

$$p(x, y) = \min\left\{ \frac{f(y)}{g(y|x)} \frac{g(x|y)}{f(x)}, 1 \right\} . \tag{6.2}$$

The probability $p(x, y)$, given by the condition (6.2), is known as the *Metropolis–Hastings acceptance probability*. It is used to decide if the new generated value Y_{i-1} is accepted as X_i.

6.2.1 Convergence properties

As it was previously noted, the chain produced via the MCMC method should be ergodic. Only in such case the ergodic theorems can be applied and then the necessary convergence of the empirical average to the expected value is achieved.

The first theorem specifies the minimal necessary condition for the instrumental density $g(.|x)$:

Theorem 6.3. *Let $(X_i)_{i=0}$ be the chain produced by MH approach described by Algorithm 6.2. For every conditional distribution $g(.|.)$, whose support includes the support of $f(.)$, $f(.)$ is the stationary distribution of this Markov chain.*

The proof of this theorem can be found in (Robert & Casella 2004).

As it was pointed out in Chapter 3, irreducibility and Harris recurrence are the most important properties of the considered Markov chain. Also the aperiodicity of the chain should be taken into account from the practical point of view. If the chain is not aperiodic, then the relevant mean, based on this chain, significantly changes from step to step. Such behaviour can increase the estimation error.

Then, the following theorems can be applied to check the above requirements:

Theorem 6.4. *If for all $x, y \in \mathcal{S} \times \mathcal{S}$ the* positivity *condition of the conditional density*

$$g(y|x) > 0 \tag{6.3}$$

is fulfilled, then the chain produced by Algorithm 6.2 is irreducible and Harris recurrent. If, additionally, the events such that $X_{i+1} = X_i$ are allowed, i.e.

$$\mathrm{P}\left(\frac{f(y)}{g(y|x)}\frac{g(y|x)}{f(x)} < 1\right) > 0 \ , \tag{6.4}$$

then this MC is aperiodic.

The proof of this theorem can be found in (Robert & Casella 2004).

The inequality (6.3) in the positivity condition is equivalent to the condition that every set can be reached in a single step. And the inequality (6.4) leads directly to the conclusion that

$$P(X_{i+1} = X_i) > 0 \ ,$$

i.e. the probability that MC remains in some state for the next step is not equal to zero.

There is also another theorem that can be applied:

Theorem 6.5. *Assume that the density $f(x)$ is bounded and positive on every compact set of its support $\mathcal{S}$. If there exist positive numbers $\epsilon, \delta > 0$ such that*

$$g(y|x) > \epsilon \text{ for } |x - y| < \delta \ , \tag{6.5}$$

then the MC produced by Algorithm 6.2 is irreducible and aperiodic. Additionally, every nonempty compact set is a small set. If $h : \mathcal{S} \to \mathbb{R}$ is such that

$$\mathbb{E}_f |h(X)| < \infty \ ,$$

then, for every initial distribution π_{X_0} there is

$$\frac{1}{n} \sum_{k=1}^{n} h(X_k) \xrightarrow[n\to\infty]{a.s.} \int_{\mathcal{S}} h(x) f(x) dx \tag{6.6}$$

and

$$\lim_{n\to\infty} \left\| \int_{\mathcal{S}} \mathcal{K}^n(x, .) d\pi_{X_0}(x) - f(.) \right\|_{\text{TV}} = 0 \ . \tag{6.7}$$

The proof of this theorem can be found in (Robert & Casella 2004). The limits (6.6) and (6.7) mean that the empirical average, based on the chain constructed by the MH algorithm, converges to the appropriate expected value and the limit distribution of this Markov chain is the same as the target, stationary distribution $f(.)$.

The condition (6.5) can be identified with the conclusion that the instrumental density of transitions in a neighbourhood of x is bounded from below.

6.2.2 Selection of the instrumental distribution

Because of the previously mentioned universality in the selection of the instrumental density, there are many ways to choose this function. Some kinds of densities are common choices for the experimenters, and therefore they will be discussed in a more detailed way.

One of the simplest approaches is the *independent MH* algorithm. In this case, the sampling density $g(y|x)$ is independent of the variable x, i.e. it is given by the function $g(y)$. This leads to the following algorithm:

Algorithm 6.6 (Independent MH)**.**

1. Sample Y_{i-1} from the density $g(.)$.
2. Generate new value X_i according to the formula:

$$X_i = \begin{cases} Y_{i-1} & \text{with probability } p(x_{i-1}, Y_{i-1}) \\ x_{i-1} & \text{with probability } 1 - p(x_{i-1}, Y_{i-1}) \end{cases} \ , \tag{6.8}$$

where

$$p(x, y) = \min \left\{ \frac{f(y)}{g(y)} \frac{g(x)}{f(x)}, 1 \right\} \ . \tag{6.9}$$

This method may be seen as a straightforward generalization of the rejection algorithm (see Section 4.1.2), because sampling from the instrumental density $g(.)$ is independent of the previous value x_{i-1}. In this case, the following convergence theorem can be additionally applied:

Theorem 6.7. *The independent MH approach, described by Algorithm 6.6, produces an ergodic chain if there exists a constant M such that*

$$f(x) \leq Mg(x)$$

for all $x \in \mathcal{S}$.

This theorem is proved in (Robert & Casella 2004).

Another candidate for the instrumental density is given by the symmetrical function $g(y|x)$, i.e. the density satisfying $g(x|y) = g(y|x)$. In this case, the relevant algorithm is similar to the simulated annealing method (see Section 5.4.2), especially because the obtained acceptance probability is equal to

$$p(x, y) = \min\left\{\frac{f(y)}{f(x)}, 1\right\} \ . \tag{6.10}$$

Therefore, in order to evaluate $p(x, y)$, information about the value of $g(y|x)$ is unnecessary. Only the ratio $f(y)/f(x)$ should be calculated, which is especially helpful if the normalizing constant for the density $f(.)$ is unknown.

The next approach is related to a local exploration of the neighbourhood of the current value of the chain. Once again, this idea is similar to the one of the simulated annealing algorithm. In this case, the first choice for the instrumental distribution is some random perturbation of X_i, i.e. application of the formula

$$Y_i = X_i + \varepsilon_i \ , \tag{6.11}$$

where ε_i is some random variable independent of X_i. The chain, constructed by applying the transformation (6.11), is known as *random walk*.

Then, the uniform distribution on some ball or the normal distribution centred at X_i can also be applied as the instrumental density. If the respective pdf is also a symmetric function, then g depends only on the distance between y and x, i.e. it has the form $g(|y - x|)$. This assumption leads to the following algorithm:

Algorithm 6.8 (Symmetric random walk MH)**.**

1. Sample Y_{i-1} from the density $g(|y - x_{i-1}|)$.
2. Generate new value X_i according to the formula:

$$X_i = \begin{cases} Y_{i-1} & \text{with probability } p(x_{i-1}, Y_{i-1}) \\ x_{i-1} & \text{with probability } 1 - p(x_{i-1}, Y_{i-1}) \end{cases} ,$$

where

$$p(x,y) = \min\left\{\frac{f(y)}{f(x)}, 1\right\} .$$

It should be noted that selection of the instrumental density can be crucial for the convergence speed of the MH algorithm. In the case of the independent MH algorithm, the choice of g should maximize the average acceptance rate, which implies the similarity between the instrumental density and the target density $f(.)$. To the contrary, in the case of the random walk MH algorithm, the higher acceptance rate corresponds to a slower convergence as the moves on the support of f are more limited (because the new, sampled point is close to the previous one). However, these conclusions are rather general ideas than the direct solutions for the convergence problem (see (Robert & Casella 2004) for additional details).

6.2.3 The Rao–Blackwellized estimator

Apart from the classical form of the estimator for the expected value, given by the "standard" average (5.8), other approximations of $\mathbb{E}_f h(X)$ are possible in the case of the MH algorithm. As in the rejection method, the Metropolis–Hastings approach does not take advantage of the entire set of random variables, which are generated. Because some of the sampled variables are rejected, it could be expected that these values bring (indirectly) some additional information about the density $f(.)$. Then the *Rao–Blackwellization* approach may be applied, leading to a special, conditional (known as *Rao–Blackwellized*) estimator $\hat{h}_f^{RB}(X)$. This estimator $\hat{h}_f^{RB}(X)$ dominates the standard estimator $\hat{h}_f(X)$ under the quadratic loss function.

It should be noted that the sequence $X_1, X_2, \ldots$, generated by the MH algorithm is based on two samples: $Y_1, Y_2, \ldots$ and $U_1, U_2, \ldots$, where $Y_i \sim g(.|x_i)$ and $U_i \sim U$. This second sequence is used to reject or accept the sampled points Y_i according to the acceptance probability $p(x_i, Y_i)$. Based on this observation, we have

$$\hat{h}_f(X) = \frac{1}{n}\sum_{i=1}^{n} h(X_i) = \frac{1}{n}\sum_{i=1}^{n} h(Y_i)\sum_{j=i}^{n} \mathbb{1}(X_j = Y_i) ,$$

which leads to the formula, referring to conditional expectation

$$\hat{h}_f(X) = \frac{1}{n}\sum_{i=1}^{n} h(Y_i)\mathbb{E}\left(\sum_{j=i}^{n} \mathbb{1}(X_j = Y_i)|Y_1, Y_2, \ldots, Y_n\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} h(Y_i)\left(\sum_{j=i}^{n} \mathrm{P}(X_j = Y_i\,|Y_1, Y_2, \ldots, Y_n\,)\right) . \quad (6.12)$$

From the practical point of view, it is an important feature of this method that the probabilities $\mathrm{P}(X_j = Y_i \,|Y_1, Y_2, \ldots, Y_n)$, introduced in the conditional formula (6.12), can be explicitly computed (see, e.g. (Casella & Robert 1996)). For example, suppose that the independent MH algorithm (see Algorithm 6.6) is applied and that $X_0 \sim f$, i.e. the starting distribution, is the same as the target density $f(.)$. If we denote

$$w_i = \frac{f(y_i)}{g(y_i)} \ , \ v_{ij} = \min\left(\frac{w_i}{w_j}, 1\right) \ \text{for } 0 \le i < j \ , \tag{6.13}$$

$$z_{ii} = 1 \ , \ z_{ij} = \prod_{k=i+1}^{j} (1 - v_{ij}) \ \text{for } i < j \ , \tag{6.14}$$

then we have the following theorem:

Theorem 6.9. *The Rao–Blackwellized estimator can be written as*

$$\hat{h}_f^{RB}(X) = \frac{1}{n+1} \sum_{i=0}^{n} \varphi_i h(Y_i) \ , \tag{6.15}$$

where

$$\varphi_i = \tau_i \sum_{j=i}^{n} z_{ij}$$

and $\tau_i = \mathrm{P}(X_i = Y_i \,|Y_0, Y_1, \ldots, Y_n)$ *is given by*

$$\tau_0 = 1 \ , \ \tau_i = \sum_{j=0}^{i-1} \tau_j z_{j(i-1)} v_{ji} \ \textit{for } i > 0 \ .$$

This theorem is proved in (Robert & Casella 2004).

6.3 The Gibbs sampler

The next MCMC algorithm, known as the *Gibbs sampler*, can be applied only in the multidimensional setting. To clarify the presentation, two different cases of this approach will be further discussed: the two-stage Gibbs sampler and the multi-stage Gibbs sampler.

6.3.1 The two-stage Gibbs sampler

We start from the simpler, two-dimensional case of the Gibbs sampler. Let us suppose that we are interested in generation of variables $(X_1, Y_1), (X_2, Y_2), \ldots$ from the two-dimensional random vector (X, Y) with a joint density $f(x, y)$. If the methods for sampling from both of the conditional densities $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$, related to the joint pdf $f(x, y)$, are known, then for the fixed starting point x_0 the following algorithm can be applied:

Algorithm 6.10 (Two-stage Gibbs sampler)**.**

1. Generate Y_i from the density $f_{Y|X}(.|x_{i-1})$.
2. Generate X_i from the density $f_{X|Y}(.|y_i)$.

Then, these two steps are repeated. In this case, the subsequent values of the random variables X or Y are generated alternately from the appropriate conditional densities, i.e. the sampler "jumps" between the two coordinates of the random vector (X, Y).

The two-stage Gibbs sampler can be applied if the joint pdf $f(x, y)$ is numerically too complex for direct simulations. It should be noted that not only the whole sequence $(X_i, Y_i)_{i=1}$ is a Markov chain, but also each of subsequences $X_0, X_1, \ldots$ and $Y_0, Y_1, \ldots$ has the same property. The stationary distributions of these subsamples are the relevant marginal densities $f_X(.)$ and $f_Y(.)$. Therefore, this method can be applied in somewhat opposite way, not as the solution for the complex joint density. If we start from a marginal density $f_X(.)$, then the second, auxiliary variable Y is added, which is not directly relevant for us from the statistical point of view. Such approach leads us to a new joint density $f(x, y)$. There are many settings ,where this natural completion of the density $f_X(.)$ (which is our aim) into $f(x, y)$ exists, e.g. missing data models (see Section 5.4.3). In this case, the conditional densities $f_{Y|X}(.)$ and $f_{X|Y}(.)$ may be easier to simulate than the joint pdf or the marginal density $f_X(.)$.

Some theorems, which are related to the convergence properties of the two-stage Gibbs sampler should be mentioned. These theorems are proved in (Robert & Casella 2004).

Definition 6.11. *Let* $\mathbb{X} = \left(X^{(1)}, X^{(2)}, \ldots, X^{(m)}\right)$ *be an m-dimensional random vector with the joint density* $f_{\mathbb{X}}\left(x^{(1)}, \ldots, x^{(m)}\right)$*, and* $f_{X^{(i)}}(.)$ *be the marginal density of the variable* $x^{(i)}$*. If*

$$f_{X^{(i)}}\left(x^{(i)}\right) > 0 \text{ for every } i = 1, \ldots, m \Rightarrow f_{\mathbb{X}}\left(x^{(1)}, \ldots, x^{(m)}\right) > 0 \ , \qquad (6.16)$$

then $f_{\mathbb{X}}$ *satisfies the* positivity condition*.*

The above definition means that the support of $f_{\mathbb{X}}\left(x^{(1)}, \ldots, x^{(m)}\right)$ is the Cartesian product of the supports of the marginal densities $f_{X^{(i)}}(.)$

Lemma 6.12. *Each of the sequences* $X_0, X_1, \ldots$ *and* $Y_0, Y_1, \ldots$*, produced by the two-stage Gibbs sampler, is a Markov chain with the corresponding stationary distribution*

$$f_X(x) = \int f(x, y)dy \ , \ f_Y(y) = \int f(x, y)dx \ .$$

If the positivity constraint on $f(x, y)$ *holds, then both of these chains are irreducible.*

Theorem 6.13. *If the positivity constraint on $f(x, y)$ holds and the transition kernel*

$$\mathcal{K}_{(X,Y)}((x_{i-1}, y_{i-1}), (x_i, y_i)) = f_{Y|X}(y_i|x_{i-1}) f_{X|Y}(x_i|y_i)$$

is absolutely continuous with respect to the dominating measure, the chain (X, Y) is Harris recurrent and ergodic with the stationary distribution $f(x, y)$.

In Section 4.8, sampling from multivariate normal distribution based on Cholesky decomposition was considered. It is also possible to apply the Gibbs sampler for such a problem.

Example 6.14. *Let us suppose that we would like to sample from the two-dimensional random variable $\mathbb{X} \sim N(0, \mathbb{W})$, where*

$$\mathbb{W} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} .$$

Then, the two-dimensional Gibbs sampler for the starting point x_0 is described by the following steps:

Algorithm 6.15 (Two-stage Gibbs sampler for the multivariate normal distribution)**.**

1. *Generate $Y_i \sim N\left(\rho x_{i-1}, 1 - \rho^2\right)$.*
2. *Generate $X_i \sim N\left(\rho y_i, 1 - \rho^2\right)$.*

6.3.2 The multi-stage Gibbs sampler

In the case of multidimensional Gibbs sampler, its aim is to sample the sequence of m-dimensional random vectors $\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_n, \ldots$ from the joint density $f(\mathbb{X})$. Let $\left(X_i^{(1)}, \ldots, X_i^{(m)}\right)$ denote the coordinates of the random vector $\mathbb{X}_i$. In a more general case, each of the coordinates $X_i^{(j)}$ could be also some multidimensional random vectors.

Denote by $\mathbb{X}_i^{(-j)}$ the random vector $\left(X_i^{(1)}, \ldots, X_i^{(j-1)}, X_i^{(j+1)}, \ldots, X_i^{(m)}\right)$, i.e. the vector $\mathbb{X}_i$ without its j-th component. Let us assume that we dispose of the numerically effective algorithms for sampling from the conditional densities $f_{X^{(j)}|\mathbb{X}^{(-j)}}\left(.|x^{(1)}, \ldots, x^{(j-1)}, x^{(j+1)}, \ldots, x^{(m)}\right)$ for $j = 1, \ldots, m$. Such densities are known as *full conditionals*.

Then, for the fixed starting point $\boldsymbol{x}_0 = \left(x_0^{(1)}, \ldots, x_0^{(m)}\right)$, the following algorithm may be applied:

Algorithm 6.16 (Gibbs sampler with systematic scan)**.**

1. Generate $X_{i+1}^{(1)}$ from the density $f_{X^{(1)}|\mathbb{X}^{(-1)}}\left(.|x_i^{(2)},\ldots,x_i^{(m)}\right)$.
2. Generate $X_{i+1}^{(2)}$ from the density $f_{X^{(2)}|\mathbb{X}^{(-2)}}\left(.|x_{i+1}^{(1)},x_i^{(3)},\ldots,x_i^{(m)}\right)$.
3. Generate $X_{i+1}^{(3)}$ from the density $f_{X^{(3)}|\mathbb{X}^{(-3)}}\left(.|x_{i+1}^{(1)},x_{i+1}^{(2)},x_i^{(4)},\ldots,x_i^{(m)}\right)$.
4. ...

m. Generate $X_{i+1}^{(m)}$ from the density $f_{X^{(m)}|\mathbb{X}^{(-m)}}\left(.|x_{i+1}^{(1)},x_{i+1}^{(2)},\ldots,x_{i+1}^{(m-1)}\right)$.

Algorithm 6.16 is the straightforward generalization of the two-dimensional case, described by Algorithm 6.10. The subsequent components of the random vector $\mathbb{X}$ are sampled in a systematic way and the algorithm proceeds in "one direction" during consecutive iterations. Therefore, this approach is known as the Gibbs sampler with *systematic scan* (or *systematic sweep*).

Example 6.17. *The following* autoexponential model *was introduced by Besag (1974). If* $\boldsymbol{x}\in\mathbb{R}_+^3$*, then the corresponding joint density is given as*

$$f(\boldsymbol{x})\propto\exp\left(-\left(x^{(1)}+x^{(2)}+x^{(3)}+\theta_{12}x^{(1)}x^{(2)}+\theta_{23}x^{(2)}x^{(3)}+\theta_{31}x^{(3)}x^{(1)}\right)\right)\ ,$$

where $\theta_{ij}>0$ *are known. In this model the full conditionals are exponential densities, e.g.*

$$X^{(3)}|x^{(1)},x^{(2)}\sim\mathrm{Exp}\left(1+\theta_{23}x^{(2)}+\theta_{31}x^{(1)}\right)\ .$$

Therefore, simulation from these distributions is straightforward. Other conditionals and the marginal distributions are more numerically complex functions.

Other approaches for the selection of the coordinates of $\mathbb{X}$, apart from the systematic one, are also possible. The first example is the symmetric scan (or the reversible Gibbs sampler) and the second one is random sweep.

The *symmetric scan* of components of the random vector $\mathbb{X}$ is described by the following algorithm, which is, as previously, started from a fixed starting point $\boldsymbol{x}_0$:

Algorithm 6.18 (Gibbs sampler with symmetric scan)**.**

1. Generate $X_*^{(1)}$ from the density $f_{X^{(1)}|\mathbb{X}^{(-1)}}\left(.|x_i^{(2)},\ldots,x_i^{(m)}\right)$.
2. Generate $X_*^{(2)}$ from the density $f_{X^{(2)}|\mathbb{X}^{(-2)}}\left(.|x_*^{(1)},x_i^{(3)},\ldots,x_i^{(m)}\right)$.
3. Generate $X_*^{(3)}$ from the density $f_{X^{(3)}|\mathbb{X}^{(-3)}}\left(.|x_*^{(1)},x_*^{(2)},x_i^{(4)},\ldots,x_i^{(m)}\right)$.

4. ...
m. Generate $X_{i+1}^{(m)}$ from the density

$$f_{X^{(m)}|\mathbb{X}^{(-m)}}\left(.|x_*^{(1)}, x_*^{(2)}, \ldots, x_*^{(m-1)}\right) .$$

m+1. Generate $X_{i+1}^{(m-1)}$ from the density

$$f_{X^{(m-1)}|\mathbb{X}^{(-(m-1))}}\left(.|x_*^{(1)}, x_*^{(2)}, \ldots, x_*^{(m-2)}, x_{i+1}^{(m)}\right) .$$

m+2. ...
2m-1. Generate $X_{i+1}^{(1)}$ from the density $f_{X^{(1)}|\mathbb{X}^{(-1)}}\left(.|x_{i+1}^{(2)}, \ldots, x_{i+1}^{(m)}\right)$.

In this case only one simulation of each component out of two samples is directly used as the final output, but the chain generated in this way has some important theoretical properties (like *reversibility* of the chain, see (Robert & Casella 2004) for additional details).

The name *random sweep Gibbs sampler* is related to the random selection of the components of the generated vector. The coordinates are chosen in a random order before each set of transitions. As previously, the algorithm is started from a fixed starting point $\boldsymbol{x}_0$, and then the following steps are applied:

Algorithm 6.19 (Random sweep Gibbs sampler)**.**

1. Generate permutation Σ of m elements (which leads to rearrangement of m components of the random vector).
2. Generate $X_{i+1}^{(\Sigma_1)}$ from the density $f_{X^{(\Sigma_1)}|\mathbb{X}^{(-\Sigma_1)}}\left(x_{\Sigma_1}|\boldsymbol{x}_i^{(-\Sigma_1)}\right)$, where Σ_1 is the first value from the permutation Σ.
3. ...

m+1. Generate $X_{i+1}^{(\Sigma_m)}$ from the density $f_{X^{(\Sigma_m)}|\mathbb{X}^{(-\Sigma_m)}}\left(x_{\Sigma_m}|\boldsymbol{x}_{i+1}^{(-\Sigma_m)}\right)$, where Σ_m is the last value from the permutation Σ.

As previously, in the case of the two-dimensional Gibbs sampler, it is necessary to asses the convergence of the generated chain $\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_n, \ldots$. Additionally, the assumptions of the relevant theorem should not be too complex to verify in practice. An example of such approach is given by the following theorem:

Theorem 6.20. *If the conditional densities $f_{X^{(j)}|\mathbb{X}^{(-j)}}$ fulfil the conditions:*

1. *let $\boldsymbol{x} = (x^{(1)}, \ldots, x^{(m)})$ and $\boldsymbol{x'} = (x'^{(1)}, \ldots, x'^{(m)})$, and there exists $\delta > 0$ for which $\boldsymbol{x}, \boldsymbol{x'} \in \text{supp}(f_{\mathbb{X}})$, $|\boldsymbol{x} - \boldsymbol{x'}| < \delta$,*

$$f_{X^{(j)}|\mathbb{X}^{(-j)}}\left(x^{(j)}\middle| x^{(1)},\ldots,x^{(j-1)},x'^{(j+1)},\ldots,x'^{(m)}\right)>0 \text{ for } i=1,\ldots,m\ ,$$

2. *there exists* $\delta'<\delta$ *such that every pair* $\boldsymbol{x},\boldsymbol{x}'\in\operatorname{supp}(f_{\mathbb{X}})$ *can be connected by a finite sequence of balls with radius* δ'*, having the (stationary) measure of the intersection of two consecutive balls positive,*

then the MC produced by the Gibbs sampler is irreducible and aperiodic. Additionally, if $h:\mathcal{S}\to\mathbb{R}$ *is such that*

$$\mathbb{E}_f|h(X)|<\infty\ ,$$

then, for every initial distribution π_{X_0} *there holds*

$$\frac{1}{n}\sum_{k=1}^{n}h(X_k)\xrightarrow[n\to\infty]{a.s.}\int_{\mathcal{S}}h(x)f(x)dx$$

and

$$\lim_{n\to\infty}\left\|\int_{\mathcal{S}}\mathcal{K}^n(x,.)d\pi_{X_0}(x)-f(.)\right\|_{\mathrm{TV}}=0\ .$$

The conditions, mentioned in Theorem 6.20, which is proved in (Robert & Casella 2004), can be more directly explained. The first one informs about the possibility of transition between two states $\boldsymbol{x}$ and $\boldsymbol{x}'$, which are close enough. The second one specifies that the whole support of the Gibbs sampler is connected via the related sequence of intersecting balls.

Other useful theorems, concerning the issues of convergence, are discussed in (Robert & Casella 2004).

6.4 Hybrid MCMC methods

From the theoretical point of view, there is an important similarity between the MH algorithm and the Gibbs sampler, which is confirmed by the following theorem:

Theorem 6.21. *The multi-stage Gibbs sampler with* m *steps is equivalent to the composition of* m *Metropolis–Hastings algorithms with acceptance probabilities uniformly equal to one.*

This theorem is proved in (Robert & Casella 2004).

However, from the practical point of view, there are important differences between these two types of MCMC methods. Firstly, the acceptance probability of the Gibbs sampler is always equal to one. Therefore any discussion as to the optimal acceptance rate is unnecessary. In the case of the MH algorithm, the rejection probability is given by $1-p(x_{i-1},Y_i)$ (see (6.2)) which is not always equal to zero.

Secondly, in the case of the Gibbs sampler, knowledge of some analytical or probabilistic properties of the joint pdf f and the numerically efficient methods

of sampling from the full conditionals $f_{X^{(j)}|\mathbb{X}^{(-j)}}\left(.|\boldsymbol{x}^{(-j)}\right)$ are required. Then, the choice of these conditional densities is limited in some way. In the case of the MH algorithm, the instrumental distribution could be more freely selected. This constitutes both an advantage and a disadvantage. For the MH algorithm it is easier to find the generation method of the instrumental density which could be directly numerically applied, but it is also possible that the selected density is incorrect or misleading (e.g. having nonconnected support).

Thirdly, the construction of the Gibbs sampler is always multidimensional. It is possible that some components of the random vector are artificially introduced or unnecessary for the considered output (e.g. like in the case of missing data models), but generation of at least two-dimensional vectors is always required.

Summarising, there are important differences in the way of construction of the MH algorithm and the Gibbs sampler for practical applications. It seems that the Gibbs sampler is favoured, because this simulation method is based on conditional densities of the *true* target density $f(.)$. In the case of the MH algorithm, sampling is related to some approximation of this density $f(.)$, and rejection of a part of the generated values is necessary. But a decomposition of the joint density $f(.)$ for a particular system of coordinates does not ensure the validity of the Gibbs sampler. A wrong choice of the components can lead to the increase of the convergence time or to getting into a trapping state (see (Robert & Casella 2004) for the relevant examples).

In the case of the MH algorithm, problems are more related to a bad agreement between the target density and the instrumental pdf. Therefore, the remedy, which takes advantage of both of these algorithms, is to implement a *hybrid* approach that uses both the Gibbs sampler and the MH algorithm. This method is known as *hybrid MCMC algorithm* and is described by the following definition:

Definition 6.22. *Let $\mathcal{K}_X^{(1)}, \mathcal{K}_X^{(2)}, \ldots, \mathcal{K}_X^{(p)}$ be the transition kernels, which correspond to different steps of some MCMC method. If $a_1, \ldots, a_p$ is some probability distribution, then a* mixture *of $\mathcal{K}_X^{(1)}, \mathcal{K}_X^{(2)}, \ldots, \mathcal{K}_X^{(p)}$ is an algorithm associated with the kernel*

$$\mathcal{K}_X = a_1\mathcal{K}_X^{(1)} + \ldots + a_p\mathcal{K}_X^{(p)} \ .$$

A cycle *of $\mathcal{K}_X^{(1)}, \mathcal{K}_X^{(2)}, \ldots, \mathcal{K}_X^{(p)}$ is the algorithm with the kernel*

$$\mathcal{K}_X = \mathcal{K}_X^{(1)} \circ \ldots \circ \mathcal{K}_X^{(p)} \ .$$

Construction of the hybrid MCMC algorithm is based on some composition of both the MH algorithm and the Gibbs sampler which are appropriately selected. For example, every p iterations, the step of the Gibbs sampler is replaced with a step of the MH algorithm (then we have a cycle) or, at each iteration, this MH step is selected with some probability (which gives a mixture). The hybrid methods are valid from the theoretical point of view, when the heterogeneity of the chains generated by the cycles is removed by considering only the appropriate subchains of the whole output (see, e.g. (Robert & Casella 2004)).

6.5 Examples of applications of MCMC methods

In this section we discuss some examples of applications of MCMC methods. We start from the general remarks concerning hierarchical models related to Bayesian analysis. Then, the Ising model and the statistical approach to noise reduction methods will be considered.

6.5.1 Hierarchical models

The Gibbs sampler is particularly well adapted to the statistical models known as the *hierarchical structures*. In these models, a special form of the density $f(x)$ is used, which can be decomposed as

$$f(x) = \int f_1(x|y_1) f_2(y_1|y_2) \dots f_i(y_i) dy_1 \, dy_2 \dots dy_i$$

for some $i \geq 1$. This decomposition can be done due to structural (i.e. related to the structure of the model itself) or computational reasons (compare with missing data models, see Section 5.4.3). If prior information is sparse, the noninformative densities (e.g. uniform distribution) can be introduced in the hierarchical modelling at various levels of hierarchy of the respective model.

The hierarchical models appear in a natural way in the Bayesian analysis (see, e.g. (Gilks, Richardson & Spiegelhalter 1996)). In this case, the structure of the model or the variability of the observations may require introduction of several levels of prior distributions. Then, these levels and relations among them can be illustrated using a graph, which is known as *DAG (direct, acyclic graph)*. The DAG consists of nodes and arrows. Parents of the node are all the nodes which are directly connected with this node and the related arrows are directed to this node. Children of the node are all the nodes which are directly connected with this node and the related arrows are directed from this node. For example, in Figure 6.1, nodes `B` and `C` are children of node `A`, node `A` is parent of node `C`, node `C` is parent of node `D`. In statistical applications, the nodes represent random distributions of various parameters of the considered model, and arrows denote the statistical relations between these distributions.

A relevant example, considered in (Robert & Casella 2004) and (Schukken, Casella & van den Broek 1991) illustrates the application of the hierarchical models:

Example 6.23. *In animal epidemiology sometimes data from groups of animals (like litters or herds) is used. Because some diseases are infectious, then usual assumptions about independence cannot be not fulfilled. Schukken et al. (1991) collected the data of the number of cases of clinical mastitis in dairy cattle herds over a one year period and applied the hierarchical structure to model these data. It is assumed that, in each herd, the occurrence of this disease is a Bernoulli random variable. Let X_i denote the number of cases in the herd i, which leads to $X_i \sim \text{Poiss}(\lambda_i)$, where λ_i is the rate of the infection in herd i. Because of the*

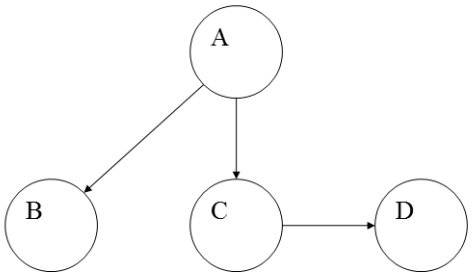

Fig. 6.1. An example of DAG

lack of independence, caused by the fact that this disease is infectious, Schukken et al. (1991) specified the following hierarchical model

$$\begin{aligned} X_i &\sim \text{Poiss}(\lambda_i), \\ \lambda_i &\sim \Gamma(\alpha, \beta_i), \\ \beta_i &\sim \Gamma(a, b), \end{aligned}$$

where α, a, b are fixed parameters. Then, the posterior density of λ_i, i.e. $f(\lambda_i)$ may be directly simulated via the Gibbs sampler and we have

$$\begin{aligned} \lambda_i &\sim \Gamma\left(x_i + \alpha, 1 + \beta_i\right), \\ \beta_i &\sim \Gamma\left(\alpha + a, \lambda_i + b\right) \ . \end{aligned}$$

6.5.2 The Ising model

In this section we discuss some examples of lattice models. We start from the *Ising model*, which is one of the simplest lattice models, but, nevertheless, can be applied in many areas (e.g. in electromagnetism, see (Cipra 1987, Binder 1988, Linna 2012)).

Let us consider the Ising model on the two-dimensional table $\mathcal{S}$ of size $D \times D$. Each term $s^{(i)}$ of this table $\mathcal{S}$, which is called spin, can take only values $+1$ or -1 ("up" and "down"). These spins interact only with their own neighbours, usually only with the nearest of them, e.g. left, right, up and down direct neighbours of the specified element. The symbol $\sim$ is used to denote this equivalence neighbourhood relation, i.e. if $x \sim y$, then x and y are neighbours. The distribution of the entire table is related to the function (which is known as *energy*)

$$h(\mathcal{S}) = -J \sum_{i \sim j} s^{(i)} s^{(j)} - H \sum_i s^{(i)} \ , \tag{6.17}$$

where J is an interaction constant and H is an external magnetic field. Both of these parameters are known. Then, e.g., we would like to obtain the most

likely configuration of the system, corresponding to the minimum of the energy function $h(\mathcal{S})$.

If $H = 0$, then this model has two degenerate ground states (i.e. states with minimum energy): either all spins are up or all spins are down.

In order to model the behaviour of the Ising model, the Gibbs sampler can be used. For the model described by the energy function (6.17), the full conditionals are given by

$$f_{S^{(i)}|\mathbb{S}^{(-i)}}\left(s^{(i)}|\mathbb{S}^{(-i)}\right) = \\ = \frac{\exp\left(-Hs^{(i)} - Js^{(i)}\sum_{j:i\sim j} s^{(j)}\right)}{\exp\left(-H - J\sum_{j:i\sim j} s^{(j)}\right) + \exp\left(H + J\sum_{j:i\sim j} s^{(j)}\right)} = \\ = \frac{\exp\left(-\left(H + J\sum_{j:i\sim j} s^{(j)}\right)(s^{(i)}+1)\right)}{1+\exp\left(-2\left(H + J\sum_{j:i\sim j} s^{(j)}\right)\right)} . \quad (6.18)$$

Then, each spin in the table $\mathcal{S}$ is sequentially updated, i.e. for each i the value of $s^{(i)} \in \{-1, +1\}$ is generated, using the density (6.18). Because in (6.18) only the neighbours of $s^{(i)}$ should be taken into account for the evaluation of this density, therefore this pdf could be directly applied even for large dimensions of $\mathcal{S}$.

6.5.3 Noise reduction models

An approach similar to the Ising model may be applied in the case of noise reduction in the images. Let us denote a digital image by $\mathcal{S}$. This is equivalent to a two-dimensional table, consisting of pixels, i.e. elements $s^{(i)}$ of the image. These pixels can take binary values (in the case of black and white images), values from some discrete space (in the case of grayscale images) or values from some more complex set (e.g. in the case of colour images). Denote by $\mathcal{A}$ the space of all possible values (known as *colours*) of the pixels. In practical applications, the dimension of $\mathcal{S}$ can be large, e.g. 2048^2. Therefore, any statistical method used to solve this issue become a numerically complex problem.

A graph K describes the structure of the neighbourhoods for all pixels of the image $\mathcal{S}$. This structure can be simpler or more complex – e.g. for the fixed pixel $s^{(i)}$ only four nearest pixels can be considered as its neighbours (as in the case of the Ising model, described in Section 6.5.2) or only eight of them (with additional four pixels at the vertices).

The observed (noisy) image is denoted by $\mathbb{Y}$ and it consists of pixels $y^{(i)}$. As previously, the possible values of each pixel $y^{(i)}$ are given by the space of colours $\mathcal{A}$, but, in general, they may be taken from some other set. The real (unobserved) image is denoted by $\mathbb{X}$ and it consists of pixels $x^{(i)}$.

Let us assume that the registered image $\mathbb{Y}$ is a noised transformation of some real image $\mathbb{X}$. The statistical model of this noise is given by the probability

$$\mathrm{P}\left(\mathbb{Y}|\mathbb{X}\right) = \prod_{i\in\mathcal{S}} \mathrm{P}\left(y^{(i)}\left|x^{(j:i\sim j)}\right.\right) \ , \tag{6.19}$$

where the conditioning at the right hand side of the formula is taken for all pixels, which are neighbours of $y^{(i)}$.

For example, the model of the noise may be described by the Gaussian noise with the direct influence of the pixel $x^{(i)}$ on the pixel $y^{(i)}$, i.e.

$$\mathrm{P}\left(y^{(i)}\left|x^{(i)}\right.\right) \sim N(x^{(i)}, \sigma^2) \ ,$$

where σ^2 is some fixed parameter (i.e. the level of the noise of background).

If the image is black and white one, then it is possible to model the noise by the random "swap" of colours (i.e. when "black" pixel is changing into the "white" one with some probability and vice versa) which could be represented via the formula

$$\mathrm{P}\left(y^{(i)}\left|x^{(i)}\right.\right) = \begin{cases} y^{(i)} = x^{(i)} & \text{with probability } p \\ y^{(i)} \neq x^{(i)} & \text{with probability } 1-p \end{cases} \ , \tag{6.20}$$

where $1-p$ is the probability that the colour "swaps".

Apart from the model of noise, also the prior probability distribution of pixels of the real image $\mathbb{X}$ should be introduced. For example, it may be assumed that this distribution is described by the *Potts model*. Its simplest case is modelled by the probability

$$\mathrm{P}(\mathbb{X}) \propto \prod_{i\in\mathcal{S}} \exp\left(-\beta \sum_{i\sim j} \mathbb{1}\left(x^{(i)} = x^{(j)}\right)\right) \ , \tag{6.21}$$

where $i \sim j$ denotes all neighbours of the pixel $x^{(i)}$, described by the previously introduced graph K. In the case of probability given by (6.21), the pixels which have the same colour in some neighbourhood are the most possible configuration.

The posterior probability is equal to

$$\mathrm{P}\left(\mathbb{X}|\mathbb{Y}\right) \propto \mathrm{P}(\mathbb{X})\,\mathrm{P}\left(\mathbb{Y}|\mathbb{X}\right) \ .$$

If we apply the models (6.19) and (6.21), it leads to the full conditional probability

$$\mathrm{P}\left(x^{(i)}\left|\boldsymbol{x}^{(-i)}, \mathbb{Y}\right.\right) \propto \exp\left(-\beta \sum_{i\sim j, i\neq j} \mathbb{1}\left(x^{(i)} = x^{(j)}\right)\right) \prod_{i\sim j} \mathrm{P}\left(y^{(j)}\left|x^{(k:j\sim k)}\right.\right) \ . \tag{6.22}$$

It follows from the assumptions of the model that if we are looking for the real, denoised image $\mathbb{X}$, then we are interested in the "most probably value" of the

distribution. Therefore, if we could sample from the full conditional probabilities (6.22), it becomes possible to use the multi-stage Gibbs sampler in a similar way as in the case of the Ising model. For example, if the neighbourhood relation for the pixel i is defined by its eight closest neighbours (four pixels at the vertices and four pixels at the edges) and the model of the noise is given by the random "swaps" as in (6.20), then we get

$$\mathrm{P}\left(x^{(i)} \left| \boldsymbol{x}^{(-i)}, \mathbb{Y}\right.\right) \propto \exp\left(-\beta \sum_{i\sim j, i\neq j} \mathbb{1}\left(x^{(i)} = x^{(j)}\right) + \ln p\ \mathbb{1}\left(x^{(i)} = x^{(j)}\right) + \right.$$
$$\left. + \ln(1-p)\ \mathbb{1}\left(x^{(i)} \neq x^{(j)}\right)\right) . \quad (6.23)$$

Because in this case the neighbourhood relation is restricted only to eight other pixels, then the normalizing constant for the probability (6.23) may be easily calculated.

The case of the Potts model, described by (6.21), is closely related to black and white images. If the image has some "real" colours or is given as a grayscale image, then prior probability can be modelled as

$$\mathrm{P}(\mathbb{X}) \propto \prod_{i\in\mathcal{S}} \exp\left(-\beta \sum_{i\sim j} \psi\left(x^{(i)} - x^{(j)}\right)\right) ,$$

where $\psi(.)$ is some symmetric function (see, e.g. (Besag 1986, Geman & MacClure 1987)). However, the relevant full conditional density can be too complex for direct sampling from it. Then, introduction of an additional step of the MH algorithm may be necessary. Such approach leads to hybrid MCMC models (see Section 6.4).

6.6 Necessity of convergence diagnosis

As it was mentioned previously, the main advantage of the MCMC method compared to the Monte Carlo approach is that there is no necessity for direct sampling from the target density $f(x)$. Instead, to generate the output, almost any instrumental density $g(y|x)$ can be selected (in the case of the MH algorithm, see Section 6.2) or full conditional densities $f_{X^{(j)}|\mathbb{X}^{(-j)}}\left(.|\boldsymbol{x}^{(-j)}\right)$ (for the Gibbs sampler, see Section 6.3) can be applied.

The ergodicity theorems, considered in Section 3.3, ensure that the Markov chain generated via the MCMC method converges to the appropriate expected value. However, the speed and quality of this convergence depend on the starting value X_0 of this chain and the number of simulations n. Of course, the chain should "forget" as soon as possible the initial value (in order to not introduce the additional bias for the final estimator) and, additionally, some necessary number of simulations should be performed (in order to assess a fixed level of

the error). These two problems are the most significant disadvantages of the MCMC methods. In the case of the MCMC approach, the most important tools for the Monte Carlo methods – the Central Limit Theorem and the Strong Law of Large Numbers – cannot be directly applied, because of dependencies among random variables in the Markov chain.

There are some theorems, which achieve the same purposes as CLT and SLLN and which are appropriate for Markov Chain Monte Carlo methods. However, usually, the assumptions of these theorems are very difficult to verify in practice. For example, information about covariance between the random variables for the subsequent steps in the Markov chain is necessary. Yet, estimation of this measure is a numerically complex problem and may require additional simulation steps.

Therefore, in the case of Markov Chain Monte Carlo methods, special *convergence diagnosis* is necessary for answering the main question: when the obtained estimator is "close enough" to the approximated value? In this section some solutions to this problem will be discussed.

6.7 Approaches to convergence diagnosis

The theorems, introduced in Sections 6.2.2 and 6.3, concerning convergence of various MCMC methods are necessary as mathematical proofs of validity of these approaches. But they cannot be directly applied as methods of controlling the chain produced by the given algorithm, i.e. they do not serve to provide the stopping rule meant to guarantee that the number of steps of the generated Markov chain is sufficient.

Therefore, we should consider various methods, which may be helpful in analysing the convergence of the given realization of the Markov chain. These methods are known as *convergence diagnosis* and their detailed review may be found in, e.g., (Robert & Casella 2004).

As it was mentioned in Section 6.6, two issues arise, when we state the problem of the quality of approximation for the considered value, e.g. the expected value $\mathbb{E}_f h(X)$, if this approximation is based on the output of the MCMC method. These issues can be summarized in the following questions:

1. There is a statistical dependency between the subsequent random variables in the Markov chain. Therefore, the obtained estimator may be influenced by the starting value of this chain. Then, the first question is: how many iterations from the beginning of the chain should be deleted to ensure that the estimator is unbiased?
2. The error of the obtained estimator depends on the number of simulations. The greater the number of iterations, the lower this error. Therefore, it is necessary to introduce some stopping rule for the chain. Then, the second question is: when we should finish the simulations?

Another issue with the convergence diagnosis is related to the problem of comparing the various methods, which give answers to the questions mentioned above. Some of these approaches are strictly theoretical ones, others are heuristic (see, e.g., (Romaniuk 2007, Romaniuk 2008) for a more detailed discussion). Additionally, they can measure different things (e.g. use different statistical measures or other stopping conditions), they are based on different simulation algorithms (e.g. in some cases parallel simulations of multiple Markov chains are necessary), etc. Therefore, it is possible that these methods give various answers, i.e. different stopping rules. In practice, the convergence of the considered MC should be checked using the whole set of convergence diagnosis methods, not just a single approach.

6.7.1 Convergence to the stationary distribution

The first considered criterion is related to the convergence of the Markov chain $X_0, X_1, \ldots$ to its stationary distribution. As it was noted, such chain is started from some initial value X_0. Therefore, if this chain is "close enough" to its target density $f(.)$, then the starting value X_0 should not influence the obtained final approximation. This "close enough" condition has various meanings, because different measures can be applied. For example, total variation difference

$$\|\mathcal{K}_X^n(x,.) - f(.)\|_{\mathrm{TV}}$$

between the n-th step transition kernel and the stationary distribution is used. Some theoretical results in this case are known (see (Jones & Hobert 2001, Robert & Casella 2004)), but the calculations necessary to obtain the analytical results may be very difficult.

Another approach is to apply some standard nonparametric test, such as the Kolmogorov-Smirnov test to check the stationarity condition of a single output of the Markov chain. The term "stationarity" is related to the simple conclusion that if $X_i \sim \pi_X$, where π_X is the stationary distribution of the MC, then also $X_{i+1} \sim \pi_X$. The same applies for the arbitrary steps X_{t_1} and X_{t_2} if the chain is in its stationary regime. Then, having the sample $X_0, X_1, \ldots, X_n$, it is possible to compare the distributions of two halves of this sample, $X_1, \ldots, X_{n/2}$ and $X_{n/2+1}, \ldots, X_n$. Because, usually, the nonparametric tests are devised for *iid* samples, then some correction for the correlation between the dependent values of the Markov chain is necessary. This can be done by the introduction of a batch size k, leading to the construction of two "more independent" samples. If only some values of the halves mentioned are used, we obtain subsamples $X_1, X_{k+1}, \ldots$ and $X_{n/2+1}, X_{n/2+1+k}, \ldots$ denoted further by $V_1, V_2, \ldots$ and $V_1^{'}, V_2^{'}, \ldots$, respectively. It should be noted that this subsampling mechanism leads to some loss of information, because only a part of the generated values is used to check this criterion.

Then the *Kolmogorov-Smirnov statistics* is applied

$$T_{\mathrm{KS}} = \frac{1}{M} \sup_x \left| \sum_{i=1}^{M} \left(\mathbb{1}(V_i \in (0,x)) - \mathbb{1}(V_i^{'} \in (0,x)) \right) \right| , \tag{6.24}$$

where M is the maximum number of values for the subsamples V_i and $V_i^{'}$. It is possible to evaluate the density of the statistics (6.24), though only in the asymptotic case. Under the stationarity assumption, as $M \to \infty$, the limiting distribution of the statistics $\sqrt{M}T_{\mathrm{KS}}$ has the cdf given by

$$D_{\sqrt{M}T_{\mathrm{KS}}}(x) = 1 - \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2x^2} ,$$

which can be easily approximated by a finite sum. But the exact solution of this statistical test is quite difficult to obtain because of the correlations between the values of the Markov chain and the influence of the subsampling mechanism. Another approach is to draw a graph of $\sqrt{M}T_{\mathrm{KS}}$ as a function of the length of the subsample and to check visually for a stable distribution around small values.

Convergence can also be verified by checking if the generated chain $X_0, X_1, \ldots$ "visited" all the states of the space state $\mathcal{S}$. In order to find how much of the support of the target distribution is explored by the chain, the integral

$$\int_{\mathcal{S}} f(x)dx \tag{6.25}$$

is evaluated. The value of the integral (6.25) should be, of course, close to one. Philippe & Robert (2001) proposed a solution based on Riemann sums (compare with Section 5.2.4). When $f(x)$ is a one-dimensional density and $X_{k:n}$ are the order statistics based on the chain $X_0, X_1, \ldots$, then

$$\sum_{k=0}^{n-1} (X_{k+1:n} - X_{k:n})\, f(X_k) \tag{6.26}$$

converges to one, even when the random variables are not generated from the density $f(x)$. Then, if the chain fails to explore some part of the support of $f(x)$, the approximation (6.26) should warn us about this problem by providing an evaluation of the mass of the previously explored region. However, this method is not so straightforward in the multidimensional case.

6.7.2 Convergence of the average

When the simulated values are "close enough" to the stationary distribution, some starting values should be deleted. The number of these deleted steps can be based on the criteria considered in Section 6.7.1 or it could be given as a fixed part of all simulated values (e.g. 10% or 20% of the total number of simulations).

Then, the next step and the second question in the convergence analysis should be taken into account.

The problem of distance between the considered MCMC estimator (e.g. empirical average) and the approximated value (e.g. the integral, which is equal to the expected value $\mathbb{E}_f h(X)$) plays a very important role in the convergence analysis. As previously, the term "distance" has various meanings, because different measures of error and norms may be applied.

Graphical methods, based on drawing the outputs of the MCMC method, can detect some obvious problems of convergence of the empirical average. Yu & Mykland (1998) proposed to use the cumulative sums (CUSUM), drawing the partial differences

$$C_n(i) = \sum_{k=1}^{i} (h(X_k) - S_n) \tag{6.27}$$

for $i = 1, \ldots, n$, where

$$S_n = \frac{1}{n} \sum_{k=1}^{n} h(X_k) \ .$$

As noted by the authors of this approach, when the mixing of the chain is high (i.e. the chain explores the support of the stationary distribution very fast), the graph of $C_n(.)$ is highly irregular and concentrated around 0. To the contrary, in the case of a slowly mixing chain (i.e. when the chain slowly explores its state space), the graph is regular with long excursions away from 0.

However, this approach has some important disadvantages. Firstly, it is based on a single chain. Therefore, it is almost impossible to detect the existence of other possible modes of $f(.)$ or of other unexplored regions of the state space. Secondly, this method is strictly heuristic and subjective. It is based only on the intuition of the observer. But there are some general advices in this case, like comparing with other, known types of time series.

Another method, proposed by Robert (1995) is to simultaneously use several estimators of $\mathbb{E}_f h(X)$, which are based on the same chain. When all of these estimator have the same value (up to a given precision), convergence can be stated. Apart from the "standard" estimator given by (5.8), the conditional (i.e. *Rao-Blackwellized*) estimator (compare with (6.15)) is used. This second estimator can be constructed, e.g., in the case of the Gibbs sampler (see (Robert & Casella 2004)) and is given by

$$\mathbb{E}_f^{\mathrm{RB}} h(X) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_f (h(X_i)|Z_i) \ , \tag{6.28}$$

where Z_i are some additional random variables, e.g. in the case of the MH algorithm they can be identified with the probabilities $\mathrm{P}(X_i = Y \,|Y_0, Y_1, \ldots, Y_n)$ (see Section 6.2.3 for additional details).

The next estimator is provided by the *importance sampling* method (compare to the approach described in Section 5.3.1). If the density $f(x)$ is known up to a constant, the importance sampling estimator is given by

$$\mathbb{E}_f^{\mathrm{IS}} h(X) = \frac{1}{n} \sum_{i=1}^{n} w_i h(X_i) \ , \tag{6.29}$$

where

$$w_i \propto \frac{f(X_i)}{g_i(X_i)}$$

and $g_i(X_i)$ is the true density used to simulate the value X_i. It should be noted that importance sampling removes the correlations between the terms in the sum at the right hand side of (6.29) (see (Robert & Casella 2004) for further details). Therefore, more "classical" convergence control methods can also be applied in this case.

Another estimator of $\mathbb{E}_f h(X)$ can be obtained via the *Riemann approximation* (compare with the method described in Section 5.2.4 and formula (6.26)). It is given by

$$\mathbb{E}_f^{\mathrm{R}} h(X) = \sum_{i=0}^{n-1} \left(X_{i+1:n} - X_{i:n}\right) h(X_{i:n}) f(X_{i:n}) \ , \tag{6.30}$$

where $X_{0:n}, X_{1:n}, \ldots$ denotes the order statistics of the chain $X_0, X_1, \ldots$. It should be noted that multidimensional extensions of such an approach have the quality decreasing with the number of dimensions, therefore, this estimator should rather be applied in the one-dimensional case.

Then, all of the obtained estimators are compared, e.g. the graph of their convergence is plotted or differences between their values are evaluated. Of course, this approach has also some important disadvantages. Firstly, it may be inapplicable in some cases, e.g. the Riemann approximation may be not available for the considered example. Secondly, it is intrinsically conservative, i.e. the value of only one estimator may indicate that additional simulations are required in distinction from the others. Lastly, it is based only on a single chain, therefore some important problems with convergence might be neglected or improper conclusions may be drawn.

An alternative convergence diagnosis was proposed by Gelman & Rubin (1992). In their approach $K > 1$ chains are simultaneously generated. Let $X_i^{(k)}$ denote the i-th element of the k-th chain for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. Then, the estimator of the between-chain variances is equal to

$$\overline{\mathrm{Var}}_B = \frac{1}{K} \sum_{k=1}^{K} \left(\overline{h(X^{(k)})} - \overline{h(X)} \right)^2 \ ,$$

where

$$\overline{h(X^{(k)})} = \frac{1}{n} \sum_{i=1}^{n} h\left(X_i^{(k)}\right) \ , \ \overline{h(X)} = \frac{1}{K} \sum_{k=1}^{K} \overline{h\left(X^{(k)}\right)} \ .$$

The estimator of within-chain variances is given by

$$\overline{\mathrm{Var}}_W = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{n}\sum_{i=1}^{n}\left(h\left(X_i^{(k)}\right) - \overline{h\left(X^k\right)}\right)^2\right) . \tag{6.31}$$

And the additional estimator of variance

$$\overline{\mathrm{Var}}^{+} = \frac{n}{n+1}\overline{\mathrm{Var}}_W + \overline{\mathrm{Var}}_B \tag{6.32}$$

is introduced. In order to diagnose the convergence, the estimators $\overline{\mathrm{Var}}^{+}$ and $\overline{\mathrm{Var}}_W$ should be compared, because they are asymptotically equivalent. Therefore, if they are close in value, convergence can be stated.

7

The generalized method of catastrophe bonds pricing

Piotr Nowak

7.1 Introduction and bibliographical overview

Nowadays, natural catastrophes occur more frequently than before. Additionally, they hit densely populated areas (e.g., tsunami in Japan (2011)). This results in high values of damages. One can mention huge losses caused by Hurricane Andrew in 1992. In Poland and in whole Central Europe devastating floods (e.g., the 1997 Oder Flood) are a significant problem.

Natural disasters have negative impact on financial stability of insurers, since classical insurance mechanisms are inadequate to deal with rare catastrophic events, which cause large-scale damages. Therefore, new insurance-linked financial instruments are introduced. *Catastrophe bonds* (also called *cat bonds*), which are examples of such instruments, are issued as means of catastrophic risk transfer to financial markets. A hedger, sponsoring the catastrophe bond, is usually an insurance or reinsurance company. It creates a special purpose entity (often called SPV) to manage the cash flows related to the catastrophe bond. The payoff received by the cat bond holder depends on a precisely specified random variable, called *triggering point*, which is connected, inter alia, with the occurrence of a catastrophic event (such as flood, earthquake, hurricane, windstorm) in a specified place and time, its real parameters, hedger's actual losses or actual losses to the insurance industry as a whole, caused by a natural disaster. Moreover, the payoff usually depends on interest rates. The hedger pays an insurance premium in exchange for a pre-specified coverage, if the triggering point occurs. The bondholders purchase cat bonds for cash. Both premium and cash flows are directed to the special purpose entity, which issues the catastrophe bonds and purchases high grade securities in order to satisfy future demands. Bondholders receive full payment if the triggering point does not occur. Otherwise, the payment for bondholders (i.e., the face value and/or interest) is partly (or fully) forgiven and the entity compensates the hedger instead.

Some further detailed remarks concerning historical and practical aspects of catastrophe bonds can be found in Section 8.1. In this chapter we apply methods of stochastic analysis and financial mathematics to find the valuation expression for catastrophe bonds. Our aim is to propose a generalized catastrophe bond pricing formula, which can be applied to many types of such instruments. In

particular, our approach can be used for various types of payoff functions. Moreover, the possibility of choice of an appropriate model of the spot interest rate makes the proposed pricing method more universal from theoretical and practical points of view.

The amount of papers, concerning catastrophe bonds and their valuation, is relatively low. The authors of (Canabarro, Finkemeier, Anderson & Bendimerad 2000), apart from a discussion concerning stochastic catastrophe modelling, present a comparative analysis of catastrophe bonds and other assets. Additionally, the authors noted absence of correlation between natural catastrophe risk and financial market risk. In (Bodoff & Gan 2009), an analysis of empirical data was reported, and the issuing price of cat bonds was described as a linear function of expected loss with parameters that depend on peril and zone. The behavioural finance method was used in (Kai, Zhong-ying & Shang-zhi 2007). The authors mentioned its applicability potential in China. Wang (2004) applied the probability transform to extend the Sharpe ratio concept to evaluate the risk-adjusted performance of cat bonds.

There are also approaches involving discrete-time stochastic processes: (Cox & Pedersen 2000) within the framework of representative agent equilibrium and (Reshetar 2008), where the authors assumed that payoff functions are linked to catastrophic property losses and catastrophic mortality.

Stochastic models in continuous time were considered in (Baryshnikov, Mayo & Taylor 1998), where compound Poisson processes were used to incorporate various characteristics of the catastrophe process. The drawback of this approach is the assumption that the arbitrage and "the real-life" measure coincide. However, although no analytical pricing formula was obtained, a method of reducing the pricing problem to an integral PDE was proposed. In (Albrecher, Hartinger & Tichy 2008) the claim index was modelled by the doubly stochastic compound Poisson process, reporting lags of the claims having occurred were incorporated into the model and the properties of catastrophe bonds' prices were analysed with application of Quasi-Monte Carlo methods. In turn, Egamia & Young (2008) used indifference pricing method for valuation of structured cat bonds. Vaugirard (2003) applied the arbitrage approach for cat bonds pricing. He addressed the problem of incompleteness of the financial market, caused by catastrophic risk, and non-traded insurance-linked underlyings in the Merton's manner (see (Merton 1976)). Lin, Shyu & Chang (2008) applied the Markov-modulated Poisson process for the description of the arrival rate of natural catastrophes. In (Nowak & Romaniuk 2013*b*) we used the approach similar to the one proposed by Vaugirard. However, in contradistinction to (Vaugirard 2003), where risk index in the form of geometric jump-diffusion process was considered, we defined a cat bond payoff function dependent on cumulative catastrophe losses, modelled by compound Poisson process. We proved catastrophe bond pricing formula at moment 0, assuming diffusion model of the risk-free interest rate and a stepwise and a piecewise linear cat bond payoff functions. Moreover, we conducted Monte Carlo simulations to analyse the behaviour of the valu-

ation expression. Finally, in the mentioned paper we summarized our earlier approaches (see (Nowak & Romaniuk 2009, Nowak & Romaniuk 2010*a*, Nowak & Romaniuk 2010*d*, Nowak & Romaniuk 2011, Nowak, Romaniuk & Ermolieva 2011)), where special cases of our pricing method were proposed. In the series of papers (Nowak & Romaniuk 2010*c*, Nowak & Romaniuk 2013*c*, Nowak & Romaniuk 2014*b*, Nowak & Romaniuk 2014*c*) we also considered the problem of pricing of the insurance-linked instrument in fuzzy framework (see, e.g., (Zadeh 1965)) for some particular models of catastrophe bonds. We combined there a stochastic and a fuzzy approach to obtain cat bond valuation expressions under uncertainty. A similar approach was applied for the case of European options in (Wu 2004, Nowak & Romaniuk 2010*b*, Nowak 2011, Nowak & Romaniuk 2013*a*, Nowak & Romaniuk 2014*a*).

In this chapter we propose a generalized method of catastrophe bonds pricing. We assume that there is no possibility of arbitrage on the market, investors are neutral toward nature jump risk and replicability of interest rate changes by financial instruments existing on the market. We apply stochastic processes in continuous time and stochastic analysis, in particular, the martingale method of pricing. The contribution of this chapter is threefold. First, we consider catastrophe bonds with a more complex payoff structure than those defined in (Nowak & Romaniuk 2013*b*). Second, we give a full analytical description of the cat bond valuation formulas for three important one-factor affine interest rate models. Third, in comparison with (Nowak & Romaniuk 2013*b*), where we focused our attention on catastrophe bond prices at the moment 0, we describe in a more detailed way the pricing formula at each moment $t \in [0, T]$, where T is the bond expiration date. The payoff structure, defined by us, enables the use of a wide class of functions, describing dependence between values of bondholder's payoff and cumulative catastrophe losses, in particular, the piecewise quadratic functions can be applied. We consider three particular cases of the one-factor affine interest rate models, i.e. the Merton, the Vasicek and the Cox–Ingersoll–Ross model with the above mentioned complex payoff structure of the insurance-linked instruments. We also show basic properties of the cat bonds prices for these models as functions of the risk-free interest rate parameters.

As we noted above, this chapter is devoted to derivation and proof of the generalized catastrophe bonds pricing formula. Section 7.2 contains basic notations, definitions and assumptions concerning the financial market, the filtered probability space and the stochastic processes used for modelling catastrophe bonds. In Section 7.3 we define catastrophe bond payoff structure. Section 7.4, which is the main section of this chapter, contains definition of the class of the spot affine interest rate models used by us and the proof of the generalized catastrophe bonds pricing formula. In Section 7.5 we present the particular forms of the catastrophe bonds valuation expressions for three models of the affine spot interest rate. For definitions of probabilistic and stochastic notions used in this chapter we refer the reader to Chapter 2.

7.2 Basic notations and definitions

We introduce basic notations, definitions and assumptions concerning catastrophic losses, the risk-free interest rate, zero-coupon bonds and financial market. At the beginning, we define stochastic processes describing the dynamics of the spot interest rate and cumulative catastrophe losses.

We use stochastic models in continuous time, assuming $\mathcal{T} = [0, T']$ for $T' > 0$. The date of maturity of the catastrophe bond, denoted by T, is not later than T', i.e. $T \leq T'$. We consider two probability measures: P and Q and we denote by $\mathbb{E}^{\mathrm{P}}$ and $\mathbb{E}^{\mathrm{Q}}$ the expected values with respect to them.

Stochastic processes and random variables, introduced in this section, are defined with respect to probability P. A probability measure Q, equivalent to P, will be defined in the next section.

Let $(W_t)_{t\in[0,T']}$ be a Brownian motion. The process $(W_t)_{t\in[0,T']}$ will be used in the stochastic model of the risk-free interest rate.

Let $(U_i)_{i=1}^{\infty}$ be a sequence of independent identically distributed random variables with finite second moments. For each $i = 1, 2, ...$ U_i equals the value of losses caused by the i-th catastrophic event. Though we focus our attention on the case of the independent catastrophic losses, it is possible to assume, more generally, that there is a dependence between U_i for $i = 1, 2, ...$

For each $t \in [0, T']$, cumulative catastrophe losses until moment t will be described by compound Poisson process $\tilde{N}_t$, defined by the equality

$$\tilde{N}_t = \sum_{i=1}^{N_t} U_i,\ t \in [0, T'],$$

where N_t is a (possibly non-homogeneous) Poisson process with an intensity function κ.

In particular, for each $t \in [0, T']$, the value of the process N_t is equal to the number of catastrophic events until the moment t and the moments of jumps of the process $(N_t)_{t\in[0,T']}$ are treated as moments of catastrophic events. In turn, the heights of jumps of the process $\tilde{N}_t$ are equal to the values of losses caused by catastrophic events.

All the above introduced processes and random variables are defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T']}, \mathrm{P}\right)$. The filtration $(\mathcal{F}_t)_{t\in[0,T']}$ and the additional filtrations $\left(\mathcal{F}_t^0\right)_{t\in[0,T']}$ and $\left(\mathcal{F}_t^1\right)_{t\in[0,T']}$ are: the natural filtration of both processes $(W_t)_{t\in[0,T']}$ and $\left(\tilde{N}_t\right)_{t\in[0,T']}$ in the first case, the natural filtration of $(W_t)_{t\in[0,T']}$ in the second case, and the natural filtration of $\left(\tilde{N}_t\right)_{t\in[0,T']}$ in the third case, augmented to encompass all the null sets from $\mathcal{F} = \mathcal{F}_{T'}$.

All the above defined processes and random variables are independent and the filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T']}, \mathrm{P}\right)$ satisfies usual conditions (see Definition 2.23).

In the remainder of this section we focus our attention on assumptions concerning financial market and financial instruments.

We assume that trading on the market takes place continuously in time, there are no transaction costs, there are no restrictions concerning borrowing and short selling. Moreover, lending and borrowing rates are equal.

We denote by $(B_t)_{t\in[0,T']}$ the banking account, satisfying the following equation:

$$dB_t = r_t B_t dt, \quad B_0 = 1,$$

where $r = (r_t)_{t\in[0,T']}$ is the risk-free spot interest rate, i.e. short-term rate for risk-free borrowing or lending at time t over the infinitesimal time interval $[t, t+dt]$. We assume that r is a specified later diffusion process. For all $t \in [0, T']$

$$B_t = exp\left(\int_0^t r_s ds\right).$$

We also assume that zero-coupon bonds are traded on the market. We denote by $B(t,T)$ the price at time t of zero-coupon bond with the maturity date $T \leq T'$ and with the face value equal to 1.

We price catastrophe bonds under the assumption of absence of arbitrage on the market. Moreover, we assume that i) investors are neutral toward nature jump risk and ii) changes in interest rate r can be replicated by the existing financial instruments (especially zero-coupon bonds). For empirical confirmation of assumption i) we refer the reader e.g. to (Canabarro et al. 2000, Vaugirard 2003).

7.3 Catastrophe bond payoff structure

We begin with the definition of the catastrophe bond payoff structure. In (Vaugirard 2003) a relatively simple form of the cat bond payoff function was applied. The author assumed that if the triggering point does not occur, the bondholder is paid the face value Fv and otherwise the payoff is equal to the face value minus a coefficient in percentage w (called a payoff decrease in further parts of this book), i.e. $Fv(1-w)$.

However, more complex cat bond payoff functions are also possible. Therefore, in this section we propose a generalized payoff structure, whose particular forms can be applied for various types of catastrophe bonds.

We fix a positive integer $n \geq 1$, the face value of catastrophe bond $Fv > 0$ and the maturity date of cat bond $T \in [0, T']$.

We denote by $\mathcal{W}$ the class of sequences of payoff decreases

$$w = (w_1, w_2, ..., w_n),$$

where $0 \leq w_1, w_2, ..., w_n$ and $\sum_{i=1}^{n} w_i \leq 1$. We denote the partial sums of $w \in \mathcal{W}$ by

$$w^{(0)} = 0,\ w^{(k)} = \sum_{i=1}^{k} w_i,\ k = 1, 2, ..., n.$$

Let Φ be the class of sequences

$$\varphi = (\varphi_1, \varphi_2, ..., \varphi_n),$$

satisfying the following conditions

(i) $\varphi_i : [0,1] \to [0,1]$, $i = 1, ..., n$;
(ii) $\varphi_i \in C([0,1])$ and is non-decreasing for each $i = 1, 2, ..., n$;
(iii) $\varphi_i(0) = 0$, $i = 1, ..., n$.

We also consider the following subclasses of Φ

$$\Phi_0 = \{\varphi \in \Phi : \varphi_i \equiv 0 \text{ for } i = 1, 2, ..., n\};$$

$$\Phi_1 = \{\varphi \in \Phi : \varphi_i(1) = 1 \text{ for } i = 1, 2, ..., n\};$$

$$\Phi_{1,l} = \{\varphi \in \Phi : \varphi_i(x) = x \text{ for } x \in [0,1] \text{ and } i = 1, 2, ..., n\};$$

and

$$\Phi_{1,q} = \left\{\varphi \in \Phi : \varphi_i(x) = x^2 \text{ for } x \in [0,1] \text{ and } i = 1, 2, ..., n\right\}.$$

Clearly, $\Phi_{1,l}$ and $\Phi_{1,q}$ are subclasses of Φ_1.

Let $\mathcal{K}$ be the class of sequences

$$K = (K_0, K_1, K_2, ..., K_n),$$

where $0 \le K_0 < K_1 < ... < K_n$.

We define now the catastrophe bond payoff function.

Definition 7.1. *Let $w \in \mathcal{W}$, $\varphi \in \Phi$ and $K \in \mathcal{W}$. Let*

$$f_{w,\varphi,K} : [0, \infty) \to \left[Fv\left(1 - w^{(n)}\right), Fv\right]$$

be a function such that

(i) $f_{w,\varphi,K}|_{[0,K_0]} \equiv Fv$;
(ii) $f_{w,\varphi,K}(x)|_{(K_{i-1},K_i]} = Fv\left(1 - w^{(i-1)} - \varphi_i\left(\frac{x - K_{i-1}}{K_i - K_{i-1}}\right) w_i\right)$, $i = 1, 2, ..., n$;
(iii) $f|_{(K_n,\infty)} \equiv Fv\left(1 - w^{(n)}\right)$.

We denote by $IB(w, \varphi, K)$ the catastrophe bond with the face value Fv, the maturity and the payoff date T, if its payoff function is the random variable $\nu_{w,\varphi,K}$, given by the equality

$$\nu_{w,\varphi,K} = f_{w,\varphi,K}\left(\tilde{N}_T\right).$$

The payoff function $\nu_{w,\varphi,K}$ of $IB(w, \varphi, K)$ is called stepwise (piecewise linear or piecewise quadratic) if $\varphi \in \Phi_0$ ($\varphi \in \Phi_{1,l}$ or $\varphi \in \Phi_{1,q}$).

Remark 7.2. *The catastrophe bond $IB(w, \varphi, K)$ has the following properties:*

1. *The general formula describing the payoff as a function of losses $\tilde{N}_T$ can be written down in the form*

$$\nu_{w,\varphi,K} = Fv\left[1 - \sum_{i=1}^{n} \varphi_i\left(\frac{\tilde{N}_T \wedge K_i - \tilde{N}_T \wedge K_{i-1}}{K_i - K_{i-1}}\right) w_i - \sum_{i=1}^{n} (1 - \varphi_i(1)) w_i I_{\left\{\tilde{N}_T > K_i\right\}}\right]. \tag{7.1}$$

In particular,

$$\nu_{w,\varphi,K} = \begin{cases} Fv\left(1 - \sum_{i=1}^{n} w_i I_{\left\{\tilde{N}_T > K_i\right\}}\right) & \text{for } \varphi \in \Phi_0; \\ Fv\left[1 - \sum_{i=1}^{n} \frac{\tilde{N}_T \wedge K_i - \tilde{N}_T \wedge K_{i-1}}{K_i - K_{i-1}} w_i\right] & \text{for } \varphi \in \Phi_{1,l}; \\ Fv\left[1 - \sum_{i=1}^{n} \left(\frac{\tilde{N}_T \wedge K_i - \tilde{N}_T \wedge K_{i-1}}{K_i - K_{i-1}}\right)^2 w_i\right] & \text{for } \varphi \in \Phi_{1,q}. \end{cases}$$

2. *If the catastrophe does not occur ($\tilde{N}_T \leq K_0$), the bondholder receives the payoff equal to the face value Fv.*
3. *If $\tilde{N}_T > K_n$, the bondholder receives the payoff equal to $Fv(1 - w^{(n)})$.*
4. *If $K_{i-1} < \tilde{N}_T \leq K_i$ for $i = 1, 2, ..., n$, the bondholder receives the payoff equal to*

$$Fv\left(1 - w^{(i-1)} - \varphi_i\left(\frac{\tilde{N}_T - K_{i-1}}{K_i - K_{i-1}}\right) w_i\right).$$

In the case of the stepwise payoff function this payoff is constant and equal to $Fv\left(1 - w^{(i-1)}\right)$ when $\tilde{N}_T \in (K_{i-1}, K_i]$. For $\varphi \in \Phi_{1,l}$ ($\varphi \in \Phi_{1,q}$), the payoff decreases linearly (quadratically) from value $Fv\left(1 - w^{(i-1)}\right)$ to value $Fv\left(1 - w^{(i)}\right)$ when $\tilde{N}_T$ increases in the interval $(K_{i-1}, K_i]$.

By applying the above properties of the catastrophe bond payoff structure, we obtain the following lemma.

Lemma 7.3. *Let $\nu_{w,\varphi,K}$ be the payoff function of $IB(w, \varphi, K)$,*

$$\psi_i = \mathrm{P}(\tilde{N}_T \leq K_i),\ i = 0, 1, 2, ..., n \quad \text{and } \psi_{n+1} = 1.$$

Let

$$e_i = E\left\{\varphi_i\left(\frac{\tilde{N}_T - K_{i-1}}{K_i - K_{i-1}}\right) I_{\left\{K_{i-1} < \tilde{N}_T \leq K_i\right\}}\right\},\ i = 1, 2, ..., n.$$

Then the following equality holds

$$\mathbb{E}^{\mathrm{P}}\, \nu_{w,\varphi,K} = Fv\left\{\psi_0 - \sum_{i=1}^{n} w_i e_i + \sum_{i=1}^{n+1} \left(1 - w^{(i-1)}\right)(\psi_i - \psi_{i-1})\right\}. \tag{7.2}$$

In particular,

$$\mathbb{E}^{\mathrm{P}}\,\nu_{w,\varphi,K} = Fv\left(\psi_0 + \sum_{i=1}^{n+1}\left(1 - w^{(i-1)}\right)(\psi_i - \psi_{i-1})\right)$$

for $\varphi \in \Phi_0$.

Formulas obtained in Lemma 7.3 can be used in the cat bond pricing formula at moment 0 (see Theorem 7.7). They will be applied to numerical considerations presented in Section 8.2.3.

7.4 Catastrophe bond valuation

7.4.1 Spot interest rate models

We shall now describe the class of the spot one-factor affine interest rate models and the probability measure Q, used in the further part of this chapter.

Let the spot interest rate $(r_t)_{t\in[0,T']}$ be a diffusion process satisfying the following equation

$$dr_t = \boldsymbol{\alpha}(r_t)\,dt + \boldsymbol{\sigma}(r_t)\,dW_t. \tag{7.3}$$

We denote by $\mathcal{S}$ the set of all the values which r can take with a strictly positive probability.

Let $\left(\bar{\lambda}_t\right)_{t\in[0,T']}$ be a stochastic process, integrable with respect to the Brownian motion $(W_t)_{t\in[0,T']}$, which satisfies the equality

$$\mathbb{E}^{\mathrm{P}}\exp\left(-\int_0^{T'}\bar{\lambda}_s dW_s - \frac{1}{2}\int_0^{T'}\bar{\lambda}_s^2 ds\right) = 1. \tag{7.4}$$

As it was mentioned in Section 2.4, one of the sufficient conditions for (7.4) is the following inequality:

$$\mathbb{E}^{\mathrm{P}}\left[\exp\left(\frac{1}{2}\int_0^{T'}\bar{\lambda}_t^2 dt\right)\right] < \infty.$$

Let a probability measure Q, equivalent to P, be defined by the Radon–Nikodym derivative

$$\frac{d\,\mathrm{Q}}{d\,\mathrm{P}} = \exp\left(-\int_0^{T'}\bar{\lambda}_s dW_s - \frac{1}{2}\int_0^{T'}\bar{\lambda}_s^2 ds\right)\ \mathrm{P}\text{-a.s.} \tag{7.5}$$

We assume that $\bar{\lambda}_t = \bar{\lambda}(r_t)\,,\ t \in [0,T']$, is the market price of risk process and it has the following interpretation. If the dynamics of $B(t,T), t \in [0,T]$, is described by the stochastic equation

$$dB(t,T) = B(t,T)(\mu_t^T dt + \sigma_t^T dW_t),$$

where $\mu^T = (\mu_t^T)_{t\in[0,T]}$ is the drift and $\sigma^T = (\sigma_t^T)_{t\in[0,T]}$ is the volatility of the bond price process, then the process $\bar{\lambda}$ satisfies the equation

$$\sigma^T \bar{\lambda} = \mu^T - r \quad \mathrm{Q}\otimes dt \ a.e.$$

Since there is no possibility of arbitrage on the market, all bonds (regardless of their maturity time T) have the same market price of risk (see e.g. (Kwok 2008)).

Theorem 2.46 implies that under Q the interest rate process is given by the equation

$$dr_t = \hat{\boldsymbol{\alpha}}(r_t)\, dt + \boldsymbol{\sigma}(r_t)\, dW_t^{\mathrm{Q}}, \tag{7.6}$$

where

$$\hat{\boldsymbol{\alpha}}(r) = \boldsymbol{\alpha}(r) - \boldsymbol{\sigma}(r)\,\bar{\lambda}(r)$$

and W_t^{Q} is Q-Brownian motion.

We consider a subclass of the above described interest rate models, which are one-factor affine models. The general theory of the affine models was developed by Duffie & Kan (1996) while Dai & Singleton (2000) provided their classification. Their popularity follows from their tractability for bond prices and bond option prices. Such popular interest rate models as the *Merton*, the *Vasicek* and the *Cox–Ingersoll–Ross* model are of this type.

A *one-factor affine model of the interest rate* is a time homogeneous diffusion model, given by (7.6), where

$$\hat{\boldsymbol{\alpha}}(r) = \hat{\phi} - \hat{\kappa} r \quad \text{and} \quad \boldsymbol{\sigma}^2(r) = \delta_1 + \delta_2 r \tag{7.7}$$

for constants $\hat{\phi}, \hat{\kappa}, \delta_1, \delta_2$ (see, e.g., (Munk 2011)). It is required that $\delta_1 + \delta_2 r \geq 0$ for all values $r \in \mathcal{S}$.

Though it would be possible to consider the whole class of one-factor affine interest rate models in our cat bond pricing approach, yet in order to avoid technicalities, we focus on the class $\mathfrak{Af}$ of the three mentioned above interest rate models, corresponding to the following settings of parameters $\hat{\phi}, \hat{\kappa}, \delta_1, \delta_2$: i) $\hat{\phi} \in \mathbb{R}, \hat{\kappa} = 0, \delta_1 > 0, \delta_2 = 0$ (the Merton model); ii) $\hat{\phi} > 0, \hat{\kappa} > 0, \delta_1 > 0, \delta_2 = 0$ (the Vasicek model); iii) $\hat{\phi} > 0, \hat{\kappa} > 0, \delta_1 = 0, \delta_2 > 0$ (the Cox–Ingersoll–Ross model).

7.4.2 The pricing formula

We apply a method of pricing similar to the one proposed in (Vaugirard 2003). As we noted, Vaugirard (2003) considered a simple form of catastrophe bond payoff function. The triggering point was defined as the first passage time through a level of losses K of risk index I, driven by a Poisson jump-diffusion process. He assumed that if the triggering point does not occur, the bondholder is paid the face value Fv; and if the triggering point occurs, the payoff is equal to the face

value minus a coefficient in percentage w, i.e. $Fv(1-w)$. Therefore, bondholders were regarded to be in a short position on a one-touch up-and-in digital option on I and, similarly as in the case of options, the martingale method could be used to price catastrophe bonds. Though we do not consider the risk index, we follow these steps of the Vaugirard's pricing method, using conditional expectation with respect to the equivalent martingale measure to obtain the analytical form of the cat bond pricing formula.

At the beginning, we recall the notions of a λ-system and a π-system as well as Dynkin's lemma.

Definition 7.4. *A λ-system is a non-empty family $\mathcal{A}$ of subsets of a set X with the following properties:*

(1) $X \in \mathcal{A}$.
(2) If $A, B \in \mathcal{A}$ *and* $A \subset B$, *then* $B \setminus A \in \mathcal{A}$.
(3) If a sequence $\{A_1, A_2, \ldots\} \subset \mathcal{A}$ *satisfies* $A_n \uparrow A$, *then* $A \in \mathcal{A}$.

A π-system Π is a non-empty family of subsets of a set X, which satisfies the following condition:

$$A, B \in \Pi \Rightarrow A \cap B \in \Pi.$$

We can now formulate Dynkin's lemma.

Lemma 7.5. *If $\mathcal{A}$ is a λ-system and a non-empty family $\Pi \subset \mathcal{A}$ is a π-system, then $\sigma(\Pi) \subset \mathcal{A}$.*

To prove the catastrophe bond pricing formula we will use the following lemma.

Lemma 7.6. *For each $t \in [0, T]$ the following equalities hold:*

a)

$$\mathbb{E}^{\mathrm{Q}}\left(\exp\left(-\int_t^T r_s ds\right)|\mathcal{F}_t\right) = \mathbb{E}^{\mathrm{Q}}\left(\exp\left(-\int_t^T r_s ds\right)|\mathcal{F}_t^0\right);$$

b)

$$\mathbb{E}^{\mathrm{Q}}\left(\nu_{w,\varphi,K}|\mathcal{F}_t\right) = \mathbb{E}^{\mathrm{Q}}\left(\nu_{w,\varphi,K}|\mathcal{F}_t^1\right);$$

c)

$$\mathbb{E}^{\mathrm{Q}}\left(\exp\left(-\int_t^T r_s ds\right)\nu_{w,\varphi,K}|\mathcal{F}_t\right) = \mathbb{E}^{\mathrm{Q}}\left(\exp\left(-\int_t^T r_s ds\right)|\mathcal{F}_t^0\right)\mathbb{E}^{\mathrm{Q}}\left(\nu_{w,\varphi,K}|\mathcal{F}_t^1\right).$$

Proof. Let $t \in [0,T]$. Let

$$\mathcal{F}_t^{0,1} = \{A \cap B : A \in \mathcal{F}_t^0 \text{ and } B \in \mathcal{F}_t^1\}.$$

Since $\mathcal{F}_t^0 = \{A \cap \Omega : A \in \mathcal{F}_t^0\}$ and $\mathcal{F}_t^1 = \{\Omega \cap B : B \in \mathcal{F}_t^1\}$, it follows that $\mathcal{F}_t^0 \cup \mathcal{F}_t^1 \subset \mathcal{F}_t^{0,1}$. Clearly, $\sigma(\mathcal{F}_t^{0,1}) = \mathcal{F}_t$ and $\mathcal{F}_t^{0,1}$ is a π-system.
To shorten notation we introduce the following symbols:

$$c = exp\left(-\int_t^T r_s ds\right), \nu = \nu_{w,\varphi,K},$$
$$\mathbb{E}_t^0(\cdot) = \mathbb{E}^{\mathrm{Q}}\left(\cdot|\mathcal{F}_t^0\right) \text{ and } \mathbb{E}_t^1(\cdot) = \mathbb{E}^{\mathrm{Q}}\left(\cdot|\mathcal{F}_t^1\right).$$

We also introduce the following families of sets:

$$\mathcal{A}_1 = \{A \in \mathcal{F}_t : \mathbb{E}^{\mathrm{Q}}(cI_A) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)I_A)\},$$
$$\mathcal{A}_2 = \{A \in \mathcal{F}_t : \mathbb{E}^{\mathrm{Q}}(\nu I_A) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^1(\nu)I_A)\}$$

and

$$\mathcal{A}_3 = \{A \in \mathcal{F}_t : \mathbb{E}^{\mathrm{Q}}(c\nu I_A) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)\,\mathbb{E}_t^1(\nu)I_A)\}.$$

$\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ are λ-systems. Indeed, $\Omega \in \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$. If $A \subset B$, then $I_{B\setminus A} = I_B - I_A$. For $A, B \in \mathcal{A}_3$

$$\begin{aligned}\mathbb{E}^{\mathrm{Q}}(c\nu I_{B\setminus A}) &= \mathbb{E}^{\mathrm{Q}}(c\nu I_B) - \mathbb{E}^{\mathrm{Q}}(c\nu I_A) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)\,\mathbb{E}_t^1(\nu)I_B)\\ &\quad - \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)\,\mathbb{E}_t^1(\nu)I_A) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)\,\mathbb{E}_t^1(\nu)I_{B\setminus A}).\end{aligned}$$

Similarly, for $A, B \in \mathcal{A}_1$

$$\mathbb{E}^{\mathrm{Q}}(cI_{B\setminus A}) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)I_B) - \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)I_A) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)I_{B\setminus A})$$

and for $A, B \in \mathcal{A}_2$

$$\mathbb{E}^{\mathrm{Q}}(\nu I_{B\setminus A}) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^1(\nu)I_B) - \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^1(\nu)I_A) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^1(\nu)I_{B\setminus A}).$$

If $A_n \uparrow A$, then $I_{A_n} \uparrow I_A$ and condition (3) from Definition 7.4 for $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ follows from Lebesgue's monotone convergence theorem. Moreover, $\mathcal{F}_t^{0,1} \subset \mathcal{A}_1$, $\mathcal{F}_t^{0,1} \subset \mathcal{A}_2$ and $\mathcal{F}_t^{0,1} \subset \mathcal{A}_3$. Indeed, let $C \in \mathcal{F}_t^{0,1}$ and $C = A \cap B$, where $A \in \mathcal{F}_t^0$ and $B \in \mathcal{F}_t^1$. Then, from the independence of $\mathcal{F}_t^0$ and $\mathcal{F}_t^1$, it follows that

$$\begin{aligned}\mathbb{E}^{\mathrm{Q}}(cI_C) &= \mathbb{E}^{\mathrm{Q}}(cI_A)\,\mathbb{E}^{\mathrm{Q}}(I_B) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)I_A)\,\mathbb{E}^{\mathrm{Q}}(I_B)\\ &= \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)I_AI_B) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)I_C).\end{aligned}$$

Furthermore,

$$\begin{aligned}\mathbb{E}^{\mathrm{Q}}(\nu I_C) &= \mathbb{E}^{\mathrm{Q}}(\nu I_B)\,\mathbb{E}^{\mathrm{Q}}(I_A) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^1(\nu)I_B)\,\mathbb{E}^{\mathrm{Q}}(I_A)\\ &= \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^1(\nu)I_C)\end{aligned}$$

and finally,

$$\begin{aligned}\mathbb{E}^{\mathrm{Q}}(c\nu I_C) &= \mathbb{E}^{\mathrm{Q}}(cI_A)\,\mathbb{E}^{\mathrm{Q}}(\nu I_B) = \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)I_A)\,\mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^1(\nu)I_B)\\ &= \mathbb{E}^{\mathrm{Q}}(\mathbb{E}_t^0(c)\,\mathbb{E}_t^1(\nu)I_C).\end{aligned}$$

By applying Dynkin's lemma for the π-system $\mathcal{F}_t^{0,1}$ and λ-systems $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ we obtain equalities a), b) and c), which finishes the proof of the lemma. □

Now we formulate and prove the main theorem concerning the catastrophe bond pricing.

Theorem 7.7. *Let $(r_t)_{t\in[0,T']}$ be a risk-free spot interest rate given by the diffusion process (7.3) and such that after the equivalent change of probability measure, described by the Radon – Nikodym derivative (7.5), it belongs to $\mathfrak{Aff}$. Let $w \in \mathcal{W}$, $\varphi \in \Phi$ and $K \in \mathcal{K}$ and let $IB_{w,\varphi,K}(t)$ be the price at time t, $0 \le t \le T \le T'$, of $IB(w,\varphi,K)$. Then*

$$IB_{w,\varphi,K}(t) = \eta(t,T,r_t,w,\varphi,K),\ 0 \le t \le T, \tag{7.8}$$

where

(i)

$$\eta(t,T,r,w,\varphi,K) = \exp(-a(T-t) - b(T-t)\,r)\,\mathbb{E}^{\mathrm{Q}}\left(\nu_{w,\varphi,K}|\mathcal{F}_t^1\right); \tag{7.9}$$

(ii) functions $a(\tau)$ and $b(\tau)$ satisfy the following system of differential equations:

$$\begin{aligned}&\frac{1}{2}\delta_2 b^2(\tau) + \hat{\kappa}b(\tau) + b'(\tau) - 1 = 0, \quad \tau > 0,\\ &a'(\tau) - \hat{\phi}b(\tau) + \frac{1}{2}\delta_1 b^2(\tau) = 0 \quad \tau > 0\end{aligned} \tag{7.10}$$

with conditions $a(0) = b(0) = 0$.

In particular,

$$\begin{aligned}IB_{w,\varphi,K}(0) &= \eta(0,T,r_0,w,\varphi,K)\\ &= \exp(-a(T) - b(T)\,r_0)\,\mathbb{E}^{\mathrm{P}}\,\nu_{w,\varphi,K}.\end{aligned} \tag{7.11}$$

Proof. According to the assumptions formulated in Subsection 7.4.1, we obtain the probability measure Q equivalent to P. The change of probability measure is described by (7.5). Under the probability measure Q, the spot interest rate dynamics is given by the equation

$$dr_t = \left(\hat{\phi} - \hat{\kappa}r_t\right)dt + \sqrt{\delta_1 + \delta_2 r_t}\,dW_t^{\mathrm{Q}} \tag{7.12}$$

with parameters $\hat{\phi}, \hat{\kappa}, \delta_1, \delta_2$, specified in definition of the class $\mathfrak{Aff}$. For Q, the family $B(t,T),\ 0 \le t \le T \le T'$, is an arbitrage-free family of zero-coupon bond

prices with respect to r, i.e. for each $T \in [0, T']$, $B(T,T) = 1$ and the process of discounted zero-coupon bond price

$$B(t,T)/B_t, \quad t \in [0,T],$$

is a Q-martingale. Then we have the following valuation formula for the zero-coupon bond:

$$B(t,T) = \mathbb{E}^{\mathrm{Q}}\left(e^{-\int_t^T r_u du}|\mathcal{F}_t^0\right), \quad t \in [0,T].$$

From the zero-coupon bond pricing formula for the one-factor affine interest rate models (see, e.g., (Munk 2011)) it follows that

$$\mathbb{E}^{\mathrm{Q}}\left(\exp\left(-\int_t^T r_u du\right)|\mathcal{F}_t^0\right) = \exp\left(-a(T-t) - b(T-t)\, r_t\right), \tag{7.13}$$

where $a(\tau)$ and $b(\tau)$ satisfy the system (7.10). Using assumptions formulated in Section 7.2 and arguments similar as in (Vaugirard 2003), we obtain the analogous equality for the price of catastrophe bond at moment $t \in [0,T]$:

$$IB_{w,\varphi,K}(t) = \mathbb{E}^{\mathrm{Q}}\left(\exp\left(-\int_t^T r_u du\right)\nu_{w,\varphi,K}|\mathcal{F}_t\right). \tag{7.14}$$

From Lemma 7.6 it follows that

$$\begin{aligned}\mathbb{E}^{\mathrm{Q}}&\left(\exp\left(-\int_t^T r_u du\right)\nu_{w,\varphi,K}|\mathcal{F}_t\right)\\ &= \mathbb{E}^{\mathrm{Q}}\left(\exp\left(-\int_t^T r_s ds\right)|\mathcal{F}_t^0\right)\mathbb{E}^{\mathrm{Q}}\left(\nu_{w,\varphi,K}|\mathcal{F}_t^1\right).\end{aligned} \tag{7.15}$$

By applying (7.13) to formula (7.15) we obtain the equality (7.8). Moreover, from formula (7.8) for $t = 0$,

$$IB_{w,\varphi,K}(0) = \exp\left(-a(T) - b(T)\, r_0\right)\mathbb{E}^{\mathrm{Q}}\,\nu_{w,\varphi,K}. \tag{7.16}$$

Since $\nu_{w,\varphi,K}$ and $\frac{d\,\mathrm{Q}}{d\,\mathrm{P}}$ are independent,

$$\begin{aligned}\mathbb{E}^{\mathrm{Q}}\,\nu_{w,\varphi,K} &= \mathbb{E}^{\mathrm{P}}\left(\nu_{w,\varphi,K}\frac{d\,\mathrm{Q}}{d\,\mathrm{P}}\right) = \mathbb{E}^{\mathrm{P}}\left(\nu_{w,\varphi,K}\right)\mathbb{E}^{\mathrm{P}}\left(\frac{d\,\mathrm{Q}}{d\,\mathrm{P}}\right)\\ &= \mathbb{E}^{\mathrm{P}}\,\nu_{w,\varphi,K}.\end{aligned}$$

From this equality, together with (7.16), we obtain (7.11). □

7.5 Catastrophe bonds pricing formulas in particular cases

As we noted earlier, very popular and often used models belonging to the class of affine spot interest rates are the Merton, Vasicek, and Cox–Ingersoll–Ross (CIR for short) models.

The Merton model is described by the following equation

$$dr_t = \varphi dt + \sigma dW_t \tag{7.17}$$

for constants φ and $\sigma > 0$. Its advantage is its relatively simple form. However, its main drawback is the possibility of obtaining negative values of r.

The model of the spot interest rate, introduced by Vasicek, is an Ornstein–Uhlenbeck process of the form

$$dr_t = \kappa\left(\theta - r_t\right) dt + \sigma dW_t, \tag{7.18}$$

where κ, θ and σ are positive constants. The process is mean reverting and the constant θ is called the long-term level of the short rate. The interest rates modelled by (7.18), similarly as in the case of the Merton model, can become negative, which is an undesirable feature.

The spot interest rate process in the CIR model is given by the equation

$$dr_t = \kappa\left(\theta - r_t\right) dt + \sigma\sqrt{r_t} dW_t, \tag{7.19}$$

where κ, θ and σ are positive constants. The CIR model also exhibits mean reversion around the long-term level θ. However, the value space of the process r is the space of non-negative numbers (positive numbers when $2\kappa\theta \geq \sigma^2$).

Theorem 7.8. *Let $w \in \mathcal{W}$, $\varphi \in \Phi$ and $K \in \mathcal{W}$. Let $IB_{w,\varphi,K}(t)$ be the price at time $t \in [0,T]$ of $IB(w,\varphi,K)$.*

(i) If $(r_t)_{t\in[0,T']}$ is described by (7.17) and $\bar{\lambda}(r_t) = \lambda$ is constant, then

$$IB_{w,\varphi,K}(t) = \exp\left(-\left(T-t\right) r_t - \frac{1}{2}\left(\varphi - \lambda\sigma\right)\left(T-t\right)^2 + \frac{1}{6}\sigma^2\left(T-t\right)^3\right) \mathbb{E}^{\mathrm{Q}}\left(\nu_{w,\varphi,K}|\mathcal{F}_t^1\right). \tag{7.20}$$

(ii) For $(r_t)_{t\in[0,T']}$ given by (7.18) and $\bar{\lambda}(r_t) = \lambda$ being constant

$$IB_{w,\varphi,K}(t) = \exp\left(-\frac{1}{\kappa}\left(1 - e^{-\kappa(T-t)}\right) r_t - y_\infty\left[\left(T-t\right) - \frac{1}{\kappa}\left(1 - e^{-\kappa(T-t)}\right)\right] - \frac{\sigma^2}{4\kappa^3}\left(1 - e^{-\kappa(T-t)}\right)^2\right) \mathbb{E}^{\mathrm{Q}}\left(\nu_{w,\varphi,K}|\mathcal{F}_t^1\right), \tag{7.21}$$

where $y_\infty = \theta - \frac{\lambda\sigma}{\kappa} - \frac{\sigma^2}{2\kappa^2}$.

(iii) If $(r_t)_{t\in[0,T']}$ *is modelled by (7.19),* $2\kappa\theta \geq \sigma^2$ *and* $\bar{\lambda}(r_t) = \frac{\lambda\sqrt{r_t}}{\sigma}$*, where* λ *is a constant, then*

$$IB_{w,\varphi,K}(t) = \exp\left(\frac{2\hat{\kappa}\hat{\theta}}{\sigma^2}\left[\ln\left(\frac{\gamma}{\frac{(\hat{\kappa}+\gamma)}{2}\left(e^{\gamma(T-t)}-1\right)+\gamma}\right)+\frac{(\hat{\kappa}+\gamma)}{2}(T-t)\right]\right.$$
$$\left.-\frac{\left(e^{\gamma(T-t)}-1\right)}{\frac{(\hat{\kappa}+\gamma)}{2}\left(e^{\gamma(T-t)}-1\right)+\gamma}r_t\right)\mathbb{E}^{\mathrm{Q}}\left(\nu_{w,\varphi,K}|\mathcal{F}_t^1\right), \tag{7.22}$$

where $\hat{\theta} = \frac{\kappa\theta}{\kappa+\lambda}$ $\hat{\kappa} = \kappa+\lambda$ *and* $\gamma = \sqrt{\hat{\kappa}^2+2\sigma^2}$.

Proof. We apply Theorem 7.7. Since the price of the catastrophe bond is the product of the form (7.9), our approach is similar to the zero-coupon bonds pricing method for the risk-free affine interest rate models (see, e.g., (Munk 2011)). In case (i), under the assumption that $\bar{\lambda}(r_t) = \lambda$, the solution of (7.10) has the form

$$a(\tau) = \frac{1}{2}(\varphi-\lambda\sigma)\tau^2 - \frac{1}{6}\sigma^2\tau^3;$$
$$b(\tau) = \tau.$$

For the Vasicek interest rate model (ii) and $\bar{\lambda}(r_t) = \lambda$, the solution of (7.10) is given by

$$a(\tau) = y_\infty[\tau - b(\tau)] + \frac{\sigma^2}{4\kappa}b^2(\tau);$$
$$b(\tau) = \frac{1}{\kappa}\left(1-e^{-\kappa\tau}\right),$$

where $y_\infty = \theta - \frac{\lambda\sigma}{\kappa} - \frac{\sigma^2}{2\kappa^2}$. In case (iii), under the assumption of $\bar{\lambda}(r_t) = \frac{\lambda\sqrt{r_t}}{\sigma}$, the equality (7.4) holds (see, e.g., (Filipovic, Cheridito & Kimmel 2007)) and

$$a(\tau) = -\frac{2\hat{\kappa}\hat{\theta}}{\sigma^2}\left(\ln(2\gamma)+\frac{1}{2}(\hat{\kappa}+\gamma)\tau - \ln\left[(\hat{\kappa}+\gamma)(e^{\gamma\tau}-1)+2\gamma\right]\right);$$
$$b(\tau) = \frac{2(e^{\gamma\tau}-1)}{(\gamma+\hat{\kappa})(e^{\gamma\tau}-1)+2\gamma}.$$

□

In the remaining part of this section we present some properties of the catastrophe bond price for the affine one-factor interest rate models. The following corollaries, which are the consequences of behaviour of zero-coupon bond prices, show properties of the price of catastrophe bond at moment 0 as a function of chosen parameters of the Merton, Vasicek and CIR interest rate models. Such properties can be important for the construction of portfolio of insurance and financial instruments by an insurance or reinsurance company.

Corollary 7.9. *For all the one-factor affine interest rate models from class* $\mathfrak{Af}$*, the first and the second derivative of* $IB_{w,\varphi,K}(0) = \eta(0,T,r_0,w,\varphi,K)$ *with respect to* r_0 *are given by the formulas:*

$$\frac{\partial \eta(0,T,r_0,w,\varphi,K)}{\partial r_0} = -b(T)\,\eta(0,T,r_0,w,\varphi,K);$$

$$\frac{\partial^2 \eta(0,T,r_0,w,\varphi,K)}{\partial r_0^2} = b^2(T)\,\eta(0,T,r_0,w,\varphi,K).$$

Corollary 7.10. *The following statements hold:*

1. *For the Merton interest rate model the price* $\eta(0,T,r_0,w,\varphi,K)$ *is a convex and decreasing function of* r_0 *and* φ*. Moreover, the price is a convex and increasing function of* λ*.*
2. *In the case of the Vasicek interest rate,* $\eta(0,T,r_0,w,\varphi,K)$ *is a convex and decreasing function of* r_0 *and* θ*. It is also a convex and increasing function of* λ*. Moreover,* $\eta(0,T,r_0,w,\varphi,K)$ *approaches* $\exp(-\theta T)\,\mathbb{E}^{\mathrm{P}}\,\nu_{w,\varphi,K}$ *as* $\kappa \to \infty$*.*
3. *For the Cox–Ingersoll–Ross risk-free interest rate model,* $\eta(0,T,r_0,w,\varphi,K)$ *is a convex and decreasing function of* r_0 *and* θ *and it is a concave and increasing function of* λ*.*

As we noted above, the catastrophe bond valuation expression, proposed and proved in this chapter, has a general form, which takes into account the possibility of a complex instrument's payoff structure and enables to choose an appropriate risk-free interest rate model. Taking into consideration the growth of the catastrophe bond market, further development of our cat bond pricing method is interesting, possible and useful.

8

Simulations and analysis of catastrophe bonds and multi-layer insurance portfolios

Maciej Romaniuk

The natural catastrophic events like floods and hurricanes are source of serious problems for the insurance and reinsurance industry. Even a single catastrophic event can lead to bankruptcy or to lack of reserves of the insurer. Therefore, new financial instruments, which connect the insurance and financial markets might constitute a remedy for these problems. In Chapter 7 the method of pricing of an example of such instrument, called the catastrophe bond (or the cat bond), was considered. Then, in the present chapter, numerical analysis of cat bond prices, based on Monte Carlo simulations will be presented. However, the insurer takes usually into account not only a single financial derivative, but a whole portfolio, consisting of various layers, such as a risk reserve process itself, an issued catastrophe bond, a possible agreement with a reinsurer etc. For examples of such complex approaches, see (Ermolieva, Romaniuk, Fischer & Makowski 2007, Nowak & Romaniuk 2009, Nowak & Romaniuk 2010*d*, Nowak et al. 2011). Therefore, there is a need to analyse such multi-layer insurance portfolio applying, e.g., Monte Carlo simulations. Thus, this problem will be considered in this chapter in a more detailed way.

8.1 Catastrophe bonds

Nowadays, the insurance and reinsurance industry are threatened by risks arising from natural catastrophes, such as hurricanes, earthquakes and floods. Losses from Hurricane Andrew, for example, reached US$30 billion in 1992, while losses from Hurricane Katrina in 2005 are estimated at $40–60 billion (see (Muermann 2008)). It means that a single catastrophic event, like the mentioned hurricanes, may result in damages worth billions of dollars. Therefore, these events can cause problems with reserves for many insurers or even bankruptcy of these enterprises (see (Cummins, Doherty & Lo 2002)). Also the prices of reinsurance policies are strongly connected with appearance of such catastrophes. It causes problems with mitigation of catastrophic risks if the insurers apply the standard, reinsurance approach.

The main problem is that the classic insurance mechanisms are unsuitable for addressing the extreme, infrequent losses caused by the natural catastrophes.

Then, even a single catastrophe reduces the reserves of the insurers or even leads to bankruptcy of these companies. For example, after Hurricane Andrew more than 60 insurance companies fell into insolvency (see (Muermann 2008)). In the case of traditional insurance models (see, e.g. (Borch 1974)) independent risk claims that are frequent, but also small in relation to the value of the whole insurance portfolio are the assumed usual situation. Car crashes or fires in the houses are the examples of such incidents. And the classic strategy of selection of an insurance contract portfolio – the more of (independent) risks, the safer the portfolio – which is based on this assumption, is justified by the law of large numbers and the central limit theorem (see, e.g. (Borch 1974)).

However, catastrophic risks mean that new approaches are needed for building insurance company portfolios. As the sources of losses caused by natural catastrophes, which are infrequent, but rather severe, are strongly dependent on time and location, the traditional portfolio-building strategy can only increase the probability of bankruptcy of the insurer (see (Ermoliev, Ermolieva, McDonald & Norkin 2001)). Additionally, classical insurance mechanisms are often criticized because of serious problems with adverse selection and moral hazard – e.g., hope for governmental help or possession of insurance policy may change people's attitude and lead them to "dangerous" behaviour like building houses in areas threatened by floods, not preventing additional losses etc. To cope with the dramatic impacts of extreme catastrophic events, like floods and hurricanes, an integrated policy that combines mitigation measures with diversified ex ante and ex post financial instruments is required (see (Nowak 1999)).

Development of new financial instruments is an example of a modern approach. As it is known, worldwide financial markets fluctuate by tens of billions of dollars on a daily basis. This is why securitization of losses, i.e., the "packaging" of catastrophic risks into tradable financial assets in the form of the so-called catastrophe derivatives, is useful for dealing with the impacts of extreme natural catastrophes (see, e.g. (Cummins, Doherty & Lo 2002, Freeman & Kunreuther 1997, Froot 2001, Harrington & Niehaus 2003, Nowak & Romaniuk 2009, Nowak & Romaniuk 2010*d*, Nowak et al. 2011, Nowak & Romaniuk 2013*b*, Nowak & Romaniuk 2014*b*)).

The example of such catastrophe-linked security is the catastrophe bond (cat bond, known also as Act-of-God bond, see, e.g., (Cox, Fairchild & Pedersen 2000, Ermolieva et al. 2007, George 1997, Nowak & Romaniuk 2013*b*, O'Brien 1997, Romaniuk & Ermolieva 2005, Vaugirard 2003)). In 1993 the first catastrophe derivatives were introduced by the Chicago Board of Trade (CBoT). These financial derivatives were based on underlying indexes, reflecting the insured property losses due to natural catastrophes, reported by insurance and reinsurance companies.

The payoff received by the cat bond holder is linked to an additional random variable, which is called *triggering point* (see (George 1997)). This triggering point may be connected, for example, with occurrence of a natural catastrophe in a specified region at a fixed time interval, the value of issuer's actual losses

from the catastrophic event (like flood), losses modelled by special software, based on the real parameters of a catastrophe, the whole insurance industry index, the real parameters of a catastrophe (e.g., earthquake magnitude or wind speeds in case of windstorms), or the hybrid index related to modelled losses (see, e.g. (Walker 1997)). Then, the triggering points are related to indemnity triggers, index and parametric triggers of such catastrophic events. For some cat bonds (like Atlas Re II issued for the SCOR Group), the triggering points are the second or even the third event during a fixed period of time. If the triggering point occurs, then the whole structure of payments is changed. The necessary parameters, like the region and the time interval for the catastrophic event, are described in detail for the catastrophe bond. The payments for cat bonds usually depend also on the interest rates, like LIBOR.

One of the earliest examples of catastrophe bonds is the A-1 USAA bond, introduced in 1997. The payoff for this instrument was connected with the hurricane on the East coast of the USA between July 15, 1997 and December 31, 1997. If there had been a hurricane in this region with more than $1 billion losses against USAA, the coupon of the bond would have been lost. The payment of this cat bond equalled LIBOR plus 282 basis points.

Another catastrophe bond was issued in 1997 by Swiss Re to cover earthquake losses. The first cat bond prepared by a non-financial firm was issued in 1999 in order to cover earthquake losses in the Tokyo region for Oriental Land Company, Ltd., the owner of Tokyo Disneyland (see (Vaugirard 2003)). The cat bond market in year 2003 hit a total issuance of $1.73 billion, a 42% increase from the record of the year 2002 which was equal to $1.22 billion (see (McGhee 2004)). Additionally, since 1997, 54 cat bond issues have been completed with total risk limits of almost 8 billion (see (Vaugirard 2003)). To the end of 2004 there were about 65 emissions of cat bonds. Insurance and reinsurance companies issued almost all of these cat bonds, with reinsurers accounted for over 50% of the issuances (see (Vaugirard 2003)). To the end of 2006, 89 cat bonds were issued and 41 of them were prepared by insurers and 43 by reinsurers. In the very year 2006 there were 20 issuances (see (*Ripples Into Waves: The Catastrophe Bond Market at Year-End 2006* 2007)). Moreover, the market of cat bonds is expected to emerge in the future, because of increasing losses from the catastrophic events.

As mentioned earlier, the catastrophe bonds are used to transfer risk from the insurance markets to the financial markets via the mechanism known as securitization of losses. Because of problems with classical insurance models, which appear in the case of the catastrophic events, like dependencies among sources of risks, potentially unlimited losses, problems with adverse selection, as well as moral hazard and reinsurance pricing cycles, new financial instruments like cat bonds are an important alternative. Apart from transferring capital, the liquid catastrophe derivatives market allows the insurance and reinsurance companies to adjust their exposure to natural catastrophic risk dynamically through hedging with those contracts at lower transaction costs.

Many authors emphasize the advantages of using cat bonds by insurers, as compared to other sources of additional capital in the case of catastrophic events. For example, Finken & Laux (2009) indicate that the cat bonds play an important role in the pricing of reinsurance contracts when there is asymmetric information between inside and outside reinsurers about an insurer's risk. Additionally, if the cat bond with a parametric trigger is used, then the adverse selection problem does not arise. On the contrary, there is basis risk, caused by some degree of independence between the cat bond's payoff and the insurer's actual loss. However empirical analysis shows that even in such cases cat bonds can be effectively used to hedge catastrophic risks (see (Cummins, Lalonde & Phillips 2002, Cummins, Lalonde & Phillips 2004, Harrington, Mann & Niehaus 1995)).

In the case of cat bond, the cash flows related to this financial instrument are usually managed by a special tailor-made fund, called a special-purpose vehicle (SPV) or special purpose company (SPC) (see, e.g. (Vaugirard 2003, Lee & Yu 2007)). The hedger (e.g. insurer or reinsurer) pays an insurance premium in exchange for coverage in the situation, when the triggering point occurs. The investors purchase an insurance-linked security for cash. The above mentioned premium and cash flows are directed to SPV (see Figure 8.1), which issues the catastrophe bonds. Usually, SPV purchases safe securities in order to satisfy future possible demands. Investors hold the issued assets whose coupons and/or principal depend on the occurrence of the triggering point, e.g. the catastrophic event. If this event occurs during the specified period, the SPV compensates the insurer and the cash flows for investors are changed (see Figure 8.2). Usually, these flows are lowered, i.e. there is full or partial forgiveness of the repayment of principal and/or interest. However, if the triggering point does not occur, the investors usually receive the full payment (see Figure 8.3). Cat bonds, similarly to other financial derivatives, are often rated by an agency, such as Standard & Poor's, Moody's, or Fitch Ratings.

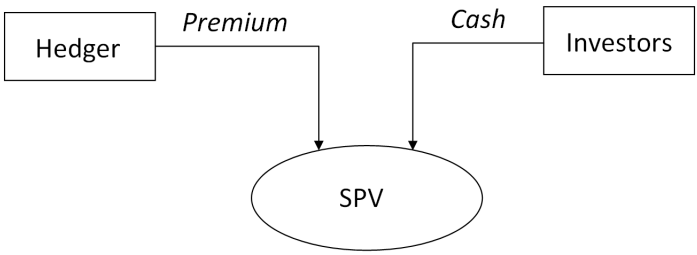

Fig. 8.1. Catastrophe bond: initial cash flows

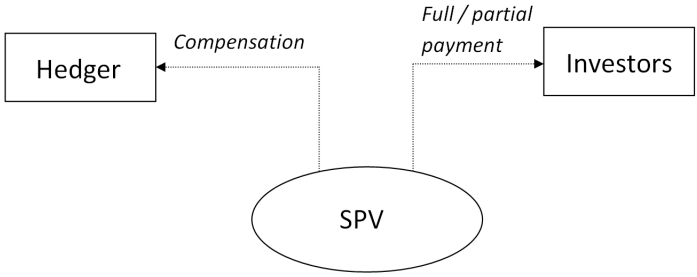

Fig. 8.2. Catastrophe bond: cash flows if the triggering point occurs

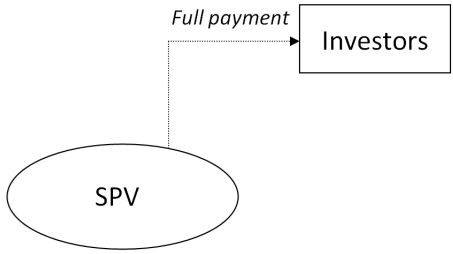

Fig. 8.3. Catastrophe bond: cash flows if the triggering point does not occur

8.2 Simulations of cat bond prices

In order to analyse the behaviour of the cat bond price, described by the formulas from Section 7.4.2, Monte Carlo simulations will be considered in this section. We start from the introduction of the relevant stochastic models of losses and of the risk-free interest rate.

8.2.1 Model of the losses

As previously noted, the triggering point of the cat bond is connected directly (e.g. via value of losses) or indirectly (e.g. via special catastrophe index of the insurance industry) with catastrophic events. Therefore, there is a necessity to apply accurate statistical models of natural catastrophes.

An example of such an approach is considered in (Chernobai, Burnecki, Rachev, Trueck & Weron 2006), where the losses resulting from natural catastrophic events in the United States are taken into account. The distributions of the single catastrophic event and the stochastic processes of the number of events are calibrated in the mentioned paper, based on data from the United States provided by the Property Claim Services (PCS) of the ISO (Insurance Service Office Inc.).

For each catastrophe, the PCS loss estimate represents anticipated industry-wide insurance payments for different property lines of insurance covering (see (Burnecki, Kukla & Weron 2000)). A catastrophic event is denoted as catastrophe if the expected claims exceed a determined threshold. In 1997, the threshold was increased from \$5 million to \$25 million due to changing economic conditions. The losses of magnitudes below this limit are not recorded in the PCS databases. Therefore, there are non-randomly missing data on the left side of the loss distribution of the catastrophic event. This problem is considered in (Chernobai et al. 2006) and the distributions of losses are modelled via conditional distributions, taking into account the effect of the missing data.

In (Chernobai et al. 2006) some of the loss distributions, widely used in statistical analysis, are considered. As noted there, especially the lognormal pdf showed a good fit – both in-sample (i.e. during goodnes-of-fit tests for the estimated parameters of the distribution) and out-of-sample (in forecasting tests). Also the Weibull distribution provided good forecasting results.

In the insurance industry the process of the aggregated losses is described by

$$\tilde{N}_t = \sum_{i=i}^{N_t} U_i \tag{8.1}$$

for $t \geq 0$, where the number of losses is modelled by some stochastic process $N_t \geq 0$ and the values of the single claims are given by the random sequence $U_1, U_2, \ldots$ (see, e.g., (Asmussen & Albrecher 2010)). Usually, the process N_t is homogeneous (abbreviated further as HPP) or non-homogeneous (NHPP) Poisson process (see Section 4.9) and the losses $\{U_i\}_{i=1}$ are *iid* random variables (see, e.g., (Asmussen & Albrecher 2010, Cox, Fairchild & Pedersen 2004, Nowak & Romaniuk 2013*b*)). In the case of the catastrophic events, like hurricanes and floods, the value of the single loss U_i is modelled by the relevant heavy-tailed distribution (see, e.g., (Asmussen & Albrecher 2010)), like the lognormal distribution, considered in (Chernobai et al. 2006).

Apart from the estimation of the parameters of distribution of U_i, the stochastic process N_t used in (8.1) should be also modelled. Then, for the HPP the intensity parameter λ_{HPP}, and for the NHPP the intensity function $\lambda_{\text{NHPP}}(t)$ are estimated. In (Chernobai et al. 2006) the annual seasonality of the events, based on periodogram, is pointed out. Therefore, the cyclic intensity function

$$\lambda_{\text{NHPP}}(t) = a + b2\pi \sin\left(2\pi(t-c)\right) \tag{8.2}$$

with parameters a, b, c was proposed there. The considered estimation procedure is based on fitting the mean value function $\mathbb{E}\, N_t$ to the accumulated quarterly number of losses in PCS data (see also (Burnecki & Weron 2005) for additional details).

In (Ma & Ma 2013), for the similar set of PCS data, adjusted for inflation, more complex cyclic intensity function

$$\lambda_{\text{NHPP}}(t) = a + b\sin^2(t+c) + d\exp\left(\cos\left(\frac{2\pi t}{\omega}\right)\right)$$

with parameters a, b, c, d, ω is considered. These parameters were then fitted to the accumulated annual number of losses using nonlinear least squares method.

In our setting we apply the parameters obtained by Chernobai et al. (2006), based on the previously mentioned real-life data provided by PCS. These data covered the values of losses resulting from natural catastrophes in USA that occurred between 1990 and 1999, specially adjusted using the Consumer Price Index, provided by the U.S. Department of Labour. Using the least squares estimation, the values of parameters for the intensity function $\lambda_{\mathrm{NHPP}}(t)$, given by (8.2) were evaluated in (Chernobai et al. 2006) as

$$a = 30.8750 \ , \ b = 1.6840 \ , \ c = 0.3396 \ . \tag{8.3}$$

These parameters will be further on used in our analysis.

As mentioned previously, Chernobai et al. (2006) considered various probability distributions for the value of the single loss U_i, like exponential, lognormal, Weibull distribution etc. Special attention was paid to the problem of truncation of the recorded data, because a catastrophic event is denoted as a catastrophe if losses caused by this event exceeded some fixed minimum level. As indicated by the authors quoted, the lognormal distribution with the density

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{LN}} t} \exp\left(-\frac{(\log t - \mu_{\mathrm{LN}})^2}{2\sigma_{\mathrm{LN}}^2} \right) \tag{8.4}$$

and the estimated parameters

$$\mu_{\mathrm{LN}} = 17.3570 \ , \ \sigma_{\mathrm{LN}} = 1.7643 \tag{8.5}$$

showed a good fit for the considered real-life data. This pdf surpasses other types of distributions discussed in (Chernobai et al. 2006). Therefore, the lognormal distribution with the mentioned parameters will be used further on in our analysis. However, other types of distributions and other intensity functions could be directly applied to model the process of losses $\tilde{N}_t$ using the Monte Carlo approach, developed in this chapter.

8.2.2 Model of the interest rates

The cat bond price, given by Theorem 7.8, which is a special case of Theorem 7.7, depends also on the selected risk-free interest rate model. As noted in Section 7.5, in our setting, the Vasicek model (see (7.18)), described by

$$dr_t = \kappa \left(\theta - r_t \right) dt + \sigma dW_t \tag{8.6}$$

is applied.

As in the case of the stochastic process of the losses $\tilde{N}_t$ (see Section 8.2.1), the model of r_t fitted to real-life data is used. Chan, Karolyi, Longstaff & Sanders (1992) considered various one-factor interest rate models for the U.S. Treasury

bill yield data. In particular, the parameters of the Vasicek model were evaluated as

$$\kappa = 0.1779 \ , \ \theta = 0.086565 \ , \ \sigma^2 = 0.0004 \tag{8.7}$$

and these values are used in our further analysis.

8.2.3 Analysis of prices

Because of the complex nature of the pricing formula, introduced in Theorem 7.8, which consists of both the discount factor and the expected value as a function of the process of losses, in the following, the Monte Carlo approach is applied to evaluate and analyse the prices of the cat bond.

As it was noted in Section 5.3, the error of the Monte Carlo method strictly depends on the number of simulations. Therefore, in order to minimize its influence, in each experiment one million simulations are conducted (i.e. $n = 100000$).

As the model of the risk-free interest rate, the Vasicek model (see Sections 7.5 and 8.2.2) is applied. Its parameters are given by (8.7) and for the simplicity, its starting value r_0 is set to 0.03.

We assume that $T = 1$, i.e. the maturity time of the considered catastrophe bond is equal to one year. The face value of this bond is one, so the *one monetary unit assumption* is introduced.

Because the *catastrophe* bond is priced, then the underlying model of catastrophic events (8.1) should be also adopted. We apply the NHPP model of the number of catastrophes N_t and the lognormal distribution of the value of the single loss U_i, given by the parameters (8.3) and (8.5), respectively. Of course, other models of catastrophic events and various distributions of the value of the loss can also be used with the Monte Carlo approach.

The payment function for the considered catastrophe bond is piecewise linear, as described in a detailed way in Section 7.1. The triggering points K_i are connected with exceeding the limits given by quantiles of the cumulated value of losses for the previously mentioned models: NHPP (the number of losses) and lognormal distribution (the value of the single loss). Such x-th quantile is denoted further by $Q^{\text{loss}}_{\text{NHPP-LN}}(x)$. The graph of the related quantiles is given with Figure 8.4.

Using the quantiles as the triggering points is intuitively very appealing, because usually the insurer is interested in transferring some value of its "too high" part of the cumulated claims into the financial markets. In the considered setting we take

$$K_0 = Q^{\text{loss}}_{\text{NHPP-LN}}(0.75), K_1 = Q^{\text{loss}}_{\text{NHPP-LN}}(0.85), K_2 = Q^{\text{loss}}_{\text{NHPP-LN}}(0.95)$$

and the values of payments decreases are equal to

$$w_1 = 0.4, w_2 = 0.6 \ .$$

Therefore, e.g. if the cumulated value of losses exceeds the 0.95th quantile, then the holder of the cat bond receives zero as the final payment.

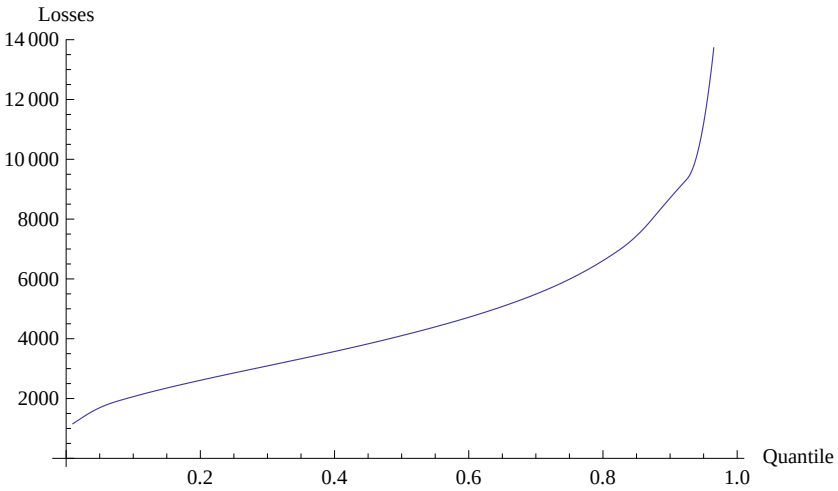

Fig. 8.4. Graph of the quantiles $Q^{\text{loss}}_{\text{NHPP-LN}}(x)$ of the cumulated value of losses (in million \$)

After conducting Monte Carlo simulations, the estimator of the catastrophe bond price is equal to 0.839936. Because of the applied numerical approach, it is possible to analyse the payments of such bond in a more detailed way. For example, the histogram (see Figure 8.5) or descriptive statistics (see Table 8.1) are evaluated. As it may be seen from this histogram, two values of the payments are the most common – zero or one (i.e. the face value of this cat bond).

Measure	Value
Mean	0.869583
Median	1
Mode	1
Standard deviation	0.282263
Skewness	-2.14197
Kurtosis	6.26528

Table 8.1. Descriptive statistics of payments for catastrophe bond

In practical cases, estimation of true parameters of the probability distribution of the single catastrophic loss is a serious problem. If there is an estimation error, i.e. the "real" cdf is not similar to the estimated distribution, then the evaluated cat bond price may be not close to its "appropriate" value. Such issue is illustrated with Figure 8.6, where the dependency between the catastrophe bond price (vertical axis) and the various sets of parameters μ_{LN} and σ_{LN} (horizontal axes) is analysed. As it may be seen, the cat bond price is a decreasing function

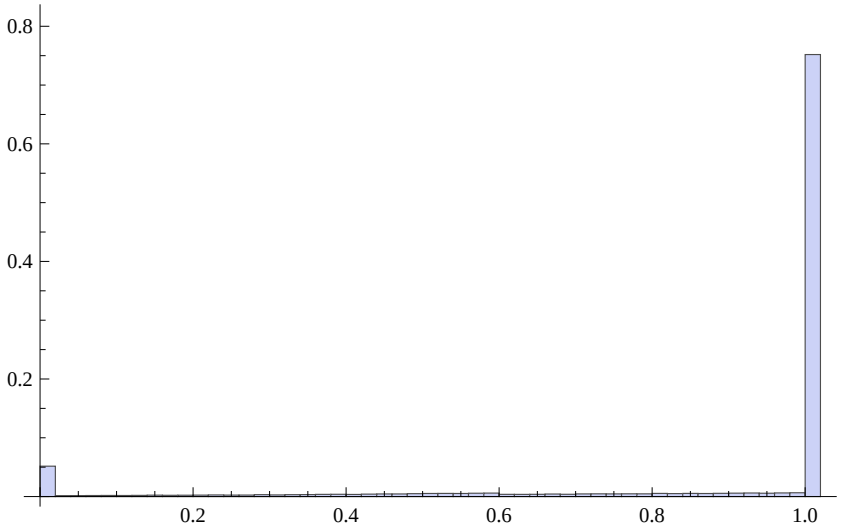

Fig. 8.5. Example of histogram of catastrophe bond payments

of both of these parameters and differences between the prices for various sets of μ_{LN} and σ_{LN} are significant.

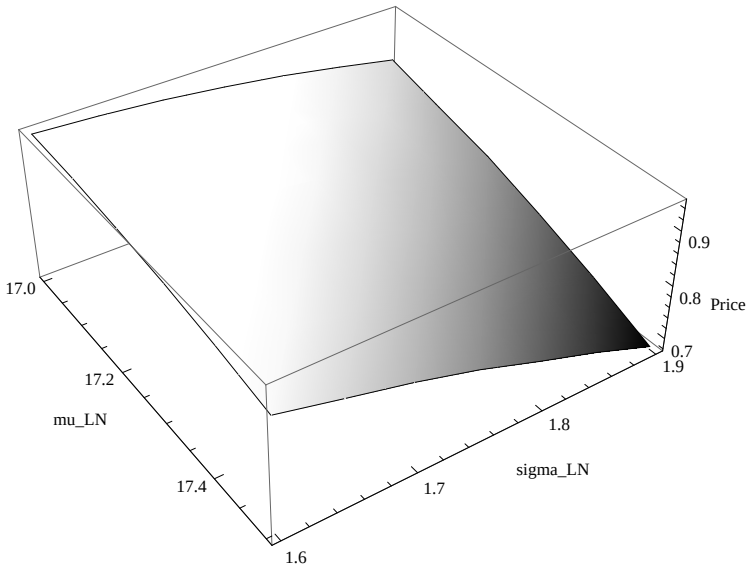

Fig. 8.6. Example of analysis of cat bond prices for different values of μ_{LN} and σ_{LN}

Apart from the possible errors in the estimation of cdf of a single loss, also the intensity function of the process of the number of losses $\lambda(t)$ may be subject to similar problems. As it is seen from Figure 8.7, if the parameter a in the function $\lambda_{\mathrm{NHPP}}(t)$, given by (8.2), is altered, then the catastrophe bond price behaves in a rather non-linear way. Nevertheless, generally speaking, such price is a decreasing function of a.

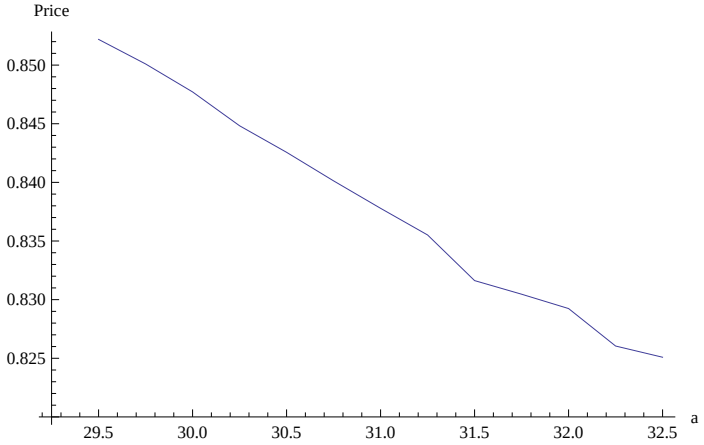

Fig. 8.7. Example of analysis of cat bond prices for different values of a in the intensity function $\lambda_{\mathrm{NHPP}}(t)$

A more complex situation is illustrated with Figure 8.8, where the dependence between the catastrophe bond price and the value of the parameter b for the intensity function $\lambda_{\mathrm{NHPP}}(t)$, given by (8.2), is considered. In this case, because of the cyclic component in (8.2), the price is a highly unstable function of b. Therefore, it is almost impossible to predict the level of error for the cat bond price if there is any problem with estimation of this parameter.

Apart from the influence of the estimation errors, the insurer may be interested also in the analysis of relation between the evaluated price of cat bond and the given set of parameters of the payment function. An example of such an examination is illustrated with Figure 8.9, where the prices for various values of w_2 are found if $w_1 = 0.1$ is set. The other parameters are the same as in the case of the first cat bond considered in this section. Then, the price is almost a linear function of the value of the payment decrease w_2.

Another example of similar analysis is given in Figure 8.10. In this case, various values for only two triggering points K_0 and K_1 are considered. Additionally, $w_1 = 1$ is set to model the bigger impact of decreasing the final payment on the price of the cat bond. As previously, the triggering points are related to the fact of exceeding the quantiles $Q^{\mathrm{loss}}_{\mathrm{NHPP\text{-}LN}}(x)$ of the cu-

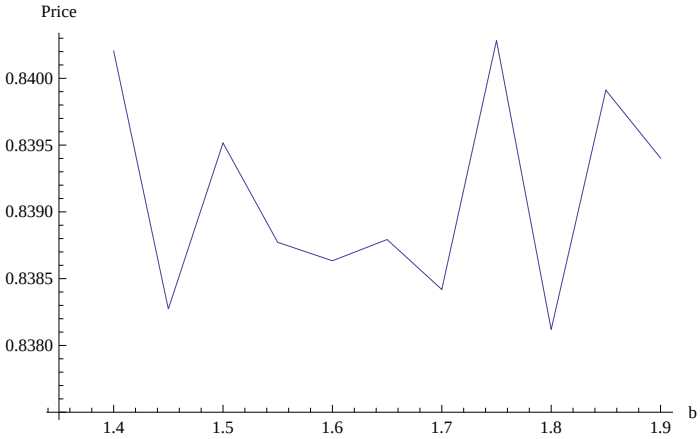

Fig. 8.8. Example of analysis of cat bond prices for different values of b in the intensity function $\lambda_{\text{NHPP}}(t)$

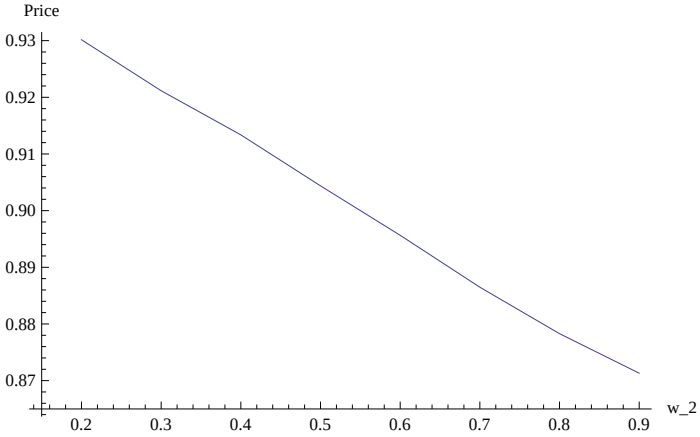

Fig. 8.9. Example of analysis of cat bond prices for different values of w_2

mulated value of losses. Then, $K_0 \in [Q^{\text{loss}}_{\text{NHPP-LN}}(0.4), Q^{\text{loss}}_{\text{NHPP-LN}}(0.65)]$ and $K_1 \in [Q^{\text{loss}}_{\text{NHPP-LN}}(0.7), Q^{\text{loss}}_{\text{NHPP-LN}}(0.95)]$. The other parameters are the same as in the first analysis. The price is a non-linear increasing function of both K_0 and K_1.

8.3 Modelling of the portfolio layers

In previous sections, cat bond prices were analysed using Monte Carlo simulations. Now we apply these results to the problem of the whole insurance portfolio,

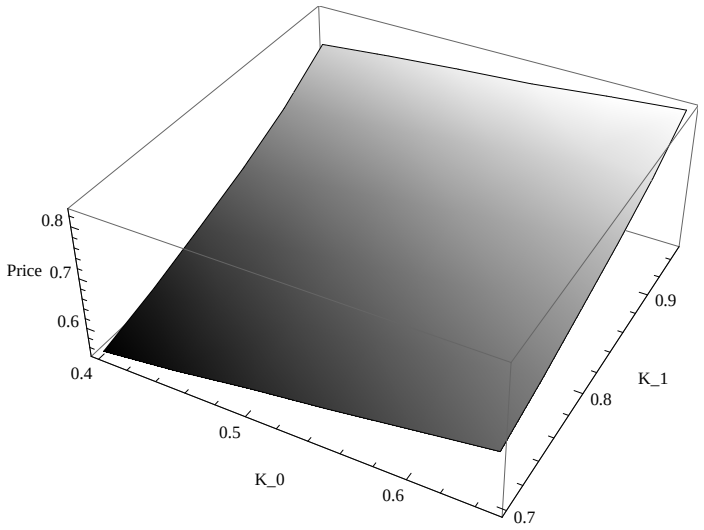

Fig. 8.10. Example of analysis of cat bond prices for different values of $K_0 = Q^{\text{loss}}_{\text{NHPP-LN}}(.)$ and $K_1 = Q^{\text{loss}}_{\text{NHPP-LN}}(.)$

which consists of several instruments (or *layers*). Before the analysis of such portfolio, these complex instrument should be modelled. Therefore, we start from the detailed description of this portfolio and some necessary formulas.

8.3.1 Risk reserve process

In insurance mathematics, a *risk reserve process* (or risk process) $(R_t)_{t\geq 0}$ is introduced as a model of evolution of the financial reserves of the insurance company over time (see, e.g., (Asmussen & Albrecher 2010)). For $t = 0$ we have $R_0 = u$, where u is called the *initial reserve* of the insurer. The number of claims, which are submitted to the insurer in time interval $[0, t]$, is modelled by the stochastic process $N_t \geq 0$. The values of these claims are given by non-negative *iid* random variables $C_1, C_2, \ldots$ (compare this notation with the discussion of the model of losses in Section 8.2.1). In return for protection, the insureds pay a premium to the insurer at the rate p per unit time. Then, the whole risk reserve process is defined as

$$R_t = u + pt - \tilde{C}_t \ , \tag{8.8}$$

where $\tilde{C}_t = \sum_{i=1}^{N_t} C_i$ is the *claim process* (compare with a similar process of losses given by (8.1)).

In the classical insurance approach, if the process of the number of claims N_t is given by the homogeneous Poisson process (HPP), then the premiums are related to the average number of claims per unit time and the expected value of the single claim C_i via the formula

$$p = (1 + \nu_p)\lambda_{\text{HPP}}\,\mathbb{E}\,C_i\ , \tag{8.9}$$

where ν_p is a *safety loading* (or a *security loading*, see, e.g. (Asmussen & Albrecher 2010)) and λ_{HPP} is a constant intensity of the process N_t. A similar notation was introduced in Section 8.2.1. In the literature, the typically assumed values of the safety loading ν_p are approximately 10% – 20%.

If the process N_t is described by the non-homogeneous Poisson process (NHPP) (as in the case of the model considered in Section 8.2.1), then, instead of the risk reserve process (8.8), its modified form

$$R_t = u + \int_0^t p^*(s)ds - \tilde{C}_t$$

with non-constant intensity of premium $p^*(s)$ should be introduced. In this case (see, e.g., (Chernobai et al. 2006) for similar approach) the related *premium function* $p(t)$ is given by the formula similar to (8.9), namely

$$p(t) = \int_0^t p^*(s)ds = (1 + \nu_p)\,\mathbb{E}\,C_i \int_0^t \lambda_{\text{NHPP}}(s)ds\ , \tag{8.10}$$

where $\lambda_{\text{NHPP}}(s)$ is the intensity function of the considered NHPP.

The formula (8.10) with time-dependent function $p(t)$ reflects a situation, when for each t the insurer continually adjusts the premium, taking into account new information about the claims. However, from the practical point of view, the premium is usually treated as "aggregation of average value of claims" and in many cases it is not possible (e.g. because of governmental regulation) or not feasible (e.g. because of necessity of introducing completely new contracts) to change its value too often. Then, the premium is a constant function for some periods of time, e.g. is equal to p till the fixed moment T, when a new value of the premium is evaluated. In such a case, the condition

$$p(T) = pT = (1 + \nu_p)\,\mathbb{E}\,C_i \int_0^T \lambda_{\text{NHPP}}(s)ds \tag{8.11}$$

is fulfilled.

In classical insurance models, the cash flows for the risk process R_t are considered without taking into account the value of money in time, i.e. without introducing the interest rate model to the financial market. Therefore, the main problem in the classical approach is related to calculation of *probability of the ruin* (see, e.g., (Asmussen & Albrecher 2010)). The probability of ultimate ruin (ruin probability with infinite horizon) $\psi(u)$ is the probability that the reserves of the insurer ever drop below zero

$$\psi(u) = \mathrm{P}\left(\inf_{t\geq 0} R_t < 0\right)$$

and the probability of the ruin before time T (ruin probability with finite horizon) is the probability that the reserves drop below zero in time interval $[0, T]$, i.e.

$$\psi(u,T) = \mathrm{P}\left(\inf_{t\in[0,T]} R_t < 0\right) .$$

However, in our setting, the relevant interest rate model for the financial market is considered. Therefore, instead of the classical version of the risk process R_t, its modified form, given by

$$R_T^{\mathrm{FV}} = \mathrm{FV}_T(u) + \mathrm{FV}_T(p(T)) - \mathrm{FV}_T\left(\sum_{i=1}^{N_T} C_i\right) \tag{8.12}$$

is used. From now on, the operator $\mathrm{FV}_T(.)$ denotes the future value of the specified cash flow evaluated at a fixed, final moment T. It should be taken into account that the moments of payments vary for particular parts of (8.12) – the initial reserve u exists at $t = 0$, the premium $p(t)$ is paid continually for all $0 \le t \le T$, the claims C_i are reported when the related events occur. Therefore, the future values for various periods should in this case be evaluated.

8.3.2 Catastrophe bond

As noted in Section 8.1, the traditional insurance approach is not adequate in the case of the catastrophic events like hurricanes and floods. Therefore, in the here modelled insurance portfolio, consideration of additional layers, i.e. supplementary financial and insurance instruments, is necessary. We start from adding first such layer – a catastrophe bond – to this portfolio.

Let us suppose that the insurer, the owner of the considered portfolio, issues a catastrophe bond. The cash flows, related to this instrument, are described in Section 8.1. The insurer pays an insurance premium p_{cb}, which is proportional to the part α_{cb} of the whole price of the single catastrophe bond I_{cb} and to the number of issued cat bonds n_{cb}, i.e.

$$p_{\mathrm{cb}} = \alpha_{\mathrm{cb}} I_{\mathrm{cb}} n_{\mathrm{cb}} . \tag{8.13}$$

This premium is paid always, regardless of the fact whether the issued instrument brings any future payments for the insurer.

The subsequent cash flows depend on the occurrence of the triggering point. If the triggering point takes place, then the insurer receives some payoff. It is equal to the payment function of the single catastrophe bond for the issuer $f_{\mathrm{cb}}^i(.)$ multiplied by the quantity of the issued bonds n_{cb}. In our setting, both the triggering point (or the triggering points) and the payment function $f_{\mathrm{cb}}^i(.)$ depend on the cumulated value of the losses $\tilde{N}_t$, defined by the process (8.1).

Otherwise, if the triggering point does not occur, the insurer receives nothing and $f_{\mathrm{cb}}^i(\tilde{N}_t) = 0$ in this case. It means that the whole payoff – the whole face value of the cat bond – is transferred to the holder of the bond.

Then, the risk process R_t is modified by an additional part, defined by

$$R_t^{\mathrm{cb}} = -p_{\mathrm{cb}} + n_{\mathrm{cb}} f_{\mathrm{cb}}^i(\tilde{N}_t) . \tag{8.14}$$

It should be noted that both the process of the losses $\tilde{N}_t = \sum_{i=1}^{N_t} U_i$ and the process of the claims $\tilde{C}_t = \sum_{i=1}^{N_t} C_i$ are driven by the same Poisson process N_t, but the value of the single claim C_i in some settings can be only a part of the whole value of the relevant loss U_i. For example, if the process $\tilde{N}_t$ and the triggering points are related to the values of losses for the whole insurance industry (as this is the case of the PCS index), then the claims of the single insurer are only some part of the registered losses. Then, the value of the claim C_i is given by

$$C_i = \alpha_{\text{claim}} U_i Z_i \ , \tag{8.15}$$

where $\alpha_{\text{claim}} \in [0,1]$, $Z_i \sim U[c_{\min}, c_{\max}]$, where U_i and Z_i are mutually independent random variables for all i. The parameter α_{claim} describes the share of the considered insurer in the whole insurance industry for the given region, for which the catastrophe bond is issued. If $\alpha_{\text{claim}} = 1$, then there is only one insurer in this region and all losses are (potentially) insured by this enterprise. The additional random variable Z_i may be seen as the random part of the value of losses U_i in the whole value of claims C_i. It is known that sometimes not all the properties are completely insured, or the losses are not reported to the insurer at all, or the specified insurer has various levels of shares in different subregions of the whole region, which may be modelled by some random distribution. As an example of such an approach, the uniform pdf for the interval $[c_{\min}, c_{\max}]$, where $0 \leq c_{\min} < c_{\max} \leq 1$ are some constant parameters, is applied. Therefore, if $Z_i \sim U[0,1]$, then one of the classical non-informative distributions is used in (8.15).

Of course, other probability distributions with support on the interval $[c_{\min}, c_{\max}]$, instead of the uniform one, could be applied for the variable Z_i. Also a simpler influence relation, given by

$$C_i = \alpha_{\text{claim}} U_i \tag{8.16}$$

can be used in the analysis of the portfolio.

It should be noted that the use of such kind of triggering point as the value of the losses of the whole insurance industry or other parametric trigger, leads to the absence of the adverse selection problem. Instead of this, in the case of dependence between the losses U_i and claims C_i, modelled by the formulae (8.15) or (8.16), the basis risk occurs. It is caused by an imperfect correlation between the payoff for the insurer and the claims reported to this enterprise. Because of this risk, the cash flow from the catastrophe bond does not completely match the claims. This leads to exposure of the insurer to uncovered claims (see, e.g., (Finken & Laux 2009, Cummins et al. 2004)). Therefore, additional layers of the portfolio are also important in the hedging of the insurer.

The flows of the cat bond R_t^{cb}, described by the formula (8.14), are usually connected with the maturity of this financial instrument, i.e. time T, as noted in Section 8.1. Additionally, the model of the interest rate is applied in our setting. Then, the introduction of this extra layer into the portfolio leads to a more complex form of the risk process (8.12) given by

$$R_T^{\mathrm{FV}} = \mathrm{FV}_T\left(u - p_{\mathrm{cb}}\right) + \mathrm{FV}_T(P(T)) - \mathrm{FV}_T\left(\tilde{C}_T\right) + n_{\mathrm{cb}} f_{\mathrm{cb}}^i(\tilde{N}_T) \tag{8.17}$$

for the fixed moment T of maturity of the cat bond.

8.3.3 The reinsurer contract

In order to improve their financial stability, the insurers usually transfer some part of their risk caused by high insured losses to other companies, known as reinsurers. A standard reinsurance contract is an *excess-of-loss policy* (see, e.g., (Lee & Yu 2007)). In this case, in exchange for some premium p_{rnsr}, the reinsurance company pays a part of claim value to the insurer if the total value of claims $\tilde{C}_T$ till the moment T exceeds some fixed amount. This minimum limit is called an attachment point A_{rnsr}. The payment from the reinsurer mentioned is also subject to the maximum value of total claims. This maximum limit is a cap level B_{rnsr}. Beyond this level, the reinsurer pays only the fixed value $B_{\mathrm{rnsr}} - A_{\mathrm{rnsr}}$ regardless of the reported claims. Then, the payment function for the insurer is equal to

$$f_{\mathrm{rnsr}}^i\left(\tilde{C}_T\right) = \begin{cases} B_{\mathrm{rnsr}} - A_{\mathrm{rnsr}} & \text{if } \tilde{C}_T \geq B_{\mathrm{rnsr}} \\ \tilde{C}_T - A_{\mathrm{rnsr}} & \text{if } B_{\mathrm{rnsr}} > \tilde{C}_T \geq A_{\mathrm{rnsr}} \\ 0 & \text{otherwise} \end{cases} . \tag{8.18}$$

Therefore, in the classical risk process R_t, defined by (8.8), we have an additional part, related to the reinsurance contract

$$R_t^{\mathrm{rnsr}} = -p_{\mathrm{rnsr}} + f_{\mathrm{rnsr}}^i\left(\tilde{C}_T\right) .$$

This leads us to a more complete form of the risk process R_T^{FV}, defined by (8.17), with two layers in the portfolio

$$R_T^{\mathrm{FV}} = \mathrm{FV}_T\left(u - p_{\mathrm{cb}} - p_{\mathrm{rnsr}}\right) + \mathrm{FV}_T(P(T)) - \mathrm{FV}_T\left(\tilde{C}_T\right) + n_{\mathrm{cb}} f_{\mathrm{cb}}^i(\tilde{N}_T) + f_{\mathrm{rnsr}}^i\left(\tilde{C}_T\right) .$$

Of course, other forms of reinsurance contracts, apart from the excess-of-loss policy, are possible in this setting.

An other issue is related to the calculation of the premium for the reinsurer p_{rnsr}. Usually, as noted by (Finken & Laux 2009), the value of this premium depends on reinsurer's knowledge about inner procedures of the insurer, like insurer's book of business, principles of underwriting and reserving policy, business model and future plans of the enterprise, etc. There is a close business relation between insurers and reinsurers, which leads to asymmetric information in the case when some "new" reinsurer is coming onto the market. For simplicity of our analysis, we assume that the premium p_{rnsr} depends on the expected value of the

present value (denoted further by PV) of the payment function for the insurer f^i_{rnsr}, namely

$$p_{\text{rnsr}} = (1 + \nu_{\text{rnsr}})\, \mathbb{E}\left(\text{PV}\left(f^i_{\text{rnsr}}\left(\tilde{C}_T\right)\right)\right) \ , \tag{8.19}$$

where ν_{rnsr} is the safety loading for the reinsurer. This approach is similar to the security loading ν_p, used for the insured's premium p (compare with formula (8.9)).

8.3.4 Other layers

As noted in (Nowak et al. 2011, Nowak & Romaniuk 2009, Nowak & Romaniuk 2010*d*) other layers may be incorporated into our model of the insurer's portfolio. For example, these layers can be related to governmental help, or foreign help, or additional infusion of capital from another enterprise (e.g. owner of the considered insurer), etc. Such cash flow may be seen as positive payment directed to the insurer, because the previously mentioned types of payoffs actually lower the value of total claims.

We assume that the value of the considered cash flow is proportional to the reported claims $\tilde{C}_T$ till the moment T with a fixed minimum limit A_{hlp}, namely

$$f^i_{\text{hlp}}\left(\tilde{C}_T\right) = \begin{cases} \rho_{\text{hlp}}\tilde{C}_T & \text{if } \tilde{C}_T \geq A_{\text{hlp}} \\ 0 & \text{otherwise} \end{cases} \ ,$$

where ρ_{hlp} is a proportion parameter. However, this additional fund is not always used, but only with some probability, modelled by an independent random variable T_{hlp}. If this binary variable is equal to one (i.e. $T_{\text{hlp}} = 1$, where $\text{P}\left(T_{\text{hlp}} = 1\right) = p_{\text{hlp}}$), then the described cash flow can occur. Otherwise, if $T_{\text{hlp}} = 0$, this additional external help is not available. This leads us to the subsequent part in the risk process R_t, which is equal to

$$R^{\text{hlp}}_T = \mathbb{1}\left(T_{\text{hlp}} = 1\right) f^i_{\text{hlp}}\left(\tilde{C}_T\right) \ .$$

Then, taking into account all of the previously described layers in the insurer's portfolio, the modified risk process R^{FV}_T is given by

$$\begin{aligned} R^{\text{FV}}_T = {} & \text{FV}_T\left(u - p_{\text{cb}} - p_{\text{rnsr}}\right) + \text{FV}_T(P(T)) - \text{FV}_T\left(\tilde{C}_T\right) \\ & + n_{\text{cb}} f^i_{\text{cb}}(\tilde{N}_T) + f^i_{\text{rnsr}}\left(\tilde{C}_T\right) + \mathbb{1}\left(T_{\text{hlp}} = 1\right) f^i_{\text{hlp}}\left(\tilde{C}_T\right) \ . \end{aligned} \tag{8.20}$$

The risk process for the multi-layer insurance portfolio R^{FV}_T is defined by the formula (8.20) in the simplified form, i.e. only when the time t is equal to the fixed moment T. Taking into account the layers described in the previous sections, a similar stochastic process could be formulated for any $0 < t \leq T$, where T is the maturity of the cat bond and the reinsurer contract. But even for

this simplified form (8.20), the process R_T^{FV} is a very complex one. Therefore, in order to analyse its behaviour, like estimation of the probability of the insurer's ruin for the fixed time T or of value of its default, Monte Carlo simulations should be applied.

8.4 Simulation of the underlying models

In order to completely describe and analyse the multi-layer insurance portfolio, introduced in Section 8.3, the relevant model of the interest rates and the model of the process of catastrophes should be selected, calibrated and then used during Monte Carlo simulations.

8.4.1 Model of losses and claims

As presented in Section 8.2.1, the process of the aggregated catastrophe losses $\tilde{N}_t$ has the standard form given by

$$\tilde{N}_t = \sum_{i=i}^{N_t} U_i \ ,$$

where N_t is the process of the number of losses, modelled by the homogeneous (HPP) or the non-homogeneous (NHPP) Poisson process, and the values of losses $U_1, U_2, \ldots$ are *iid* random variables, usually described by some heavy-tailed distributions, e.g. lognormal or Weibull distribution.

In our case we apply the model of losses based on parameters estimated in (Chernobai et al. 2006) for catastrophic data provided by the Property Claim Services of the Insurance Service Office Inc. (see Section 8.2.1 for additional details), namely the cyclic intensity function $\lambda_{\mathrm{NHPP}}(t)$ for NHPP

$$\lambda_{\mathrm{NHPP}}(t) = a + b2\pi \sin\left(2\pi(t-c)\right) \tag{8.21}$$

with the parameters

$$a = 30.8750 \ , \ b = 1.6840 \ , \ c = 0.3396 \ , \tag{8.22}$$

and the lognormal distribution of the single loss U_i, given by the density (8.4) with the parameters

$$\mu_{\mathrm{LN}} = 17.3570 \ , \ \sigma_{\mathrm{LN}} = 1.7643 \ .$$

As discussed in Section 8.3.2, apart from the losses, the risk process R_t^{FV} for the considered portfolio is also influenced by claims based on the catastrophic losses. Then the process of the aggregated claims

$$\tilde{C}_t = \sum_{i=1}^{N_t} C_i \tag{8.23}$$

is described by the same NHPP of the number of events given by the intensity function $\lambda_{\text{NHPP}}(t)$ of the form (8.21) with parameters (8.22). Further to evaluate the value of the single claim C_i for the given value of loss U_i, the formula (8.15) or (8.16) can be used.

There are various possible types of relations between the losses and the claims:

Claims are equal to losses – The insurer is influenced by the full value of losses from the catastrophic events, and so $C_i = U_i$ is set. This approach may be applied if e.g. the considered insurer has complete monopoly for the given region.

Claims are a deterministic part of losses – Only some given, deterministic part of losses is transferred into claims. Then, the formula (8.16) with the specified parameter α_{claim} is used. Such approach may be applied if e.g. some expert knowledge about the share in the whole market for the considered insurer is available.

Claims are a random part of losses – Only some random part of losses is equal to claims. Then, the formula (8.15) with the fixed value $\alpha_{\text{claim}} = 1$ and the specified limits $c_{\min}, c_{\max}$ is used. This model may be appropriate in many circumstances, e.g. if only some general knowledge of the characteristics of share in the market for the insurer is available.

Claims are both a deterministic and a random part of losses – In this case the formula (8.15) with some specified parameters $\alpha_{\text{claim}} < 1, c_{\min}, c_{\max}$ is used.

8.4.2 Model of interest rates

In order to simulate the risk process R_T^{FV}, given by (8.20), it is necessary to evaluate the future values of the considered cash flows.

If the Vasicek model is applied (see Section 7.5 and Section 8.2.2 for additional details), then a special iterative formula should be used to simulate such risk-free interest rate process r_t. The relevant algorithm is discussed in (Glasserman 2004). Now we describe the most important conclusions, related to this approach.

To evaluate r_t at fixed moments $0 = t_0 < t_1 < t_2 < \ldots < t_n$, we apply the formula

$$r_{t_{i+1}} = e^{-\kappa(t_{i+1}-t_i)} r_{t_i} + \theta\left(1 - e^{-\kappa(t_{i+1}-t_i)}\right) + \sigma\sqrt{\frac{1 - e^{-2\kappa(t_{i+1}-t_i)}}{2\kappa}} Z_i \ , \quad (8.24)$$

where $Z_1, Z_2, \ldots, Z_n$ are *iid* samples from $N(0,1)$. As noted by Glasserman (2004), such approach is an exact simulation, i.e. the distribution of $r_{t_1}, r_{t_2}, \ldots, r_{t_n}$ is the same as of the Vasicek process at times $t_1, t_2, \ldots, t_n$ for the fixed value r_0. Then, the cdf of $r_{t_{i+1}}$ for the given value of r_{t_i} is equal to

$$r_{t_{i+1}} \sim N\left(e^{-\kappa(t_{i+1}-t_i)} r_{t_i} + \theta\left(1 - e^{-\kappa(t_{i+1}-t_i)}\right), \right.$$
$$\left. \sigma^2 \frac{1 - e^{-2\kappa(t_{i+1}-t_i)}}{2\kappa}\right) \ . \quad (8.25)$$

Instead of the formula (8.24), the simpler Euler scheme based on the approximation of the Vasicek model, given by (8.6), via the straightforward relation

$$r_{t_{i+1}} = r_{t_i} + \kappa\left(\theta - r_{t_i}\right)\left(t_{i+1} - t_i\right) + \sigma\sqrt{t_{i+1} - t_i}$$

can also be used. However, this approach leads to some discretisation error, as described by Glasserman (2004). Therefore, the exact solution (8.24) is further applied in our simulations.

Then, it is possible to use the simulated trajectory $r_{t_1}, r_{t_2}, \ldots, r_{t_n}$ to approximate the factor

$$\mathrm{fv}_{(t_i,t_{i+1})} = \int_{t_i}^{t_{i+1}} r_s ds \ ,$$

which is necessary to evaluate the future value. Also in this case, the simple formula $r_{t_i}\left(t_{i+1} - t_i\right)$ can be applied. However, as discussed by Glasserman (2004), a better approach is to simulate the whole paths of pairs $\left(r_t, \mathrm{fv}_{(0,t)}\right)$ applying a specially selected two-dimensional Gaussian distributions. The distribution of $r_{t_{i+1}}$ for the fixed value of r_{t_i} is described by (8.25). The cdf of $\mathrm{fv}_{(0,t_{i+1})}$ for the given $\mathrm{fv}_{(0,t_i)}$ and r_{t_i} is equal to

$$\begin{aligned}\mathrm{fv}_{(0,t_{i+1})} \sim N\Bigg(&\mathrm{fv}_{(0,t_i)} + \frac{1}{\kappa}\left(1 - e^{-\kappa(t_{i+1}-t_i)}\right) r_{t_i} \\ &+ \frac{\theta}{\kappa}\left(e^{-\kappa(t_{i+1}-t_i)} + \kappa\left(t_{i+1} - t_i\right) - 1\right), \\ &\frac{\sigma^2}{\kappa^2}\left(\left(t_{i+1} - t_i\right) + \frac{1}{2\kappa}\left(1 - e^{-2\kappa(t_{i+1}-t_i)}\right) + \frac{2}{\kappa}\left(e^{-\kappa(t_{i+1}-t_i)} - 1\right)\right)\Bigg) \ . \end{aligned} \tag{8.26}$$

The conditional covariance inside the pair $\left(r_{t_{i+1}}, \mathrm{fv}_{(0,t_{i+1})}\right)$ for the fixed value of $\left(r_{t_i}, \mathrm{fv}_{(0,t_i)}\right)$ is equal to

$$\frac{\sigma^2}{2\kappa}\left(1 + e^{-2\kappa(t_{i+1}-t_i)} - 2e^{-\kappa(t_{i+1}-t_i)}\right) \ . \tag{8.27}$$

Taking into account the known distribution (8.25) of $r_{t_{i+1}}$, the distribution (8.26) of $\mathrm{fv}_{(0,t_{i+1})}$ and the covariance (8.27) between the $r_{t_{i+1}}$ and $\mathrm{fv}_{(0,t_{i+1})}$, it is possible to apply Cholesky decomposition (see Section 4.8) to simulate the joint trajectory of both the risk-free rate r_t and the relevant integral $\mathrm{fv}_{(0,t)}$.

The last step in the process of modelling the trajectory of the interest rates and the evaluation of future values is to select the necessary parameters of the Vasicek model. As in Sections 8.2.2 and 8.2.3, the parameters estimated in Chan et al. (1992) for the U.S. Treasury bill yield data, namely

$$\kappa = 0.1779 \ , \ \theta = 0.086565 \ , \ \sigma^2 = 0.0004 \tag{8.28}$$

with $r_0 = 0.03$, will be used in the following.

8.4.3 Evaluation of the cat bond price

As described in Section 8.3.2, the catastrophe bond is one of the most important financial instruments considered in our portfolio. Therefore, it is necessary to evaluate its price, which is used in the formula (8.13) to obtain the insurance premium p_{cb}. In order to do this, the approach presented in Chapter 7 is directly applied.

The relevant analysis of the cat bond price is described in Section 8.2.3. Let us recall that we assumed that $T = 1$, i.e. the maturity time of the considered catastrophe bond is equal to one year. The face value of this bond is equal to one (the one monetary unit assumption). The payment function for the considered catastrophe bond has piecewise linear form (see Section 7.3). The triggering points K_i are given by the quantiles of the cumulated value of losses as

$$K_0 = Q^{\mathrm{loss}}_{\text{NHPP-LN}}(0.75), K_1 = Q^{\mathrm{loss}}_{\text{NHPP-LN}}(0.85), K_2 = Q^{\mathrm{loss}}_{\text{NHPP-LN}}(0.95)$$

and the values of payments decreases are set to

$$w_1 = 0.4, w_2 = 0.6 \ .$$

After applying Monte Carlo simulations, the obtained value of the estimator of the catastrophe bond price is equal to 0.839936.

8.4.4 Model of the reinsurance contract

As described in Section 8.3.3, the reinsurer contract is the next important layer in the considered portfolio of the insurer.

As previously, one-year horizon is considered, i.e. $T = 1$. If excess-of-loss policy, defined by (8.18), is taken into account, then the attachment point A_{rnsr} and the cap level B_{rnsr} should be set. In order to do this, we adopt the approach similar to the one used for the triggering point of the catastrophe bond, i.e. the values of A_{rnsr} and B_{rnsr} are given by quantiles of the cumulated claims. Such quantiles are further denoted by $Q^{\mathrm{claim}}_{\text{NHPP-LN}}(x)$. They are related to the process of the aggregated claims $\tilde{C}_t$, described by the formula (8.23), instead of the process of the aggregated losses $\tilde{N}_t$, connected with the quantiles $Q^{\mathrm{loss}}_{\text{NHPP-LN}}(x)$ (see Section 8.2.3).

If the claims are not exactly equal to the catastrophic losses, then the shape of the function $Q^{\mathrm{claim}}_{\text{NHPP-LN}}(x)$ is similar to the shape of $Q^{\mathrm{loss}}_{\text{NHPP-LN}}(x)$, but with lower values, in general. An example of such a graph is illustrated with Figure 8.11 where $c_{\min} = 0.25, c_{\max} = 0.5, \alpha_{\mathrm{claim}} = 1$ are set.

As noted in Section 8.3.3, prior to the analysis of the portfolio, the premium for the reinsurer p_{rnsr} should be evaluated for the given set of parameters. In order to do this, as indicated by the formula (8.19), the expected value $\mathbb{E}\left(\mathrm{PV}\left(f^i_{\mathrm{rnsr}}\left(\tilde{C}_T\right)\right)\right)$ should be estimated, using, e.g., the numerical approach.

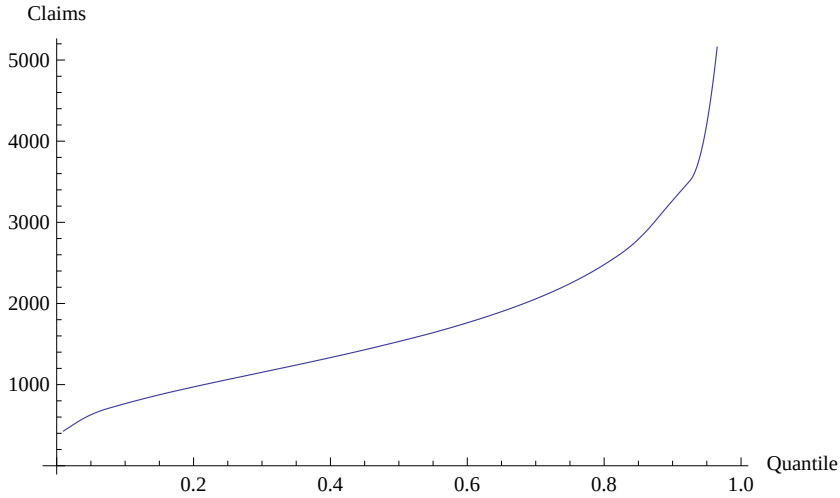

Fig. 8.11. Graph of the quantiles $Q^{\text{claim}}_{\text{NHPP-LN}}(x)$ of the cumulated value of claims (in million \$)

As previously, the model of the losses, discussed in Section 8.4.1, and the Vasicek model of the interest rates, presented in Section 8.4.2, are applied in our considerations.

An example of such estimation of the premium p_{rnsr} may be found in Figure 8.12. In this case the claims are exactly equal to losses, the safety loading for the insurer ν_{rnsr} is set to zero, and the cap level B_{rnsr} is fixed at $Q^{\text{claim}}_{\text{NHPP-LN}}(0.99) = Q^{\text{loss}}_{\text{NHPP-LN}}(0.99)$, i.e. the 0.99-th quantile. As it can be seen from Figure 8.12, the relation between the premium p_{rnsr} and the attachment point A_{rnsr} for various values of quantiles of the cumulated claims is an almost linear function.

Of course, if instead of quantiles the normal values of cumulated claims are considered, then the premium p_{rnsr} is a strictly convex, decreasing function of $\tilde{C}_1$ (see Figure 8.13).

8.4.5 Calibration of other parameters

There are some other parameters of the entire portfolio, which should be also set before the analysis. In the case of the risk reserve process, the initial reserve of the insurer u is one of the most important variables. In practice, this value is sometimes given by the relevant legal regulations, like e.g. the requirement for the minimal necessary limit of u.

In the model of the portfolio, considered in this chapter, the value of u is related to some quantile of the cumulated value of the claims $Q^{\text{claim}}_{\text{NHPP-LN}}(x)$. This assumption is consistent with the triggering points for the catastrophe bond and the limits for the reinsurer contract, introduced previously.

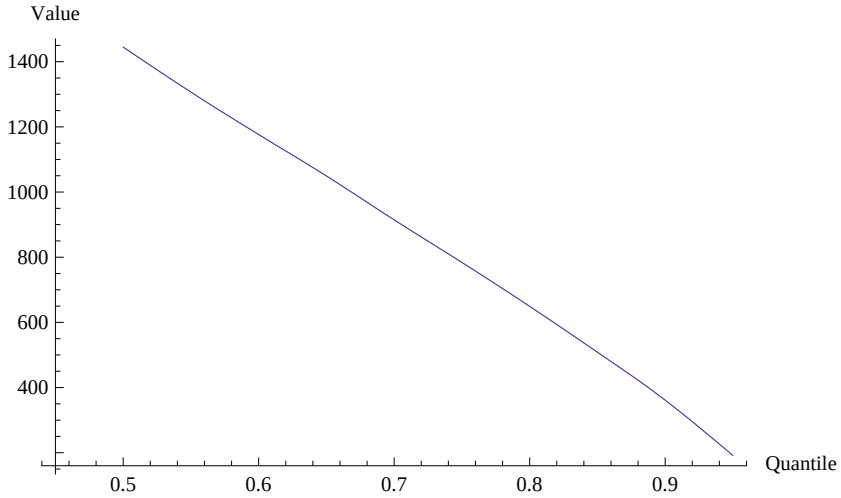

Fig. 8.12. Graph of the value of the reinsurance contract for the given quantiles $Q^{\text{claim}}_{\text{NHPP-LN}}(x)$ (in million \$)

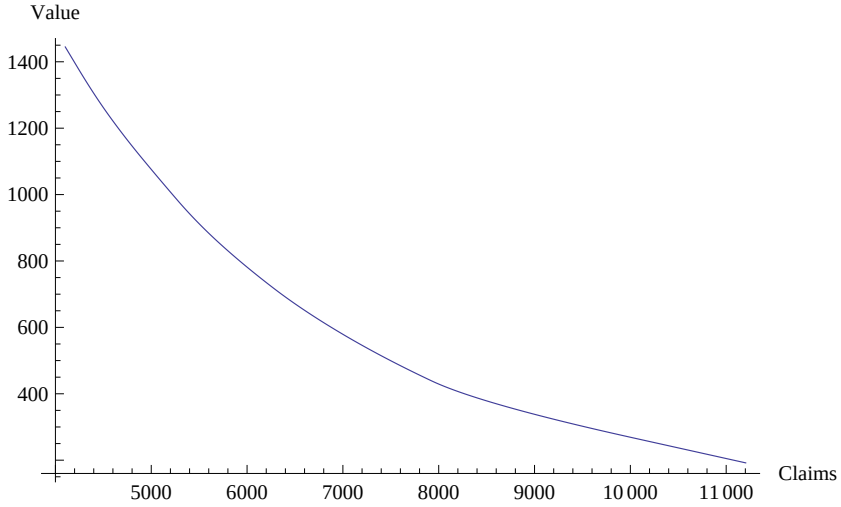

Fig. 8.13. Graph of the value of the reinsurance contract for the given values of $\tilde{C}_1$ (in million \$)

The other parameter is the safety loading ν_p for the insurance premium p, introduced in (8.11). As mentioned in Section 8.3.1, the typical values of ν_p are about 0.1 – 0.2. Also similar parameter for the reinsurer, the safety loading ν_{rnsr}, should be set. In our considerations, this parameter has rather small values, contained in the range 0.05 – 0.1.

8.5 Analysis of the portfolio

After the discussion, concerning the basis of the numerical methods applied in our approach and introduction of the necessary assumptions for models and their parameters, it is possible to analyse the behaviour of the portfolio, which consists of the layers described in Section 8.3. This analysis will be divided into a few diversified cases – and thus relations among various assumptions and simulated outputs are directly verified.

8.5.1 First insight – one layer

We start from the simplest case – when there is only one layer in the portfolio, i.e. only the classical risk reserve process is considered, without additional layers, described in Sections 8.3.2 – 8.3.4. Of course, in the course of our analysis, the future value of such a portfolio for $T = 1$ for the Vasicek model with parameters given by (8.28) is assumed.

Analysis I.1

We first consider the case when the claims are equal to losses, i.e. the processes $\tilde{C}_t$ and $\tilde{N}_t$, with parameters described in Section 8.4.1, are exactly the same. As noted in Section 8.4.5, in our analysis the initial reserve u is related to the quantile of the cumulated value of the claims, therefore we set $u = Q^{\text{claim}}_{\text{NHPP-LN}}(0.20) = 2609.79$. As indicated by Jensen (1986), the initial reserve should not be too high, because the insurers with high values of their own funds are very often acquired by other enterprises. The safety loading is set to rather a standard value, such as $\nu_p = 0.1$.

For the assumptions mentioned above, the numerical procedure described in Section 8.4 is applied. Based on its output, various measures and graphs interesting for practitioners are acquired and some important conclusions can be drawn.

In the case considered, the probability of the insurer's bankruptcy in time $T = 1$ is equal to 11.5333%. Further on, this event, i.e. the occurrence of $R^{\text{FV}}_T < 0$, is called *final ruin*.

However, it is possible that the value of the risk process R_t falls below zero for some $t \leq T$ even if its value is *above zero* for T. Then, instead of the "standard" bankruptcy, verified only in a fixed time moment, the possibility of "earlier"

bankruptcy exists. In our analysis, the estimated probability of the event $R_t < 0$ for some $t < T$ (which is called further on *earlier ruin*) is equal to 17.0583%. Therefore, there is also a possibility that the earlier bankruptcy does not lead to the insolvency in the final considered moment, i.e. that $R_T^{\mathrm{FV}} < 0$. The probability of such an event is numerically evaluated as 5.525% in this setup.

Measure	Value
Minimum	-676250
First quartile	2292.88
Median	4209.15
Third quartile	5474.94
Maximum	8381.82
Mean	3245.54
Std. deviation	4366.24

Table 8.2. Statistical measures of the portfolio for $T = 1$ from Analysis I.1

The most important statistical measures of the value of the whole portfolio in this case, i.e. of the process R_T^{FV}, can be found in Table 8.2. As it may be noted, the minimum value is significantly lower (259.12 times!) than the initial reserve. The values of other measures, like first quartile, median, mean and third quartile are similar to u or even higher. Therefore, the catastrophic events have important impact on the value of the portfolio and they lead to high probability of bankruptcy or of low income of the insurer. This can also be seen when the relevant quantiles for the values of the portfolio are plotted (see Figure 8.14). For better readability only the quantiles from the range 0.01 – 0.99 are used there. The "left tail" of this plot for $R_T^{\mathrm{FV}} < 0$ is a very quickly increasing function, which indicates the possibility of very high insolvency.

But if the insurer is "lucky enough", some strictly positive income is also possible. It is indicated by the middle, linear, slowly increasing part of the plot from Figure 8.14. The probability of extra profits with much higher values than the initial reserve is rather limited – the rapidly increasing "right tail" of the plot with exponential shape is very short (about 5% of the simulated samples).

From the insurer's point of view, very high probability of ruin with potentially almost unlimited losses constitutes an important problem. Therefore, we analyse the behaviour of the portfolio in the case of insolvency in a more detailed way. The histogram of values of the portfolio if final ruin occurs may be found at Figure 8.15. However, due to a very low minimum value for such an event and a significant number of outliers, such plot is almost unreadable. A better insight is provided if histogram of only 90% of the whole output is plotted, without the lowest 10% of samples (see Figure 8.16). Then, the shape of such a modified histogram is similar to the exponential function. As it can be seen, the most probable events are the ruins with their values close to zero.

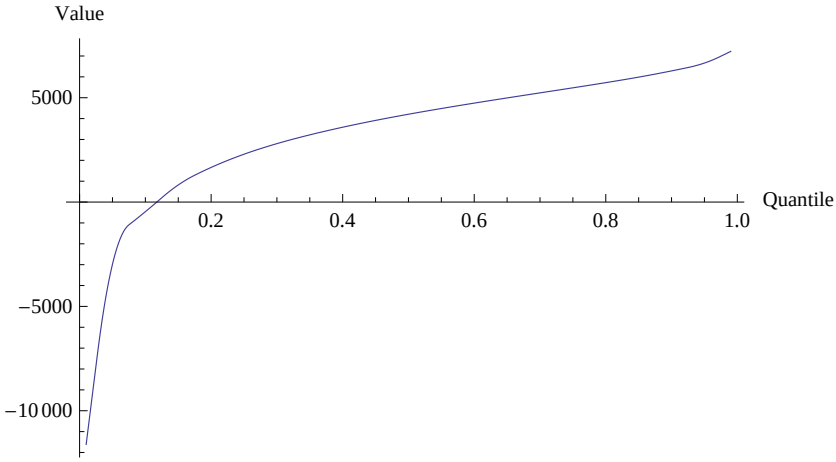

Fig. 8.14. Graph of quantiles of the final value of the portfolio from Analysis I.1 (in million $)

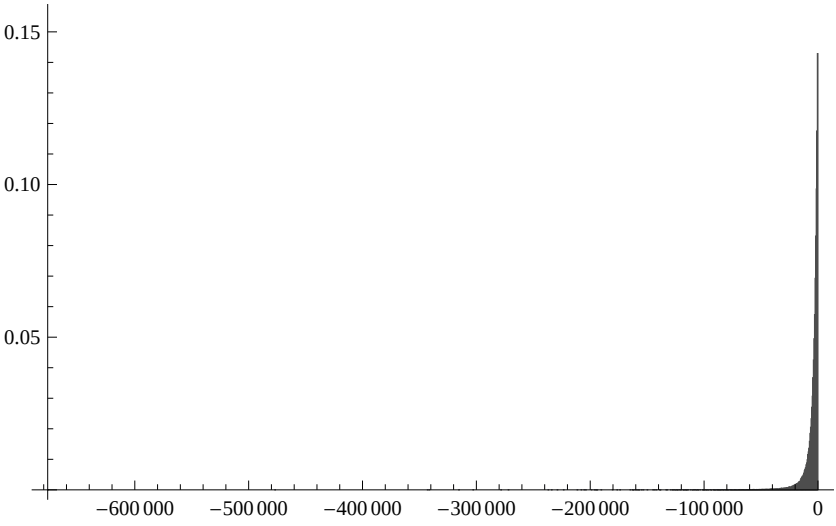

Fig. 8.15. Histogram of the values of the portfolio in the case of the final ruin from Analysis I.1 (in million $)

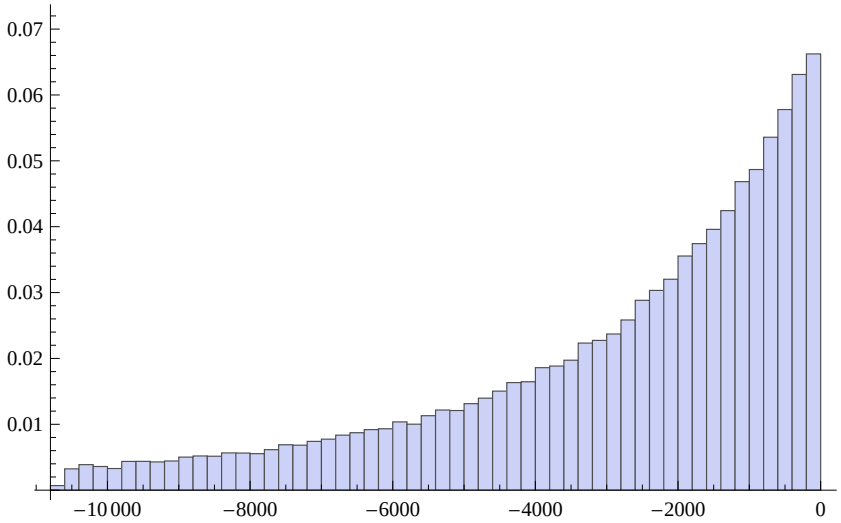

Fig. 8.16. Histogram of the 90% of values of the portfolio in the case of the final ruin from Analysis I.1 (in million $)

Other statistical measures are also numerically evaluated for the case of the final ruin and are presented in Table 8.3. As it can be seen from these data, even the mean is 1.79084 times higher than the initial reserve and the median has a similar value to u (with minus sign of course). Therefore, the final ruin has important consequences for the insurer in the considered case, because of high possible value of such insolvency.

Measure	Value
Minimum	-676250
First quartile	-5399.38
Median	-2412.39
Third quartile	-951.83
Maximum	-0.0561358
Mean	-4673.71
Std. deviation	8458.6
Skewness	-14.4552
Kurtosis	609.121

Table 8.3. Statistical measures of the portfolio in the case of the final ruin from Analysis I.1

Apart from the final ruin, the previously mentioned possibility of the *earlier ruin* should also be investigated. The values of the portfolio if the earlier ruin

occurs are illustrated with the histogram (see Figure 8.17) and the relevant statistical measures (see Table 8.4). As previously, because of very low minimum of value of the portfolio for such an event, the complete histogram is almost unreadable. Therefore, also in this analysis, the histogram of 90% of samples, without 10% of the lowest values, is plotted (see Figure 8.18). Both of these figures are similar to the histograms for the final ruin (compare with Figures 8.15 and 8.16).

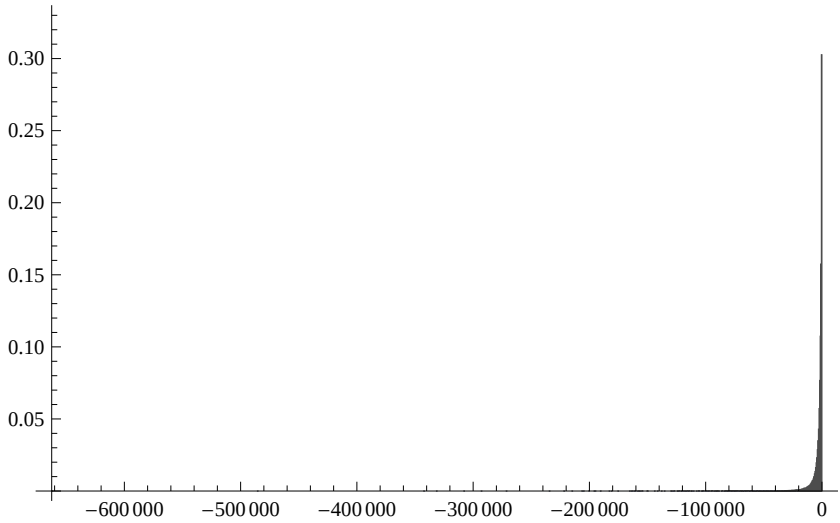

Fig. 8.17. Histogram of the values of the portfolio in the case of the earlier ruin from Analysis I.1 (in million $)

There are some small differences in statistical measures between the cases of the final and the earlier ruin (compare Table 8.3 with Table 8.4). As indicated by the minimum, quartiles and mean, the overall value of the portfolio is higher for the earlier ruin. It is also less varied. However, minimum still has a very low value, and so the earlier ruin is also an important problem for the insurer.

For the earlier ruin, not only the value of the portfolio is interesting for the insurer. The timing of these events may be also analysed. It can be seen from the related histogram (see Figure 8.19) that the distribution of the moments of the earlier ruins is not a uniform one. It rather follows some trigonometrical function, which seems to be connected with the model of the number of losses (8.21), given by the sinus function.

There is no strict dependency between the value of the earlier ruin and the timing of such event – for the considered case the correlation coefficient is equal to 0.0058815, hence it is indeed very close to zero.

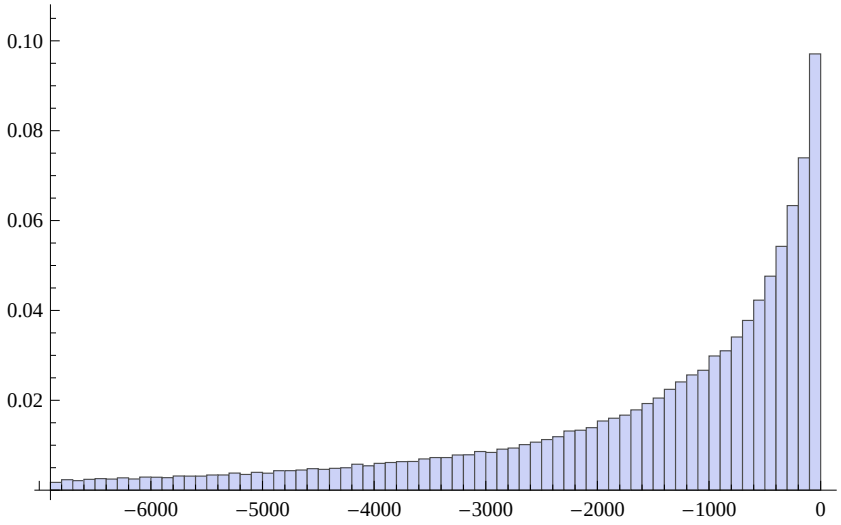

Fig. 8.18. Histogram of the 90% of values of the portfolio in the case of the earlier ruin from Analysis I.1 (in million $)

Measure	Value
Minimum	-662240
First quartile	-3071.59
Median	-1167.68
Third quartile	-379.584
Maximum	-0.00361414
Mean	-4673.71
Std. deviation	6673.69
Skewness	-18.34
Kurtosis	987.258

Table 8.4. Statistical measures of the portfolio in the case of the earlier ruin from Analysis I.1

Analysis I.2

One of the remedies to the problem of the insurer's insolvency is to increase the initial reserve u. As it was noted previously, this may be completely impractical in the real world, because of the possibility of a hostile takeover by another enterprise. However, from the theoretical point of view it is interesting to analyse the respective probabilities of bankruptcy for the cases of the final ruin and the earlier ruin. Such issue is illustrated with Figure 8.20, where the estimated probabilities of the final ruin (circles) and the earlier ruin (squares) as the functions of u (given by quantiles $Q^{\text{claim}}_{\text{NHPP-LN}}(x)$) are plotted.

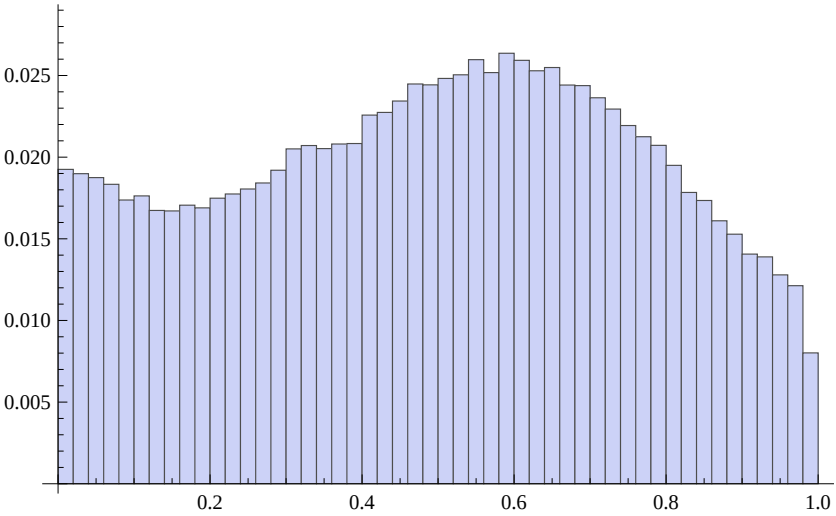

Fig. 8.19. Histogram of the moments of the earlier ruins from Analysis I.1

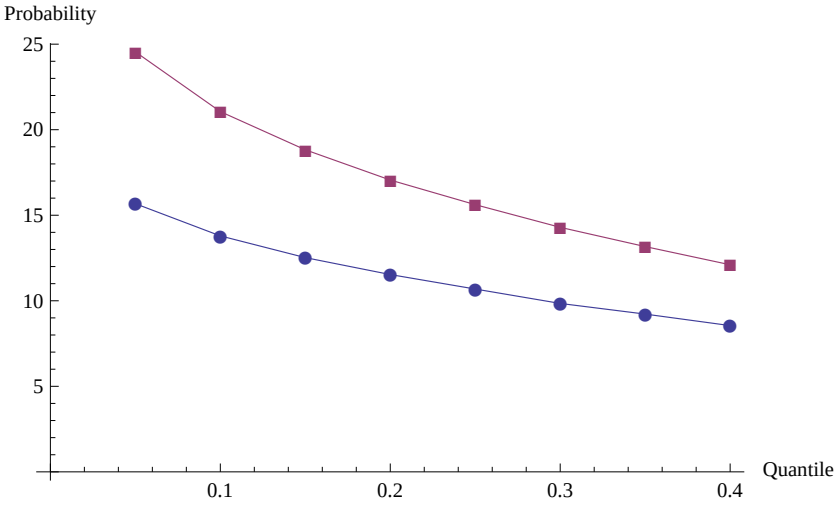

Fig. 8.20. Probabilities of the final (circles) and earlier (squares) ruins depending on the initial reserve from Analysis I.2

As it can be seen, a higher value of u is not a suitable solution – even for the initial reserve as high as 0.4 quantile of the whole cumulated value of the claims, the evaluated probabilities of the ruins are still very significant (8.549% for the final ruin and 12.1058% for the earlier ruin). Additionally, these plots are not linear. This indicates that the linear or even exponential increase of the value of u (compare with the related quantiles in Figure 8.11) does not lead to the same scale of decrease for the probability of the ruin. Therefore, other solutions for the problem of insolvency are necessary.

8.5.2 Comparison of portfolios with two layers

It is known that the problem of insolvency of the insurer may be solved using additional reinsurance contract, because – as it was shown also in Analysis I.2 – the increase of the initial reserve is not an optimal remedy. Therefore, in this section we start from the modelling of portfolio of the insurer with addition of such a contract.

Another solution, also considered in this section, is to apply the catastrophe bond instead of the reinsurance contract. Because we assume in this part that the analysed portfolio is limited to only two layers, therefore these two different approaches are compared afterwards.

Analysis II.1

We start our considerations from the case similar to the one discussed in Analysis I.1 (see Section 8.5.1). As previously, we assume $T = 1$, the processes $\tilde{C}_t$ and $\tilde{N}_t$ are exactly the same, the initial reserve u is equal to $Q^{\text{claim}}_{\text{NHPP-LN}}(0.20)$ and the safety loading is set to $\nu_p = 0.1$.

The reinsurance contract is modelled using the approach described in Section 8.4.4. Then, the safety loading ν_{rnsr} is equal to 0.05 and the cap level B_{rnsr} is fixed at $Q^{\text{claim}}_{\text{NHPP-LN}}(0.99)$. The attachment point A_{rnsr} is set as $Q^{\text{claim}}_{\text{NHPP-LN}}(0.8)$. Therefore, the reinsurance contract is used only if relatively high value of total claims is achieved.

Introduction of this additional reinsurance contract changes the behaviour of the portfolio in an essential way. Based on numerical simulations, the probability of the final ruin is estimated as 0.896%, the probability of the earlier ruin is equal to 22.272%, and the probability that the earlier ruin does not lead to the final ruin is evaluated as 21.376%. It means that the probability of the final insolvency of the insurer is significantly lowered – a major part of the risk, related to the high value of the cumulated claims is transferred to the reinsurer. However, due to the necessity of paying the reinsurer's premium p_{rnsr} at the very beginning of the valuation of the portfolio (i.e. at $t = 0$), the probability of the earlier ruin is higher by about 5.2137%.

Table 8.5 presents the statistical measures related to the value of the portfolio for $T = 1$. They should be compared with the relevant data for Analysis I.1 (see

Table 8.2). The first quartile, median and third quartile are lower in the case with the reinsurance contact. This behaviour seems to be related to the necessity of paying the reinsurer's premium p_{rnsr} always, regardless of the fact whether such contract is at all used. The mean is similar in both cases, while the minimum and maximum are higher for the portfolio considered here.

Measure	Value
Minimum	-663837
First quartile	1602.43
Median	3506.85
Third quartile	4770.81
Maximum	19702.8
Mean	3211.98
Std. deviation	2959.35

Table 8.5. Statistical measures of the portfolio for $T = 1$ from Analysis II.1

Similarly, the graph of the quantiles of the final value of the portfolio for the presented case should be compared with the similar plot prepared during Analysis I.1 (compare Figures 8.14 and 8.21). The graph for the portfolio with application of the reinsurance contract has a completely different shape – it is almost linear, especially for quantiles higher than 0.2. This "breaking point" may be easily related to the value of the attachment point A_{rnsr}, used in our considerations. However, we should keep in mind the fact that Figure 8.21 is plotted only for quantiles in the range of 0.01 – 0.99, and so the minimum value of such portfolio is still very low (as indicated in Table 8.5).

As previously, both of the final ruin and the earlier ruin should be more deeply analysed. The statistical measures for the final ruin can be found in Table 8.6. After comparison with Table 8.3, where the relevant estimators for Analysis I.1 are evaluated, it appears that the final ruin is far less probable if the reinsurer's contract is used, but this insolvency has also a greater impact. All of the important location measures (quartiles, median, mean) are significantly lower, some of them even at half of the values from Table 8.3. The same applies to standard deviation. Such behaviour may be related to the existence of the cap level B_{rnsr} – the reinsurer transfers only a fixed amount in excess of this limit, but the total value of losses could be significantly higher than $Q^{\text{claim}}_{\text{NHPP-LN}}(0.99)$. Then, this uncovered "excess" has an important impact on the insolvency of the portfolio.

The histogram of the 90% of values of the portfolio if the final ruin occurs is shown in Figure 8.22. This plot has a very similar shape to the one from Figure 8.16, which illustrates the same event for Analysis I.1. The same applies for the histogram of values for the case of the earlier ruin (compare Figures 8.18 and 8.23).

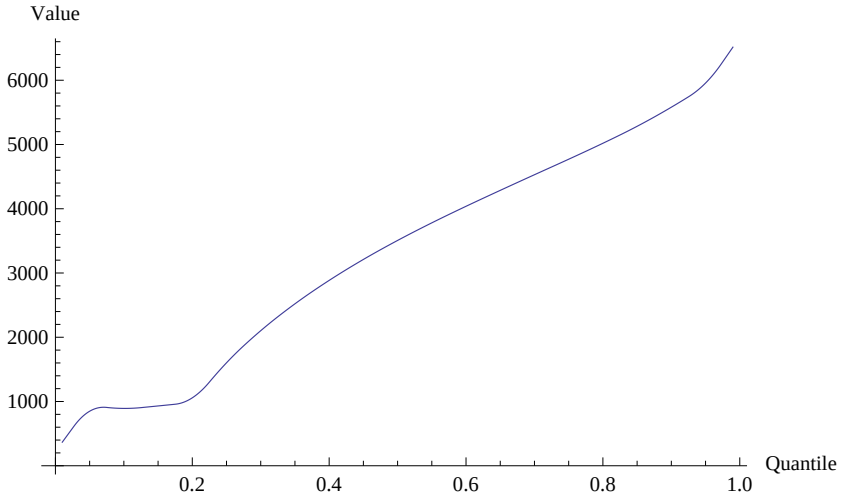

Fig. 8.21. Graph of quantiles of the final value of the portfolio from Analysis II.1 (in million $)

Measure	Value
Minimum	-663837
First quartile	-13217.1
Median	-5766.65
Third quartile	-2145.22
Maximum	-0.356304
Mean	-11524.8
Std. deviation	20584.6
Skewness	-9.30748
Kurtosis	182.724

Table 8.6. Statistical measures of the portfolio in the case of the final ruin from Analysis II.1

Likewise, the statistical measures of value of the portfolio in the case of the earlier ruin (see Table 8.7) are very similar to the respective data for Analysis I.1 (see Table 8.4). Mean is here one important exception.

Interesting conclusions are drawn when the time instants of occurrence of the earlier ruin are compared. As illustrated by the envelopes of histograms (see Figure 8.24), the probability of the earlier ruin is higher for the initial moments if the reinsurance contract is applied. This is indicated by the higher plot, marked by circles for the times up to about $t = 0.4$. Afterwards, especially for $t \geq 0.6$, the probability of such event is lower than in the case when there is no insurance contract (which is marked by squares). Such behaviour may be once again related to the necessity of paying the reinsurer's premium at $t = 0$.

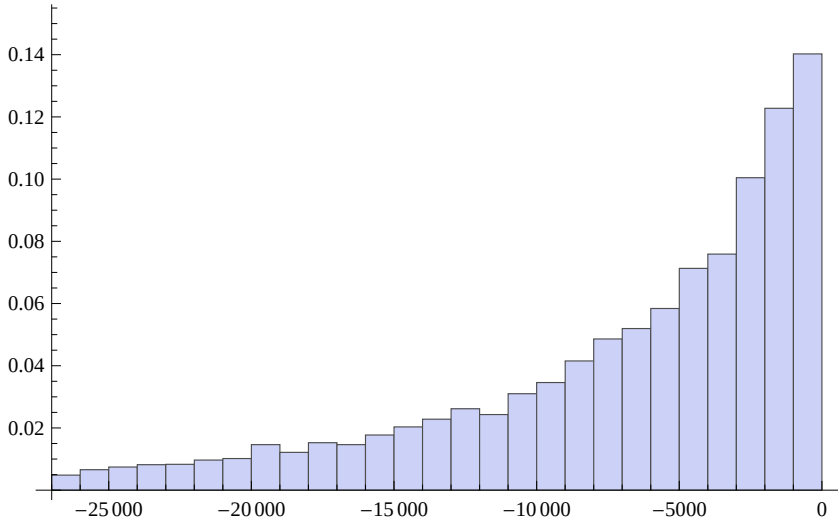

Fig. 8.22. Histogram of the 90% of values of the portfolio in the case of the final ruin from Analysis II.1 (in million $)

Measure	Value
Minimum	-662924
First quartile	-2660.13
Median	-1010.46
Third quartile	-333.874
Maximum	-0.00454387
Mean	-2564.1
Std. deviation	5959.28
Skewness	-19.9909
Kurtosis	1200.72

Table 8.7. Statistical measures of the portfolio in the case of the earlier ruin from Analysis II.1

Therefore, the portfolio is in the considered case highly sensitive to the existence of reinsurance contract at the beginning of the time period. Overall, this contract is an important advantage compared to the case discussed in Section 8.5.1.

The correlation coefficient between the time moments and the values of the earlier ruin is still close to zero and is evaluated as -0.00751083.

In the considered case also payments from the additional layer, i.e. from the reinsurance contract, are numerically evaluated and then analysed. Because of the applied values of the attachment point and the cap level, the probability that this layer is used is relatively very high and is estimated as equal 19.9891%. Therefore, from the practical point of view it may be questionable if such contract

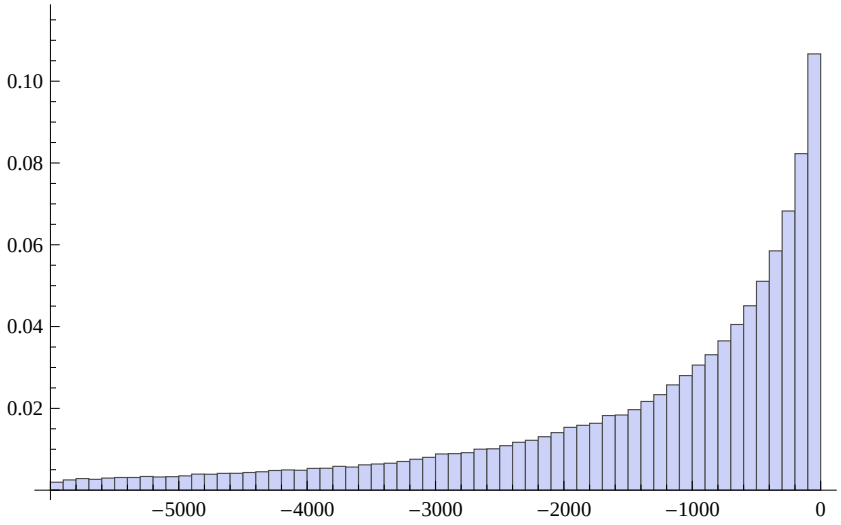

Fig. 8.23. Histogram of the 90% of values of the portfolio in the case of the earlier ruin from Analysis II.1 (in million $)

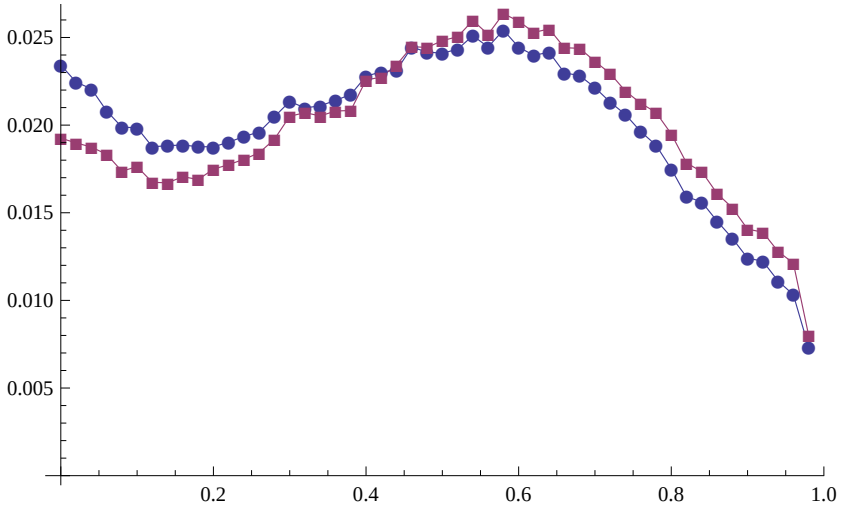

Fig. 8.24. Comparison of envelopes of histograms of moments of the earlier ruin, obtained in Analysis I.1 (squares) and Analysis II.1 (circles)

is even profitable for the reinsurer and if it could be actually available for the insurer.

The graph of the quantiles of the value of payments from the reinsurance contract, if such payments are above zero, is shown in Figure 8.25. This plot is almost an exponential function. Additionally, the most important statistical measures of these payments can be found in Table 8.8. As it can be seen from the value of the standard deviation and especially of the maximum, these payoffs may be very high. However, the median and the mean are not very far from the insurer's initial reserve u.

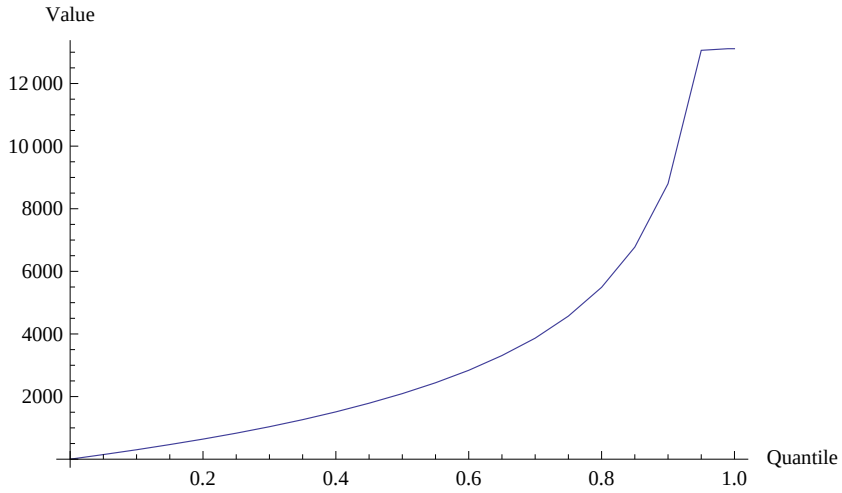

Fig. 8.25. Graph of quantiles of the value of payments from the reinsurance contract from Analysis II.1 (in million $)

Measure	Value
Minimum	0.00844779
First quartile	830.573
Median	2094.37
Third quartile	4570.63
Maximum	13111.9
Mean	3364.22
Std. deviation	3502.36

Table 8.8. Statistical measures of the value of payments from the reinsurance contract from Analysis II.1

Analysis II.2

As it can be easily seen from the description of models, presented in Sections 8.3.2 and 8.3.3, the reinsurance contract and the catastrophe bond are different instruments in many ways. Therefore, it is not possible to directly compare outputs from *the same* models, but only *similar* ones can be considered. Keeping this in mind, in the following we apply the parameters of the catastrophe bond which yield the setting as similar as possible to the one discussed in Analysis II.1.

The model of the fundamental risk reserve process and the model of the claims are the same as in Analysis I.1 and Analysis II.1. We use the catastrophe bond (see Section 8.4.3) with the piecewise linear payment function. In order to achieve the earlier mentioned similarity to the reinsurance contract, considered in Analysis II.1, only two triggering points are set. They have the same values as the attachment point A_{rnsr} and the cap level B_{rnsr}, so that $K_0 = Q^{\text{claim}}_{\text{NHPP-LN}}(0.8)$ and $K_1 = Q^{\text{claim}}_{\text{NHPP-LN}}(0.99)$. The face value, as usually, is equal to one and the sole value of payment decrease is set as $w_1 = 1$. Then, if the cumulated value of claims is beyond K_1, the whole payment for the policyholder is lost (see Chapter 7 for additional details).

In order to ensure similarity to Analysis II.1, the payment paid by the insurer α_{cb} (see formula (8.13)) is similar to the previously used safety loading for the reinsurer ν_{rnsr}, i.e. it is equal 0.05. Then the quantity n_{cb} of catastrophe bonds, which are sold to policyholders, is evaluated as equal 14881. Therefore, the insurer pays the premium for the bonds p_{cb}, which is equal to the reinsurer's premium p_{rnsr} in Analysis II.1. The cat bond price, which is also necessary for the formula (8.13), is evaluated via the approach introduced in Chapter 7.

Using Monte Carlo simulations it was estimated that the probability of the final ruin is equal to 0.7059%, so it is close to but also lower than in the case of Analysis II.1. The probability of the earlier ruin is equal to 22.272%, and the probability of the earlier ruin which does not lead to the final insolvency is evaluated as 21.5661%. Therefore, using such catastrophe bond as an additional layer in the portfolio is similar to the introduction of the reinsurance contract, considered in the preceding analysis. The necessity of paying some additional premium by the insurer (reinsurer's premium or special payment for SPC, as described in Section 8.1) leads to higher probability of the insolvency in the beginning, but the overall influence is positive and the final ruin is much less probable.

As previously, the statistical measures of the value of the portfolio at $T = 1$ are also analysed. By comparing Table 8.5 (measures for Analysis II.1) with the relevant Table 8.9 (data for Analysis II.2), we can see that the overall behaviour of these two portfolios is very similar. However, a more detailed analysis leads to noticing some important differences. Figure 8.26 shows the plots of the quantiles (from the 0.01 quantile to the 0.99 quantile) for the cases of Analysis II.1 (squares) and Analysis II.2 (circles). The obtained graph for the lower quantiles (up to the 0.3 quantile) is much smoother and indicates higher values when the catastrophe bond is used instead of the reinsurance contract.

Measure	Value
Minimum	-662068
First quartile	1795.64
Median	3507.67
Third quartile	4771.05
Maximum	21471.9
Mean	3302.71
Std. deviation	2831.99

Table 8.9. Statistical measures of the portfolio for $T = 1$ from Analysis II.2

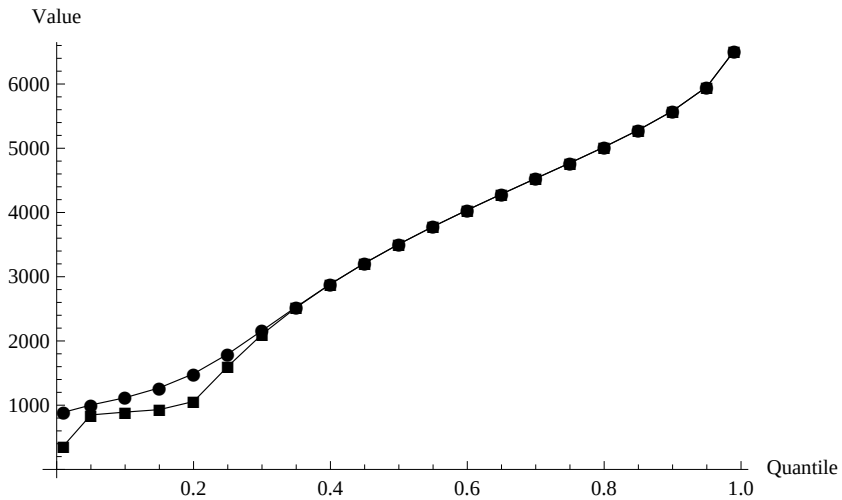

Fig. 8.26. Graph of quantiles of the final value of the portfolio from Analysis II.1 (squares) and Analysis II.2 (circles) (in million $)

For Analysis II.1, the events of the final ruin and the earlier ruin are also described. As it can be seen from the comparison of Table 8.6 with Table 8.10 and from Figure 8.27, where the related histograms of 90% of the values of the portfolios are shown together, the cases of the final ruins are comparable for the use of the reinsurer contract or the catastrophe bond. Overall, statistical measures of position have lower values for the portfolio with the catastrophe bond.

Of course, the conclusions which could be drawn for the earlier ruin in both of the cases are identical, because of the assumed equality of premiums p_{cb} and p_{rnsr}.

It should be noted that the probability that at least one of the triggering points for the catastrophe bond is surpassed (and the cat bond is "used" as the source of funds for the insurer) is estimated as equal 19.9891%. It means that with such probability the payment for the policyholder is lower than the

Measure	Value
Minimum	-662068.
First quartile	-14534.9
Median	-6331.59
Third quartile	-2427.59
Maximum	-0.776527
Mean	-12633.1
Std. deviation	22332.7
Skewness	-8.89261
Kurtosis	162.715

Table 8.10. Statistical measures of the portfolio in the case of the final ruin from Analysis II.2

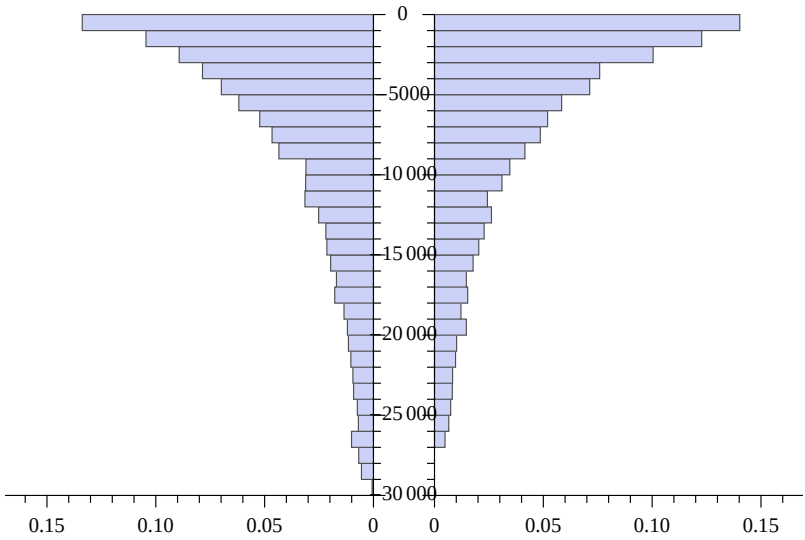

Fig. 8.27. Paired histogram of the 90% of values of the portfolios in the case of the final ruin for Analysis II.2 (left hand side) and for Analysis II.1 (right hand side), in million $

face value of the bond. Of course, this result is directly related to the parameter K_0 applied in this setting. But from the practical point of view, it seems that such catastrophe bond (even with relatively high probability of lowering its payment) should be more available than the reinsurance contract, considered in Analysis II.1. Therefore, it may be also easier to use for the insurer.

Apart from the probability, a more detailed analysis of the payments from the catastrophe bond was also conducted. The relevant statistical measures may be found in Table 8.11, and they can be easily compared with the data for the payoffs from the reinsurance contract from Table 8.8. Generally, the funds which

are paid to the insurer are higher if the catastrophe bond is used. The same is seen from Figure 8.28, where the graphs of quantiles are compared. For the two cases they have a similar shape, but quantiles for the payments from the cat bonds are clearly higher. High volatility, measured by the ratio of the standard deviation to the mean (see Table 8.11), should be also noted.

Measure	Value
Minimum	0.00958761
First quartile	942.638
Median	2376.96
Third quartile	5187.33
Maximum	14881
Mean	3818.14
Std. deviation	3974.91

Table 8.11. Statistical measures of the value of payments from the catastrophe bond from Analysis II.2

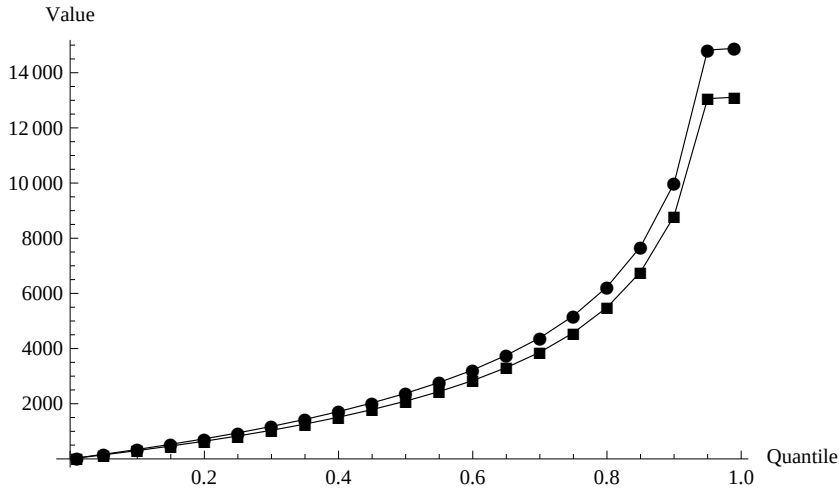

Fig. 8.28. Graph of quantiles of the value of payments from the reinsurance contract in Analysis II.1 (squares) and from the catastrophe bond in Analysis II.2 (circles) in million $

8.5.3 The whole portfolio – introduction of three layers

After taking into account only one or two layers, in the following, the portfolio, which consists of three layers is analysed. Into this portfolio, side by side with the classical risk process, both the catastrophe bond and the reinsurance contract are incorporated.

For a better comparison of outcomes, the overall setting is the same as assumed in Section 8.5.2. The main differences concern some specific details of both the catastrophe bond and the reinsurance contract.

We assume that the two triggering points for the catastrophe bond are set to $K_0 = Q^{\text{claim}}_{\text{NHPP-LN}}(0.8)$ and $K_1 = Q^{\text{claim}}_{\text{NHPP-LN}}(0.95)$, and the value of the payment decrease is given as $w_1 = 1$. For the reinsurance contract, the attachment point A_{rnsr} is equal to $Q^{\text{claim}}_{\text{NHPP-LN}}(0.95)$ and the cap level B_{rnsr} is set to $Q^{\text{claim}}_{\text{NHPP-LN}}(0.99)$. It means that the funds from the reinsurance contract are used *after* the cash flow from the catastrophe bond occurs, i.e. if the cumulated value of claims turns out to be even higher than K_1. Such assumption is closer to practice than the one considered during Analysis II.1, because it seems that the catastrophe bond, with the desired set of parameters from the insurer's point of view, should be more available than similar reinsurance contract.

Analysis III.1

We start from the assumption corresponding to the one used during Analysis II.2. Therefore, the new portfolio is constructed in such a way that the sum of the premium for the bonds p_{cb} and the reinsurer's premium p_{rnsr} is equal to the payment for the reinsurance contract used in Analysis II.1. As previously, $\alpha_{\text{cb}} = \nu_{\text{rnsr}} = 0.05$ and the expected present value of the reinsurance contract and the price of the cat bond are evaluated using Monte Carlo simulations (see Chapter 7 and Section 8.4 for additional details). Then, the estimated quantity n_{cb} of the catastrophe bonds, which are sold for the portfolio is equal to 11090. Of course, since the reinsurance contract is also one of the layers, this number is significantly lower than in Analysis II.2.

Basing on Monte Carlo simulations, the probability of the final ruin is estimated as equal 0.4149% – meaning that it is significantly lower than in Analysis II.1 and Analysis II.2. Due to the applied parameters of the cat bond and the reinsurance contract, the probabilities of the earlier ruin (22.2718%) and of the earlier ruin which does not lead to the final insolvency (21.8569%) are very similar to the ones in Analysis II.2. All of the statistical measures of position, presented in Table 8.12, are significantly higher than in Analysis II.1 and Analysis II.2 (as given in Tables 8.5 and 8.9, respectively).

The same is also indicated by Figure 8.29, where the plots of quantiles (from the 0.01 to the 0.99 quantile) of the final value of the portfolio for the now considered case and Analysis II.2 are compared. For the former the plot is clearly higher. Therefore, it should be profitable to apply both the catastrophe bond and the reinsurance contract. Such approach considerably lowers the risk of the

Measure	Value
Minimum	-657341
First quartile	2701.09
Median	3991.41
Third quartile	5116.74
Maximum	26199.4
Mean	3886.81
Std. deviation	2667.54

Table 8.12. Statistical measures of the portfolio for $T = 1$ from Analysis III.1

final insolvency and improves the overall behaviour of the portfolio of the insurer for $T = 1$.

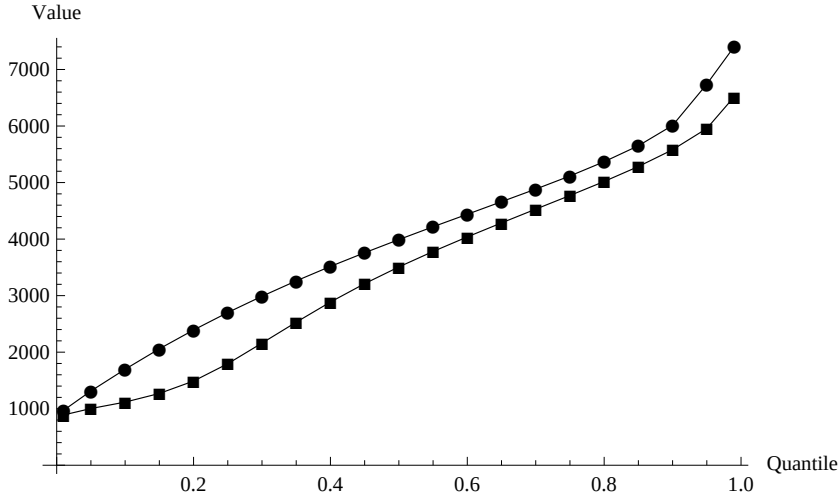

Fig. 8.29. Graph of quantiles of the final value of the portfolio from Analysis II.2 (squares) and Analysis III.1 (circles) (in million $)

The statistical measures of the value of the portfolio if the final ruin occurs can be found in Table 8.13. By comparing them with the similar data for Analysis II.1 (see Table 8.6) and for Analysis II.2 (see Table 8.10), one can see that the final insolvency has a greater impact for the portfolio considered now. The respective histogram is similar to the ones discussed previously.

Due to the applied parameters, the estimated probability of using the catastrophe bond is equal to 19.9891%. We can compare the payments directed to the insurer from the cat bonds for Analysis II.2 and the instance considered here. As it can be seen from Tables 8.11 and 8.14, all statistical measures except for maximum are higher now. A more detailed analysis is shown in Figure 8.30 –

Measure	Value
Minimum	-657341.
First quartile	-17696.4
Median	-7672.96
Third quartile	-2913.22
Maximum	-0.245904
Mean	-15324.21
Std. deviation	26717.1
Skewness	-8.02908
Kurtosis	125.776

Table 8.13. Statistical measures of the portfolio in the case of the final ruin from Analysis III.1

the quantile plot for this case is higher than in Analysis II.2 up to about the 0.9 quantile.

Measure	Value
Minimum	0.0203985
First quartile	2005.54
Median	5057.18
Third quartile	11036.5
Maximum	11089
Mean	5762.33
Std. deviation	4013.49

Table 8.14. Statistical measures of the value of payments from the catastrophe bond for Analysis III.1

Because of the assumed settings, the estimated probability of using funds from the reinsurance contract is equal to 4.9704%. As for the catastrophe bonds, payments from this source are noticeably higher in this case than in Analysis II.1. This fact can be easily seen from the comparison of Table 8.15 with Table 8.8, and from Figure 8.15. Apart from the maximum and some higher ranks of quantiles, the funds from the reinsurance contract used are higher for the considered portfolio. On the other hand, maximum values of payoffs from the catastrophe bond and from the reinsurance contract are lower now – therefore these instruments can be more available for the insurer because of lower potential risks for policyholders and the reinsurer.

Analysis III.2

There are also other approaches to the construction the relevant portfolio which consists of three layers. For example, another number of the catastrophe bonds

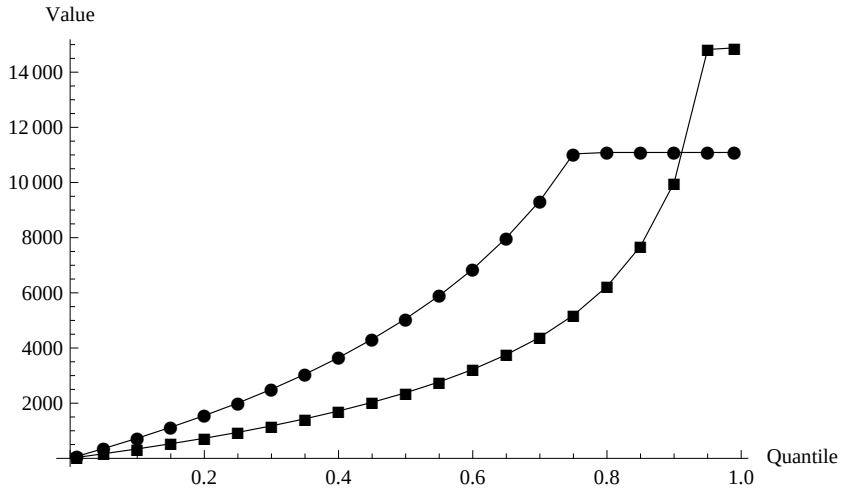

Fig. 8.30. Graph of quantiles of the payments from the catastrophe bond for Analysis II.2 (squares) and Analysis III.1 (circles) (in million $)

Measure	Value
Minimum	0.0203985
First quartile	1197.8
Median	3088.65
Third quartile	7042.49
Maximum	8519.5
Mean	3932.43
Std. deviation	3041.73

Table 8.15. Statistical measures of the value of payments from the reinsurance contract for Analysis III.1

n_{cb} could be used. Then, instead of previous assumption, concerning the fixed sum of premiums p_{cb} and p_{rnsr} (which then turns into the fixed value of n_{cb}), during the following analysis a value of n_{cb} lower than in Analysis III.1, is set in order to analyse the impact of such a setting.

Let us suppose that the number of the sold catastrophe bonds n_{cb} is equal to the difference between the triggering points K_1 and K_0, so that if the cumulated value of losses is higher than K_1, then all of the losses from the gap $K_1 - K_0$ are refunded. On the basis of Monte Carlo simulations, n_{cb} is estimated to be equal to 4593 for this assumption. Then, an analysis similar to the ones performed for the previous cases can be carried out.

Because the number of the catastrophe bonds n_{cb} is lower than in Analysis III.1, the evaluated probability of the final insolvency (0.86%) is higher now, but is still limited and comparable with the relevant value in Analysis II.1. On

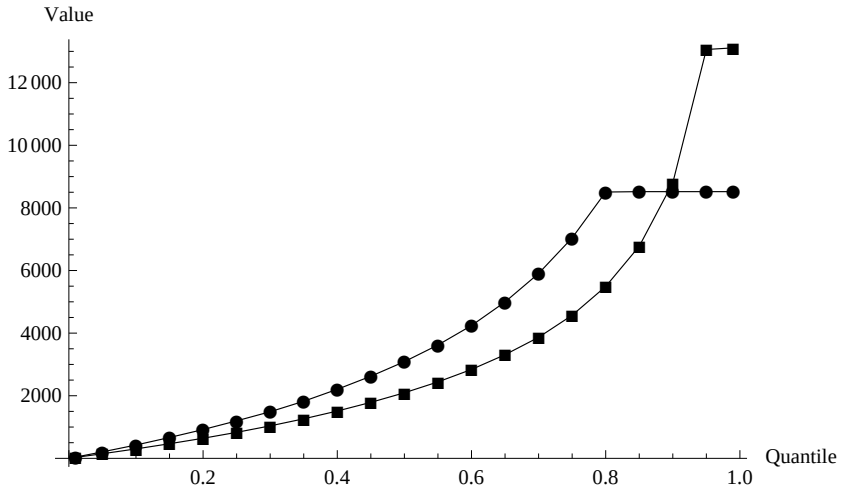

Fig. 8.31. Graph of quantiles of the payments from the reinsurance contract for Analysis II.1 (squares) and Analysis III.1 (circles) (in million $)

the other hand, the estimated probabilities of the earlier ruin (19.9057%) and the earlier ruin which does not lead to the final insolvency (19.9057%) are significantly lesser than in Analysis III.1. It can be easily seen that the additional burden for the insurer at $t = 0$, i.e. the necessary payments related to issuing of the catastrophe bond and entering into the reinsurance contract, exert important influence on these probabilities.

When comparing this instance and Analysis III.1, we can see that all of the statistical measures (see Table 8.16) and the relevant quantile plot (see Figure 8.32) indicate that the final value of the portfolio is lower now. Such behaviour is, of course, related to the applied parameter n_{cb}. This is especially seen for lower and higher quantiles.

Measure	Value
Minimum	-663548
First quartile	1893.1
Median	3798.09
Third quartile	5062.07
Maximum	20002.9
Mean	3503.41
Std. deviation	2959.44

Table 8.16. Statistical measures of the portfolio for $T = 1$ from Analysis III.2

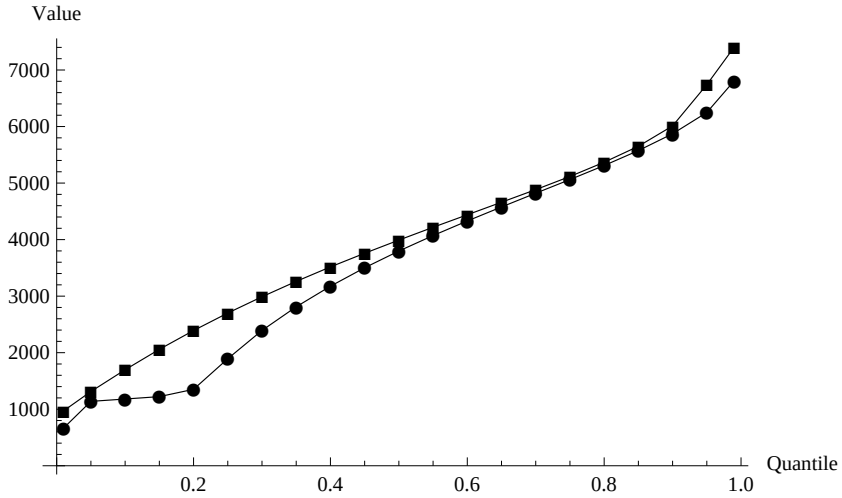

Fig. 8.32. Graph of quantiles of the final value of the portfolio for Analysis III.1 (squares) and Analysis III.2 (circles) (in million $)

As it was noted, the probability of the final ruin in the considered case is significantly higher. However, in the light of data from Table 8.17, the final insolvency has lower impact than in Analysis III.1 (see Table 8.13).

Measure	Value
Minimum	-663548.
First quartile	-13419.2
Median	-5868.27
Third quartile	-2182.69
Maximum	-0.533395
Mean	-11709.4
Std. deviation	20876
Skewness	-9.23511
Kurtosis	179.12

Table 8.17. Statistical measures of the portfolio in the case of the final ruin for Analysis III.2

The time moments of the earlier ruin can also be compared. Figure 8.33 shows that for both Analysis III.1 and the present instance, the overall shapes of the envelopes of the related histograms are highly similar. There are some small differences – because of the higher burden in the beginning, the earlier ruins are more frequent during the initial moments in Analysis III.1.

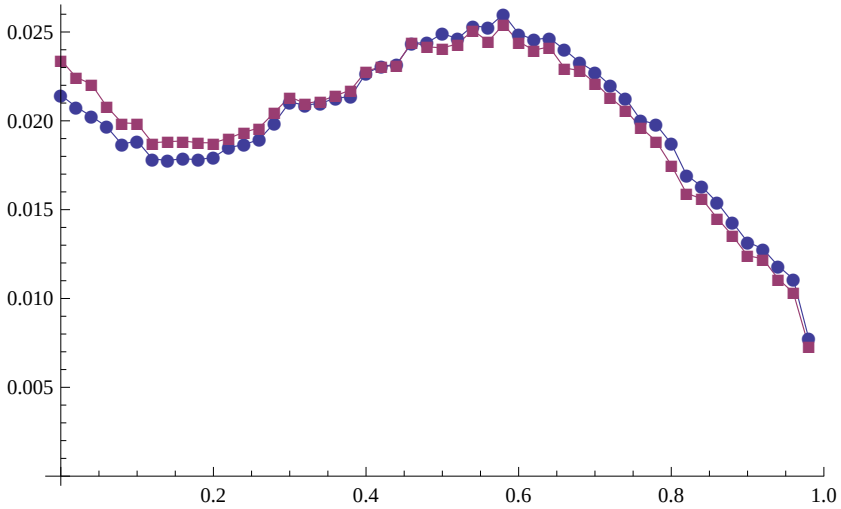

Fig. 8.33. Comparison of envelopes of histograms of the time moments of the earlier ruin for Analysis III.1 (squares) and Analysis III.2 (circles)

Since now the parameter n_{cb} has a lower value, it seems that the possible payments from the catastrophe bond should also be lower. This is confirmed by Table 8.18 and Figure 8.34. Both the statistical measures and the quantile plot take lower values than in the case of Analysis III.1.

Measure	Value
Minimum	0.00844893
First quartile	830.685
Median	2094.66
Third quartile	4571.25
Maximum	4593
Mean	2386.72
Std. deviation	1662.37

Table 8.18. Statistical measures of the value of payments from the catastrophe bond for Analysis III.2

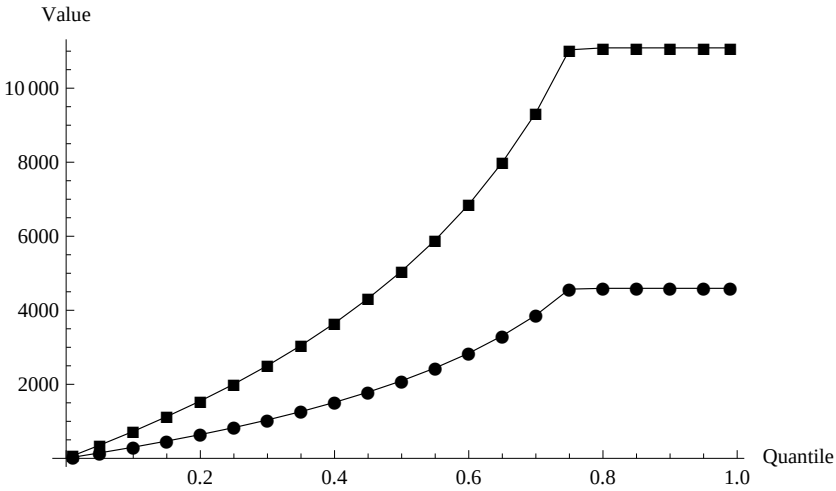

Fig. 8.34. Graph of quantiles of the payments from the catastrophe bond for Analysis III.1 (squares) and Analysis III.2 (circles) (in million $)

References

Albrecher, H., Hartinger, J. & Tichy, R. (2008), 'QMC techniques for CAT bond pricing', *Monte Carlo Methods and Applications* **10**, 197–211.

Asmussen, S. & Albrecher, H. (2010), *Ruin probabilities*, World Scientific Publishing Co. Pte. Ltd.

Atkinson, K. E. (1989), *An Introduction to Numerical Analysis*, John Wiley & Sons, New York.

Baryshnikov, Y., Mayo, A. & Taylor, D. (1998), Pricing of CAT Bonds, Working paper.

Berdnikov, A. S., Compagner, A. & Turtia, S. B. (1996), 'A MathLink program for high-quality random nubers', *The Mathematica Journal* **6**(3), 65–69.

Besag, J. (1974), 'Spatial interaction and the statistical analysis of lattice systems (with discussion)', *J. Royal Statist. Soc. Series B* **36**, 192–326.

Besag, J. E. (1986), 'On the statistical analysis of dirty pictures (with discussion)', *J. Royal Statist. Soc. Series B* **48**, 259–302.

Billingsley, P. (1995), *Probability and Measure*, Wiley New York.

Binder, K. (1988), *Monte Carlo Simulations in Statistical Physics: An Introduction*, Springer.

Blum, L., Blum, M. & Shub, M. (1986), 'A simple unpredictible pseudo-random number generator', *SIAM Journal on Computing* **15**, 364–383.

Bodoff, N. & Gan, Y. (2009), 'Analysis of the market price of cat bonds', *Casualty Actuarial Society E-Forum*, 1–26.

Borch, K. (1974), *The Mathematical Theory of Insurance*, Lexington Books, Lexington.

Box, G. E. P. & Muller, M. E. (1958), 'A note on the generation of random normal deviates', *Ann. Math. Statist.* **29**, 610–611.

Brémaud, P. (1999), *Markov Chains. Gibbs Fields, Monte Carlo Simulation and Queues*, Springer.

Burnecki, K., Kukla, G. & Weron, R. (2000), 'Property insurance loss distributions', *Physica A* **287**, 269–278.

Burnecki, K. & Weron, R. (2005), Modeling of the risk process, *in* P. Cizek, W. Härdle & R. Weron, eds., 'Statistical tools for finance and insurance', Springer, 319–339.

Canabarro, E., Finkemeier, M., Anderson, R. & Bendimerad, F. (2000), 'Analyzing insurance-linked securities', *Journal of Risk Finance* **1**, 49–78.

Capasso, V. & Bakstein, D. (2012), *An Introduction to Continuous–Time Stochastic Processes*, Birkhäuser, Boston.

Casella, G. & Robert, C. (1996), 'Rao-Blackwellisation of sampling schemes', *Biometrika* **83**(1), 81–94.

Chan, K. C., Karolyi, G. A., Longstaff, F. A. & Sanders, A. B. (1992), 'An empirical comparison of alternative models of the short-term interest rate', *The Journal of Finance* **47**, 1209–1227.

Chernobai, A., Burnecki, K., Rachev, S., Trueck, S. & Weron, R. (2006), 'Modeling catastrophe claims with left-truncated severity distributions', *Computational Statistics* **21**, 537–555.

Cipra, B. (1987), 'An introduction to the Ising model', *American Mathematical Monthly* **94**, 937–959.

Cox, S. H., Fairchild, J. R. & Pedersen, H. W. (2000), 'Economic aspects of securitization of risk', *ASTIN Bulletin* **30**, 157–193.

Cox, S. H., Fairchild, J. R. & Pedersen, H. W. (2004), 'Valuation of structured risk management products', *Insurance: Mathematics and Economics* **34**, 259–272.

Cox, S. & Pedersen, H. (2000), 'Catastrophe risk bonds', *North American Actuarial Journal* **4**, 56–82.

Cummins, J. D., Doherty, N. & Lo, A. (2002), 'Can insurers pay for the "big one"? measuring the capacity of insurance market to respond to catastrophic losses', *Journal of Banking and Finance* **26**, 557–583.

Cummins, J. D., Lalonde, D. & Phillips, R. D. (2002), Managing risk using index-linked catastrophic loss securities, *in* M. Lane, ed., 'Alternative Risk Strategies', Risk Books.

Cummins, J. D., Lalonde, D. & Phillips, R. D. (2004), 'The basis risk of catastrophic-loss index securities', *Journal of Financial Economics* **71**, 77–111.

Dai, Q. & Singleton, K. (2000), 'Specification analysis of affine term structure models', *Journal of Finance* **55**, 1943–1978.

Davies, R. (2000), 'Hardware random number generators'. `http://www.robertnz.net/hwrng.htm`.

Dempster, A., Laird, N. & Rubin, D. (1977), 'Maximum likelihood from incomplete data via the EM algorithm (with discussion)', *J. Royal Statist. Soc. Series B* **39**, 333–350.

Devroye, L. (1986), *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.

Doyle, P. G. (2006), 'Grinstead and Snell's introduction to probability'. `http://www.math.dartmouth.edu/~prob/prob/prob.pdf`.

Duffie, D. & Kan, R. (1996), 'A yield-factor model of interest rates', *Mathematical Finance* **6**, 379–406.

Eckhardt, R. (1987), 'Stan Ulam, John von Neumann, and the Monte Carlo method', *Los Alamos Science, Special Issue* **15**, 131–137.

Egamia, M. & Young, V. (2008), 'Indifference prices of structured catastrophe (cat) bonds', *Insurance: Mathematics and Economics* **42**, 771–778.

Eichenauer-Hermann, J. (1993), 'Statistical independence of a new class of inverse congruential psuedorandom numbers', *Math. Comput.* **60**, 375–384.

Eichenauer, J. & Lehn, J. (1986), 'A non-linear congruential pseudo random number generator', *Statist. Pap.* **27**, 315–326.

Ermoliev, Y. M., Ermolieva, T. Y., McDonald, G. & Norkin, V. (2001), 'Problems on insurance of catastrophic risks', *Cybernetics and Systems Analysis* **37**, 220–234.

Ermolieva, T., Romaniuk, M., Fischer, G. & Makowski, M. (2007), Integrated model-based decision support for management of weather-related agricultural losses, *in* O. Hryniewicz, J. Studzinski & M. Romaniuk, eds., 'Environmental informatics and systems research. Vol. 1: Plenary and session papers – EnviroInfo 2007', Shaker Verlag.

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications, Vol. 1*, Wiley, New York.

Filipovic, D., Cheridito, P. & Kimmel, R. (2007), 'Market price of risk specifications for affine models: Theory and evidence', *Journal of Financial Economics* **83**, 123–170.

Finken, S. & Laux, C. (2009), 'Catastrophe bonds and reinsurance: the competitive effect of information-insensitive triggers', *The Journal of Risk and Insurance* **76**, 579–605.

Fishman, G. S. (1973), *Concepts and Methods in Discrete Event Digital Simulation*, John Wiley, New York.

Freeman, P. K. & Kunreuther, H. (1997), *Managing Environmental Risk Through Insurance*, Kluwer Academic Press, Boston.

Froot, K. A. (2001), 'The market for catastrophe risk: A clinical examination', *Journal of Financial Economics* **60**, 529–571.

Gelman, A. & Rubin, D. (1992), 'Inference from iterative simulation using multiple sequences (with discussion)', *Statis. Science* **7**, 457–511.

Geman, S. & MacClure, D. E. (1987), 'Statistical methods for tomographic image reconstruction', *Bull. Int. Statist. Inst.* **LII-4**, 5–21.

George, J. B. (1997), 'Alternative reinsurance: Using catastrophe bonds and insurance derivatives as a mechanism for increasing capacity in the insurance markets', *CPCU Journal* .

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.

Glasserman, P. (2004), *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, New York.

Golomb, S. W. (1967), *Shift Register Sequences*, Holden-Day, San Francisco.

Graham, R., Knuth, D. & Patashnik, O. (1994), *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley.

Harrington, S. E., Mann, S. V. & Niehaus, G. (1995), 'Insurer capital structure decisions and the viability of insurance derivatives', *Journal of Risk and Insurance* **62**, 483–508.

Harrington, S. E. & Niehaus, G. (2003), 'Capital, corporate income taxes, and catastrophe insurance', *Journal of Financial Intermediation* **12**, 365–389.

Hull, T. E. & Dobel, A. R. (1962), 'Random number generators', *Society for Industrial and Applied Mathematics Review*, 230–254.

Jacod, J. & Shiryaev, A. (2003), *Limit Theorems for Stochastic Processes*, Springer-Verlag.

Jakubowski, J., Palczewski, A., Rutkowski, M. & Stettner, L. (2004), *Matematyka finansowa instrumenty pochodne*, Wydawnictwa Naukowo-Techniczne.

Jakubowski, J. & Sztencel, R. (2000), *Wstęp do teorii prawdopodobieństwa*, Script, Warszawa.

Jansson, B. (1966), *Random Number Generators*, Victor Pettersons Bokindustri Aktiebolag, Stockholm.

Jensen, M. C. (1986), 'Agency cost of free cash flow, corporate finance, and takeovers', *American Economic Review* **76**.

Jones, G. & Hobert, J. (2001), 'Honest exploration of intractable probability distributions via Markov chain Monte Carlo', *Statist. Science* **16**(4), 312–334.

Kai, Y., Zhong-ying, Q. & Shang-zhi, Y. (2007), Study on catastrophe bond pricing model based on behavioral finance, *in* 'Management Science and Engineering, 2007. ICMSE 2007. International Conference on', 1725–1736.

Karlin, S. & Taylor, H. M. (1975), *A First Course in Stochastic Processes*, Academic Press.

Kinderman, A. J. & Monahan, J. F. (1977), 'Computer generation of random variables using the ratio of uniform deviates', *ACM Transactions on Mathematical Software* **3**, 257–260.

Kirkpatrick, S. & Stoll, E. P. (1981), 'A very fast shift-register sequence random number generator', *J. Computational Physics* **40**(2), 517–526.

Knuth, D. E. (1966), *The Art of Computer Progamming. Vol 2: Seminumerical Algorithms*, Addison-Wesley, Reading, MA.

Kwapień, S. & Woyczyński, W. A. (1991), 'Semimartingale integrals via decoupling inequalities and tangent processes', *Probab. Math. Statist.* **12**, 165–200.

Kwok, Y. (2008), *Mathematical Models of Financial Derivatives*, Springer-Verlag, Berlin Heildelberg.

Law, A. M. (2007), *Simulation Modeling & Analysis*, McGraw-Hill, New York.

L'Ecuyer, P. & Simard, R. (2001), 'On the performance of birthday spacings tests with certain families of random number generators', *Math. Comput. Simul.* **55**(1-3), 131–137.

L'Ecuyer, P. & Simard, R. (2007), 'Testu01: A C library for empirical testing of random number generators', *ACM Transactions on Mathematical Software* **33**(7), 22–40.

Lee, J. P. & Yu, M. T. (2007), 'Valuation of catastrophe reinsurance with catastrophe bonds', *Insurance: Mathematics and Economics* **41**, 264–278.

Lehmer, D. H. (1951), Mathematical methods in large-scale computing units, *in* 'Proceedings of a Second Symposium on Large-Scale Digital Calculating Machinery, 1949', Vol. 26, Harvard University Press, 141–146.

Lewis, P. A. W. & Shedler, G. S. (1979), 'Simulation of nonhomogeneous Poisson process by thinning', *Nav. Res. Logist. Quart.* **26**, 403–413.

Lewis, T. G. & Payne, W. H. (1973), 'Generalized feedback shift register pseudorandom number algorithms', *J. ACM* **20**, 47–54.

Lin, S., Shyu, D. & Chang, C. (2008), 'Pricing Catastrophe Insurance Products in Markov Jump Diffusion Models', *Journal of Financial Studies* **16**, 1–33.

Linna, R. (2012), 'Becs-114.1100 computational science'. `http://www.lce.hut.fi/teaching/S-114.1100/`.

Ma, Z.-G. & Ma, C.-Q. (2013), 'Pricing catastrophe risk bonds: A mixed approximation method', *Insurance: Mathematics and Economics* **52**, 243–254.

Maclaren, N. M. (1992), 'A limit on the usable length of a pseudorandom sequence', *J. Statist. Comput. Sim.* **42**, 47–54.

Marsaglia, G. (1977), 'The squeeze method for generating gamma variates', *Comp. Math. App.* **3**, 321–325.

Marsaglia, G. (1984), A current view of random number generators, *in* 'Computer Science and Statistics: 16th Symposium in the Interface, Atlanta'.

Marsaglia, G. (1995), 'Diehard battery of tests of randomness'. `http://www.stat.fsu.edu/pub/diehard/`.

Marsaglia, G. & Bray, T. A. (1964), 'A convenient method for generating normal variables', *SIAM Rev.* **6**, 260–264.

Marsaglia, G. & Tsay, L. H. (1985), 'Matrices and the structure of random number sequences', *Linear Alg. App.* **67**, 147–156.

Matsumoto, M. & Nishimura, T. (1998), 'Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator', *ACM Trans. Model. Comput. Simul.* **8**(1), 3–30.

McGhee, C. (2004), *Market Update: The Catastrophe Bond Market at Year-End 2003*, Guy Carpenter & Company, Inc. and MMC Security Corporation.

Merton, R. (1976), 'Option pricing when underlying stock returns are discontinuous', *Journal of Financial Economics* **3**, 125–144.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. (1953), 'Equations of state calculations by fast computing machines', *J. Chem. Phys.* **21**, 1087–1092.

Metropolis, N. & Ulam, S. (1949), 'The Monte Carlo method', *Journal of American Statistical Association* **44**(247), 335–341.

Mikosch, T. (2009), *Non-Life Insurance Mathematics*, Springer-Verlag, Berlin Heidelberg.

Muermann, A. (2008), 'Market price of insurance risk implied by catastrophe derivatives', *North American Actuarial Journal* **12**, 221–227.

Munk, C. (2011), *Fixed Income Modelling*, Oxford University Press.

Nowak, P. (1999), Analysis of applications of some ex-ante instruments for the transfer of catastrophic risks, Technical Report IR-99-075, IIASA.

Nowak, P. (2002), 'On Jacod-Grigelionis characteristics for Hilbert space valued semimartingales', *Stochastic Analysis and Applications* **20**, 963–998.

Nowak, P. (2003), 'Integration with Respect to Hilbert Space-Valued Semimartingales via Jacod–Grigelionis Characteristics', *Stochastic Analysis and Applications* **21**(5), 1141–1168.

Nowak, P. (2011), Option pricing with Lévy process in a fuzzy framework, *in* K. Atanassov, W. Homenda, O. Hryniewicz, J. Kacprzyk, M. Krawczak, Z. Nahorski, E. Szmidt & S. Zadrozny, eds., 'Recent Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics', Polish Academy of Sciences Warsaw, 155–167.

Nowak, P. & Romaniuk, M. (2009), Portfolio of financial and insurance instruments for losses caused by natural catastrophes, *in* Z. Wilimowska, L. Borzemski, A. Grzech & J. Świątek, eds., 'Information Systems Architecture and Technology. IT Technologies in Knowledge Oriented Management Process', Wrocław University of Technology, 27–36.

Nowak, P. & Romaniuk, M. (2010*a*), 'Analiza własności portfela złożonego z instrumentów finansowych i ubezpieczeniowych', *Studia i materiały polskiego stowarzyszenia zarządzania wiedzą* **31**, 65–76.

Nowak, P. & Romaniuk, M. (2010*b*), 'Computing option price for Lévy process with fuzzy parameter', *European Journal of Operational Research* **201**, 206–210.

Nowak, P. & Romaniuk, M. (2010*c*), Fuzzy approach to evaluation of portfolio of financial and insurance instruments, *in* K. T. Atanassov, O. Hryniewicz, J. Kacprzyk, M. Krawczak, Z. Nahorski, E. Szmidt & S. Zadrozny, eds., 'Developments in Fuzzy Sets, Intuitionistics Fuzzy Sets, Generalized Nets and Related Topics', Warsaw, 189–205.

Nowak, P. & Romaniuk, M. (2010*d*), On analysis and pricing of integrated multilayer insurance portfolio, *in* Z. Wilimowska, L. Borzemski, A. Grzech & J. Świątek, eds., 'Information Systems Architecture and Technology', Wrocław University of Technology, 339–349.

Nowak, P. & Romaniuk, M. (2011), On pricing and numerical simulations of some catastrophe bond, *in* J. Świątek, L. Borzemski, A. Grzech & Z. Wilimowska, eds., 'Systems Architecture and Technology. System Analysis Approach to the Design, Control and Decision Support', Wrocław University of Technology, 99–108.

Nowak, P. & Romaniuk, M. (2013*a*), 'A fuzzy approach to option pricing in a Lévy process setting', *International Journal of Applied Mathematics and Computer Science* **23**, 613 – 622.

Nowak, P. & Romaniuk, M. (2013*b*), 'Pricing and simulations of catastrophe bonds', *Insurance: Mathematics and Economics* **52**, 18–28.

Nowak, P. & Romaniuk, M. (2013*c*), Pricing of catastrophe bond in fuzzy framework, *in* C. Borgelt, M. A. Gil, J. M. C. Sousa & M. Verleysen, eds., 'Towards Advanced Data Analysis by Combining Soft Computing and Statistics', Springer-Verlag, Berlin Heidelberg, 137–150.

Nowak, P. & Romaniuk, M. (2014*a*), 'Application of Lévy processes and Esscher transformed martingale measures for option pricing in fuzzy framework', *Journal of Computational and Applied Mathematics* **263**, 129 – 151.

Nowak, P. & Romaniuk, M. (2014*b*), 'Application of the one-factor CIR interest rate model to catastrophe bond pricing under uncertainty', *Journal of Automation, Mobile Robotics & Intelligent Systems* **8**, 19–27.

Nowak, P. & Romaniuk, M. (2014*c*), Catastrophe bond pricing with fuzzy volatility parameters, *in* L. T. Koczy, C. R. Pozna & J. Kacprzyk, eds., 'Issues and Challenges of Intelligent Systems and Computational Intelligence', Vol. 530 of *Studies in Computational Intelligence*, Springer International Publishing, 27–44.

Nowak, P., Romaniuk, M. & Ermolieva, T. (2011), Evaluation of portfolio of financial and insurance instruments – simulation of uncertainty, *in* Y. Ermoliev, M. Makowski & K. Marti, eds., 'Managing Safety of Heterogeneous Systems: Decisions under Uncertainties and Risks', Springer-Verlag, Berlin Heidelberg.

O'Brien, T. (1997), 'Hedging strategies using catastrophe insurance options', *Insurance: Mathematics and Economics* **21**, 153–162.

Odeh, R. E. & Evans, J. O. (1974), 'The percentage points of the normal distribution', *Appl. Statist.* **23**, 96–97.

Philippe, A. (1997), Importance sampling and Riemann sums, Technical report VI, Univ. de Lille. Prepub. IRMA.

Philippe, A. & Robert, C. (2001), 'Riemann sums for MCMC estimation and convergence monitoring', *Statistics and Computing* **11**(1), 103–115.

Protter, P. (2005), *Stochastic Integration and Differential Equations*, Springer-Verlag, Berlin Heidelberg.

Reshetar, G. (2008), Pricing of multiple-event coupon paying cat bond, Working paper, Swiss Banking Institute, University of Zurich.

Revuz, D. & Yor, M. (1999), *Continuous martingales and Brownian motion*, Springer-Verlag, Berlin Heidelberg.

Ripley, B. D. (1987), *Stochastic Simulation*, John Wiley & Sons.

Ripples Into Waves: The Catastrophe Bond Market at Year-End 2006 (2007), Guy Carpenter & Company, Inc. and MMC Security Corporation.

Robert, C. (1995), 'Convergence control techniques for MCMC algorithms', *Statis. Science* **10**(3), 231–253.

Robert, C. P. & Casella, G. (2004), *Monte Carlo Statistical Methods*, Springer, New York.

Romaniuk, M. (2007), 'On some method for diagnosing convergence in MCMC setups via atoms and renewal sets', *Control and Cybernetics* **36**, 985–1008.

Romaniuk, M. (2008), 'Convergence diagnosis to stationary distribution in MCMC methods via atoms and renewal sets', *Control and Cybernetics* **37**, 205–229.

Romaniuk, M. (2015), Application of Markov chain and interest rate process for forecasting of costs of maintenance of pipelines, *in* J. Wittmann & R. Wieland, eds., 'Simulation in Umwelt- und Geowissenschaften. Workshop Müncheberg 2015', Umweltinformatik, Shaker Verlag Aachen, 147–156.

Romaniuk, M. & Ermolieva, T. (2005), 'Application of EDGE software and simulations for integrated catastrophe management', *International Journal of Knowledge and Systems Sciences* **2**, 1–9.

Rubinstein, R. (1981), *Simulation and the Monte Carlo Method*, John Wiley, New York.

Saito, M. & Matsumoto, M. (2008), SIMD-oriented fast Mersenne Twister: a 128-bit pseudorandom number generator, *in* A. Keller, S. Heinrich & H. Niederreiter, eds., 'Monte Carlo and Quasi-Monte Carlo Methods 2006', Springer, Berlin Heidelberg, 607–622.

Schmidt, T. (2006), 'Coping with copulas'. `http://www.math.uni-leipzig.de/~tschmidt/TSchmidt_Copulas.pdf`.

Schukken, Y., Casella, G. & van den Broek, J. (1991), 'Overdispersion in clinical mastitis data from dairy herds: a negative binomial approach', *Preventive Veterinary Medicine* **10**, 239–245.

Shiryaev, A. (1999), *Essentials of stochastic finance, Facts, Models, Theory*, World Scientific.

Taussky, O. & Todd, J. (1956), Generation and testing of pseudo-random numbers, *in* H. A. Meyer, ed., 'Symposium on Monte Carlo Methods', Wiley, Chapman & Hall.

Tausworthe, R. C. (1965), 'Random number generated by linear reccurence modulo two', *Math. Comp.* **19**, 201–209.

Thomas, A., Spiegelhalter, D. J. & Gilks, W. R. (1992), BUGS: a program to perform bayesian inference using Gibbs sampling, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds., 'Bayesian Statistics 4', Clarendon Press, 837–842.

Thompson, J. R. (2000), *Simulation – A Modeler's Approach*, John Wiley & Sons, New York.

Vaugirard, V. E. (2003), 'Pricing catastrophe bonds by an arbitrage approach', *The Quarterly Review of Economics and Finance* **43**, 119–132.

von Neumann, J. (1951), Various techniques used in connection with random digits, *in* A. Householder, G. Forsythe & H. Germond, eds., 'Monte Carlo Method', Vol. 12, National Bureau of Standards Applied Mathematics Series, 36–38.

Walker, A. J. (1977), 'An efficient method for generating discrete random variables with general distributions', *ACM Trans. Math. Soft.* **3**, 253–256.

Walker, G. (1997), Current developments in catastrophe modelling, *in* N. R. Britton & J. Olliver, eds., 'Financial Risks Management for Natural Catastrophes', Griffith University.

Wang, S. (2004), 'Cat bond pricing using probability transforms', *Geneva Papers, Special issue on Insurance and the State of the Art in Cat Bond Pricing* **278**, 19–29.

Wieczorkowski, R. & Zieliński (1997), *Komputerowe generatory liczb losowych*, WNT, Warszawa.

Wu, H.-C. (2004), 'Pricing European options based on the fuzzy pattern of Black-Scholes formula', *Computers & Operations Research* **31**, 1069–1081.

Yu, B. & Mykland, P. (1998), 'Looking at Markov samplers through cusum path plots: A simple diagnostic idea', *Statistics and Computing* **8**(3), 275–286.

Zadeh, L. (1965), 'Fuzzy sets', *Information and Control* **8**, 338–353.

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

The Project is co-financed by the European Union from resources of the European Social Fund

ISBN 978-83-63159-14-6
e-ISBN 978-83-63159-15-3