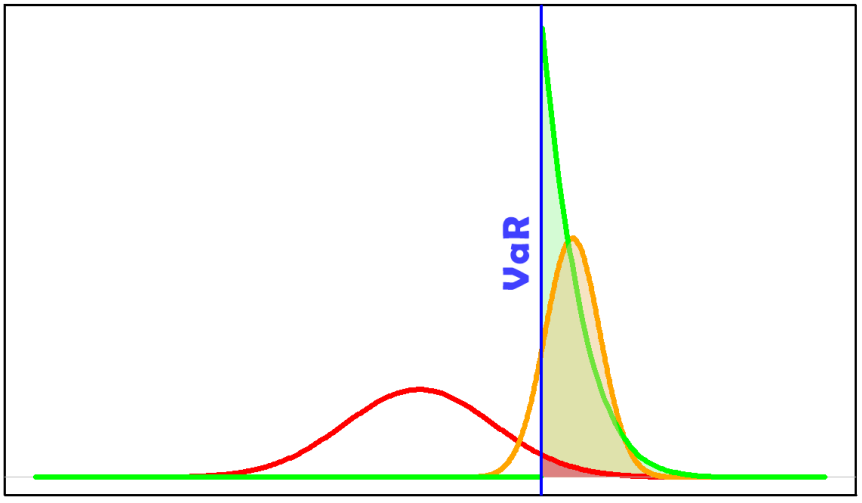
TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Variance Reduction Procedures for Market Risk Estimation

Mykhailo Pupashenko



Datum der Disputation: 29. September 2014

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern zur Verleihung des akademischen Grades Doktor der Naturwissenschaften (Doctor rerum naturalium, Dr. rer. nat.) genehmigte Dissertation.

1. Gutachter: Prof. Dr. Ralf Korn
2. Gutachter: Prof. Dr. Andreas Neuenkirch

D 386

"You have to take risks. We will only understand the miracle of life fully when we allow the unexpected to happen."

Paulo Coelho

Abstract

Monte Carlo simulation is one of the commonly used methods for risk estimation on financial markets, especially for option portfolios, where any analytical approximation is usually too inaccurate. However, the usually high computational effort for complex portfolios with a large number of underlying assets motivates the application of variance reduction procedures. Variance reduction for estimating the probability of high portfolio losses has been extensively studied by Glasserman et al. A great variance reduction is achieved by applying an exponential twisting importance sampling algorithm together with stratification. The popular and much faster Delta-Gamma approximation replaces the portfolio loss function in order to guide the choice of the importance sampling density and it plays the role of the stratification variable. The main disadvantage of the proposed algorithm is that it is derived only in the case of Gaussian and some heavy-tailed changes in risk factors.

Hence, our main goal is to keep the main advantage of the Monte Carlo simulation, namely its ability to perform a simulation under alternative assumptions on the distribution of the changes in risk factors, also in the variance reduction algorithms. Step by step, we construct new variance reduction techniques for estimating the probability of high portfolio losses. They are based on the idea of the Cross-Entropy importance sampling procedure. More precisely, the importance sampling density is chosen as the closest one to the optimal importance sampling density (zero variance estimator) out of some parametric family of densities with respect to Kullback - Leibler cross-entropy. Our algorithms are based on the special choices of the parametric family and can now use any approximation of the portfolio loss function. A special stratification is developed, so that any approximation of the portfolio loss function under any assumption of the distribution of the risk factors can be used. The constructed algorithms can easily be applied for any distribution of risk factors, no matter if light- or heavy-tailed. The numerical study exhibits a greater variance reduction than of the algorithm from Glasserman et al. The use of a better approximation may improve the performance of our algorithms significantly, as it is shown in the numerical study.

The literature on the estimation of the popular market risk measures, namely VaR and CVaR, often refers to the algorithms for estimating the probability of high portfolio losses, describing the corresponding transition process only briefly. Hence, we give a consecutive discussion of this problem. Results necessary to construct confidence intervals for both measures under the mentioned variance reduction procedures are also given.

Acknowledgements

First of all I would like to thank my supervisor, Prof. Dr. Ralf Korn, for the opportunity to be one of his PhD students and for his constant support throughout my studies. For his help and cooperation in our joint papers I am also very thankful.

I would also like to express my gratitude to my colleagues in the project 'Robust Risk Estimation', especially Dr. habil. Peter Ruckdeschel, Prof. Dr. Matthias Kohl, Dr. Bernhard Spangl and M.Sc. Daria Pupashenko, as well as all others whom I did not mention by name. I am also very thankful to the Volkswagen Stiftung for their financial support.

Further, I would like to express my appreciation to Dr. habil. Christoph Lossen (Department Manager) and Dr. Falk Triebsch (Manager of the Graduate School) for their helpful advices during my studies. A special thank goes to Jessica Borsche for her constant support and her helpfulness.

I deeply wish to thank the International School for Gradate Studies (ISGS), especially Dipl.-Math. Arthur Harutyunyan, for making my life as a foreign student in Kaiserslautern much easier.

I would also like to express my deepest gratitude to my family for constantly supporting me and believing in me, my sister, Daria Pupashenko, for her patience and my wife, Olga Pupashenko, for her love and for encouraging me in becoming better.

Contents

List of Figures

List of Tables

Abbreviations

ACE	**A**pproximate **C**ross-**E**ntropy algorithm
ACES	**A**pproximate **C**ross-**E**ntropy with **S**tratification algorithm
ACESI	**A**pproximate **C**ross-**E**ntropy with **S**tratification in **I**deal setting algorithm
AICE	**A**pproximate **I**mproved **C**ross-**E**ntropy algorithm
AICES	**A**pproximate **I**mproved **C**ross-**E**ntropy with **S**tratification algorithm
AICESI	**A**pproximate **I**mproved **C**ross-**E**ntropy with **S**tratification in **I**deal setting algorithm
AOIS	**A**pproximate **O**ptimal **I**mportance **S**ampling algorithm
AOISI	**A**pproximate **O**ptimal **I**mportance **S**ampling in **I**deal setting algorithm
AOISS	**A**pproximate **O**ptimal **I**mportance **S**ampling with **S**tratification algorithm
AOISSI	**A**pproximate **O**ptimal **I**mportance **S**ampling with **S**tratification in **I**deal setting algorithm
ATM	**A**t **T**he **M**oney
CE	**C**ross-**E**ntropy algorithm
CVaR	**C**onditional **V**alue-**a**t-**R**isk
ES	**E**xpected **S**ortfall
ETL	**E**xpected **T**ail **L**oss
FCE	**F**ull **C**ross-**E**ntropy algorithm
FICE	**F**ull **I**mproved **C**ross-**E**ntropy algorithm

HITRO	Hit-and-Run sampler with Ratio-of-Uniforms
ICE	**I**mproved **C**ross-**E**ntropy algorithm
IS	**I**mportance **S**ampling algorithm
ISS	**I**mportance **S**ampling with **S**tratification algorithm
MC	**M**onte **C**arlo
MCMC	**M**arkov **C**hain **M**onte **C**arlo
TCE	**T**ail **C**onditional **E**xpectation
VaR	**V**alue-**a**t-**R**isk
VM	**V**ariance **M**inimization algorithm
WCE	**W**orst **C**onditional **E**xpectation

a.s.	**a**lmost **s**urely
e.g.	for example
i.e.	that is
i.i.d.	**i**ndependent **i**denticaly **d**istributed
w.r.t.	**w**ith **r**espect **t**o

Notations

x, γ	non-random numbers
$\mathbf{x}$	non-random vector
X	random number
$\mathbf{X}, \mathbf{Y}, \mathbf{Z}$	random vectors
$f(\mathbf{x}; \mathbf{u})$	density function of $\mathbf{X}$
$\mathbf{u}$	original vector parameter of the density of $\mathbf{X}$
$\mathbf{v}, \mathbf{w}$	reference parameter in CE algorithm
$\mathbf{v}^*$	optimal reference parameter in CE algorithm
$\hat{\mathbf{v}}^*$	estimate of the optimal reference parameter in CE algorithm
$f_{\boldsymbol{\theta}}(\mathbf{x}; \mathbf{u}), f_\theta(\mathbf{x}; \mathbf{u})$	exponential change of measures for $f(\mathbf{x}; \mathbf{u})$
$g^*(\mathbf{x})$	optimal importance sampling density
$g_Q^*(\mathbf{x}, \mathbf{v})$	change of measure of the AOIS and AOISS algorithms
$I_A(\mathbf{x})$	indicator function, i.e. it equals 1 if $\mathbf{x} \in A$ and 0 otherwise
t	starting time for market risk estimation
$\mathbf{S} = \mathbf{S}(t)$	vector of market prices (risk factors) at time t
d	dimension of the vector of market prices
Δt	risk-measurement horizon
$\Delta \mathbf{S}$	change in market prices (risk factors) over the horizon Δt
$V(\mathbf{S}, t) = V(\mathbf{S}(t), t)$	portfolio value at time t and market price $\mathbf{S}$ at time t
ΔV	$= V(\mathbf{S} + \Delta \mathbf{S}, t + \Delta t) - V(\mathbf{S}, t)$, changes in portfolio value
$L(\Delta \mathbf{S}) = L(\Delta \mathbf{S}, \Delta t)$	$= -\Delta V$, portfolio losses over the time horizon Δt
Q	approximation function of losses, often Delta-Gamma
Q_x	indirect Delta-Gamma approximation function of losses
$N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	d-dimensional Normal distribution
$N_d(\mathbf{0}, \mathbf{I})$	standard d-dimensional Normal distribution

$t_d(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$	d-dimensional Student t-distribution
T	maturity of options (expiration date)
v_α	portfolio VaR with confidence level α
c_α	portfolio CVaR with confidence level α
$\mathcal{G} = \{f(\mathbf{x}; \mathbf{v}), \mathbf{v} \in \mathcal{V}\}$	parametric family of $\mathbf{X}$
$\mathcal{V}$	parametric set of $\mathbf{X}$
$\mathcal{D}$	Kullback-Leibler cross-entropy
$\mathbb{E}$	expectation w.r.t. original density
$\mathbb{E}_g, \mathbb{E}_h$	expectation w.r.t. other densities $g(\mathbf{x})$ and $h(\mathbf{x})$ respectively
$\mathbb{E}_\mathbf{u}, \mathbb{E}_\mathbf{v}$	expectation w.r.t. densities $f(\mathbf{x}; \mathbf{u})$ and $f(\mathbf{x}; \mathbf{v})$ respectively
$\mathbb{E}_{\boldsymbol{\theta}}, \mathbb{E}_\theta$	expectation w.r.t. densities $f_{\boldsymbol{\theta}}(\mathbf{x}; \mathbf{u})$ and $f_\theta(\mathbf{x}; \mathbf{u})$ respectively
$\mathbb{E}^*_{\mathbf{u},Q}$	expectation w.r.t. density $g^*_Q(\mathbf{x}; \mathbf{u})$
$\mathbb{P}$	probability measure
$\mathbb{P}_\theta$	probability measure w.r.t. density $f_\theta(\mathbf{x}; \mathbf{u})$
$\mathbb{P}_\mathbf{v}$	probability measure w.r.t. density $f(\mathbf{x}; \mathbf{v})$
$\mathbb{P}^*_{\mathbf{u},Q}$	probability measure w.r.t. density $g^*_Q(\mathbf{x}; \mathbf{u})$
$\mathbb{V}\mathbf{ar}$	variance
$\mathbb{R}$	real line
$\mathbb{R}_{\Delta\mathbf{S}}$	support of the distribution of $\Delta\mathbf{S}$
$\mathbb{F}(x) = \mathbb{P}(L \leq x)$	cumulative distribution function of L
$\hat{\mathbb{F}}_N(x)$	empirical distribution function of L
$\mathbf{x}^T$	transposed vector
$\nabla\phi(\boldsymbol{\theta})$	gradient of a vector function
$\hat{\mathbf{v}}$	estimator of the parameter $\mathbf{v}$
$\frac{\partial L(\mathbf{S},t)}{\partial t}$	partial derivative
$\mathrm{diag}(\lambda_i)$	diagonal matrix with λ_i being on the diagonal
$Re(\cdot)$	real part of complex number
$Im(\cdot)$	imaginary part of complex number
$O(\cdot)$	big O notation (asymptotic upper bound)
$\xrightarrow{\mathbb{P}}$	convergence in probability

$\overset{\text{a.s.}}{\to}$	almost sure convergence (with probability 1)
$\overset{\mathbb{D}}{\to}$	convergence in distribution
$\phi(\boldsymbol{\theta})$	moment generating function
$\phi_x(\theta)$	moment generating function of Q_x
$\hat{\phi}(\boldsymbol{\theta})$	characteristic function
$\hat{\phi}_\theta(u)$	characteristic function w.r.t. densities $f_\theta(\mathbf{x};\mathbf{u})$
$\psi(\boldsymbol{\theta})$	cumulant generating function
$\psi_\theta(u)$	cumulant generating function w.r.t. densities $f_\theta(\mathbf{x};\mathbf{u})$
ρ	rarity parameter in CE algorithm
$\boldsymbol{\Sigma}_S$	covariance matrix of $\boldsymbol{\Delta}\mathbf{S}$
$\boldsymbol{\delta}$	Delta of a portfolio value V, d-dimensional vector
$\boldsymbol{\Gamma}$	Gamma of a portfolio value V, $d \times d$ matrix
$\Gamma(k,s)$	gamma distribution with shape k and scale s
χ^2_ν	chi-squared distribution with ν degrees of freedom
$\Phi(\cdot)$	cumulative distribution function of $N_1(0,1)$

This thesis is dedicated to my older brother, Mykola Pupashenko, who guided me all the way from the childhood and played a key role in my development.

Chapter 1

Introduction

1.1 Motivation

1.1.1 Why Market Risk Management?

In this thesis we concentrate on the problem of measuring a special kind of financial risk, called the market risk. Such a risk is associated with changes in the value of a financial position due to unforeseen changes in the value of underlying risk factors. As a result of the market economics development, many financial institutions around the world are exposed to such a risk. These include banks, institutions that deal with numerous risky financial instruments, asset managers and investors. A regular person often takes some market risk by deciding to buy a property or to save for a later day. It is therefore of a great importance to understand and measure risks arising from adverse movements in the prices of the financial instruments. Despite a young age of the risk management field, the literature shows a wide variety of the different methods for quantifying and controlling market risk. We refer an interested reader to Alexander (2008a,b,c,d), Dowd (2002), Hull (2012a), McNeil et al. (2005) for a general overview of the corresponding techniques and tools. There is, however, quite an obvious need in further research on how to measure the market risk correctly and precisely, as many financial crises have demonstrated.

1.1.2 What are the popular Market Risk measures?

During the last few decades there have been intensive discussions on how financial institutions should evaluate risk. Improvements are continuously being sought to oblige financial institutions to cover taken risks by providing the regulatory capital. There is

no chance to eliminate the possibility of a financial institution failing, but the probability of default for a particular institution can be set to be small. By doing this, the governments hope to create a stable economic environment. Hence the regulations have played a major role in the development of risk measurement techniques.

The modern era of the bank regulation began in late 1980s with the 1988 Basel Accord, known now as Basel I. Nowadays there are considerations of implementing an improved modification of this document, known as Basel III. For those, who are interested in the evolution of the bank regulations we suggest to read Hull (2012a) and literature cited there in.

According to the implemented regulations, almost all financial institutions now use some form of Value-at-Risk (VaR) as a measure of risk. VaR measures the worst expected loss over a given time horizon at a given confidence level (see Jorion (2001)). More precisely, losses greater than VaR are suffered only with a specified small probability. We give a more rigorous mathematical definition of VaR later in Section 5.1. The greatest advantage of VaR is that it summarizes the total exposure of an investment with respect to market risk in a single, easy to understand number. No doubt this explains why VaR has become an industry standard in finance for quantifying and controlling the risk of a portfolio of securities.

There are, however, many disadvantages of the use of VaR. Perhaps the most striking problem of VaR, is that it does not satisfy all desired properties of a risk measure. The notion of the coherent risk measure, which satisfies all axioms that a 'good' risk measure should have, is first introduced by Artzner et al. (1999). As it is not the topic of this thesis to compare or investigate the particular risk measures and their properties, we refer an interested reader to Artzner et al. (1999) for the original definitions of the coherent risk measures and to Alexander (2008d), McNeil et al. (2005) for other material regarding the definition of a coherent risk measure and some examples, as well as counterexamples. At this point we merely remark that VaR fails to satisfy the property of sub-additivity (see Artzner et al. (1999)), namely it can not be guaranteed that the VaR of a combined position will not be greater than the sum of the VaR of positions individually considered. This is a great disadvantage, as for some reasons (see Dowd (2002)):

- If risks are sub-additive, then adding risks together would give us an overestimate of combined risk, which can be used as a conservative estimate of a total risk. Otherwise, the sum of risks is useless as a total risk measure;
- If regulators use non-sub-additive risk measures to set capital requirements, a financial institution might tend to break risks up to reduce its capital requirements.

This is because the sum of the capital requirements then can be less than the total capital requirement.

Another drawback of VaR is that it tells us nothing about the size of the losses in the event that VaR is exceeded.

The information about the average level of loss, given that the VaR is exceeded, can be achieved from another market risk measure called Conditional Value-at-Risk (CVaR). Artzner et al. (1999) suggest CVaR as an alternative to VaR which satisfies the axioms of coherent risk measure. This measure has become less popular than VaR because it is harder to understand and estimate. There are a number of variants on the CVaR with a variety of names, such as expected tail loss (ETL), tail conditional expectation (TCE), worst conditional expectation (WCE) and expected shortfall (ES) (see McNeil et al. (2005)). They all coincide for continuous loss distributions, which is the case in this thesis. The precise mathematical definition of CVaR is given in Section 5.1.

1.1.3 How to estimate popular Market Risk measures?

There are basically three main approaches of estimating VaR, which can also be applied for estimation of CVaR. These are variance-covariance or parametric approach, historical simulation and Monte Carlo (MC) simulation. More information about all these methods an interested reader can find in Alexander (2008d), Dowd (2002), Jorion (2001), McNeil et al. (2005). Here, we briefly give an overview of these methods emphasizing on their strong and week points.

We start with variance-covariance approach which is usually based on the assumption of Gaussian distributed changes in risk factors. Under this assumption it is possible to determine the distribution of the portfolio losses. The problem of estimating the VaR then becomes a simple problem of estimating the quantile.

Historical simulation is a simple approach that requires relatively few assumptions about the distribution of the changes in risk factors. It rather uses the historical information about the past changes in risk factors to construct the distribution of a potential future portfolio loss, out of which a simple percentile is taken as a VaR estimation. The distribution of losses is constructed by evaluating the losses under the actual changes in the risk factors during each of the last, say N, periods.

The MC simulation methodology is very similar to the historical simulation. The main difference is that instead of taking observed changes in risk factors over the last N periods, they are generated under some chosen distribution that is believed to adequately

capture the real market. The distribution of losses is constructed by evaluating them under the generated changes in the risk factors. VaR is then determined from this distribution in a similar way.

The variance-covariance method has one clear advantage, that it is quick and easy to compute. However, it has very limited applicability for several reasons. First of all it is suitable only for linear portfolios or instruments with very small holding period, as it uses the linearisation of the portfolio which otherwise can be too unreliable. The second problem is an assumption of Gaussian distributed changes in risk factors. It is well known, that more heavy-tailed distributions seem to fit the behavior of many markets better than the normal distribution (see Hull (2012a)). Moreover, an assumption that all historic data, including possibly complex dependencies between risk factors, are captured by the covariance matrix is very limited, not mentioning the difficulty of their estimation.

Both simulation methods overcome most of the problems of the variance-covariance method. They are extremely effective when any analytical approach is too unreliable. Besides, they are good at dealing with path-dependency, heavy-tails, non-linearity, optionality and multidimensionality. Historical simulation, in particular, is maybe the most popular tool for many financial institutions due to its ease of implementation and reporting. Its great advantage compared to MC method is that it does not require any assumption about the distribution of the changes in risk factors. Historical simulation takes information about underlying distribution from the past data. This is, however, also a big disadvantage, in particular, when the past data is not complete or does not represent the future market movements in a reliable manner. Although, such a data is usually used for estimation of the parameters of assumed distribution for MC methodology, it is very easy to carry out a simulation under alternative assumptions in case of MC method in contrast to historical simulation.

MC simulation is, therefore, maybe the most flexible tool not only for market risk estimation, but also for a wide variety of other applications in finance. It is often a method of a last resort when any analytic solution or approximation does not work, or when other numerical methods fail. Its greatest disadvantage lies in the amount of time necessary to resolve the problem with required precision, especially in case of market risk estimation, when the revaluation of complex pricing models takes usually too much time. Good variance reduction procedures have been, therefore, a focus of recent research in order to speed up MC simulation, keeping the accuracy at the desired level. The current thesis deals with the development of efficient and flexible variance reduction procedures for estimating market risks. At first, such procedures are developed in a more general framework, therefore, they can also have a particular value in other financial applications.

1.2 Outline

As we already mentioned, MC simulation often is the method of choice in many applications of finance. In case of market risk estimation, MC methods involve a high number of usually time-costly portfolio revaluations. There are basically two possible ways to speed up MC simulation, either to reduce the time necessary for portfolio revaluation by applying some approximation or to reduce the number of simulations (and hence revaluations) necessary for a specified precision by reducing the variance of an estimator. Because an approximation of the portfolio losses may not always yield an accurate VaR estimates (see e.g. Glasserman et al. (2002)), we find the second approach more promising. Hence, we are interested in a good variance reduction procedure in order to speed up MC estimation.

Variance reduction procedures have been a focus in the recent research of many authors. Among them we would like to point out the research of Glasserman et al. (2000a,b). The variance reduction techniques are presented there in case of Gaussian distributed changes in risk factors. A substantial variance reduction rate, often of more than two orders of magnitude, is achieved by using an exponential twisting (see Section 2.1.1) combined with stratification (see Section 2.2) based on the Delta-Gamma approximation (see Section 3.1.2) of the loss function. The corresponding method is called Importance Sampling with Stratification (ISS) algorithm (see Sections 4.2.1 and 4.2.2). This approach is quite easy to understand and implement. However, it has very limited applicability for few reasons. First of all, this method is based on the explicitly known distribution of the Delta-Gamma approximation, which is found under the assumption of Gaussian distributed changes in risk factors (see Section 3.3.1). In the case of general distributions it might not be that easy or even not possible. But, even if it is possible to derive the distribution of Delta-Gamma approximation in other cases, the exponential twisting method is restricted to light-tailed densities (see Section 2.1.1). As changes in risk factors on real financial markets typically have heavy tails (see Hull (2012a)), ISS algorithm is generalized by Glasserman et al. (2002) to cover multivariate Student t-distribution. This, however, gives a slightly more difficult modification of ISS algorithm also based on the exact distribution of the Delta-Gamma approximation. Such a distribution is found by Glasserman et al. (2002) and is presented in Section 3.3.2.

Summarizing the above mentioned research of Glasserman et al., a substantial variance reduction rate is achieved in two special cases of Gaussian and Student t-distributed changes in risk factors. As we mentioned before, the main advantage of MC simulation compared to historical simulation lies in the ability to carry out simulation under alternative assumptions of the distribution of the changes in risk factors without an extra

adaptation. However, it seems to be hard to generalize the ISS algorithm for other distributions due to the limited applicability of the exponential twisting and the necessity to derive the distribution of the Delta-Gamma approximation.

The main purpose of this thesis is to derive a unique tool for variance reduction in the context of market risk estimation. It should be easy to implement and understand. For that purpose we present all methods involved in our successful implementation process. It should also have the flexibility of the MC simulation, i.e. without an extra adaptation of the algorithm, the user should be able to use the method for any distribution of changes in risk factors, especially heavy-tailed distributions. Moreover, our algorithm should give at least the same variance reduction rate as already known methods in the cases for which they are developed.

The main idea of using an approximation of the portfolio losses to guide the selection of an effective importance sampling distribution, that samples risk factors so that large losses occur more often, is borrowed from Glasserman et al. In order to achieve desired flexibility, one of the most recent importance sampling methods, namely Cross-Entropy (CE) method (see Section 2.1.2), is used. This method has been successfully applied in many application areas, however, up to our knowledge, not for market risk estimation. A substantial variance reduction rate of the ISS algorithm can mainly be thankful to the stratification part, that is possible only for a known distribution of the Delta-Gamma approximation. Therefore, in Section 4.3.1 we suggest a simple, but very efficient variant of stratification. This special stratification does not involve any information about the distribution of the approximation of the portfolio loss function, hence, gives us an extra flexibility of using any kind of approximation. As the numerical study in Chapter 6 shows, the precision of the Delta-Gamma approximation can be improved, leading to an increase in the variance reduction rate of the corresponding algorithms. For further increase in the performance of the proposed algorithms, in Section 4.4, we improve the choice of the importance sampling density by using an approximation to an optimal importance sampling density (see Section 2.1).

The new approach, developed in this thesis, has the following characteristics.

1. It is a flexible approach:
    - it can be applied for any type of distribution of the risk factors, no matter light- or heavy-tailed,
    - any approximation of the portfolio loss function can be used,
    - special stratification does not involve an extra adaptation of the algorithm for different approximations of the portfolio losses,

- original stratification still can be applied in cases when it is possible.

2. It is highly efficient, as numerical study in Chapter 6 shows:
   - it outperforms the ISS algorithm in all examples and sometimes significantly,
   - alternatives to Delta-Gamma approximation can significantly increase the variance reduction rate,
   - any improvement of the ISS algorithm can also be applied here, hence, generally, it has at least the same performance.
3. It is slightly more complicated than ISS in implementation, but all tools necessary for a successful implementation process are presented here.

The first ideas of using the CE importance sampling method together with a special stratification are introduced in the publication of mine (see Pupashenko (2014)). Further improvements of the algorithm by using an approximation to the optimal importance sampling density, as well as improved special stratification are presented in the joint work with R. Korn (see Korn and Pupashenko (2014)).

The rest of the thesis is organized as follows. Chapter 2 gives a general overview of the variance reduction approaches that became useful in particular applications of market risk estimation, namely importance sampling and stratification methods. In Chapter 3 we consider some approximations of the portfolio loss function which we further use in the numerical study. Similarly to the ISS algorithm our approach originally estimates the probability of high portfolio losses. This is a more general problem than estimating the VaR or CVaR and can also be applied in many other applications. Sections 4.1.1, 4.1.2, 4.2.1 and 4.2.2 give an overview of the methods introduced by Glasserman et al. The rest of the sections in Chapter 4 are devoted to a step by step construction of the new approaches and contain algorithms of all developed methods. The Appendices A and B give an overview of helpful methods for the implementation of the corresponding algorithms. In Chapter 5 we start from the definitions of VaR and CVaR in Section 5.1. Then, in Section 5.2 we discuss how to perform the variance reduction of VaR and CVaR estimation with the help of the algorithms introduced in Chapter 4. The confidence intervals of the corresponding estimates are also discussed in Section 5.3.3. Chapter 6 performs a deep numerical study in light- and heavy-tailed cases in order to compare the performance of the new approaches with those given by Glasserman et al. Pricing formulas, necessary for the corresponding examples are given in Appendix C. Finally, Chapter 7 concludes.

Chapter 2

Variance Reduction of Monte Carlo simulation

As we already mentioned in Chapter 1, the main disadvantage of the MC simulation method is its slow convergence. A high number of simulations is necessary in order to resolve the problem with the required precision. In case of market risk estimation this problem becomes even more crucial because of the, usually time-costly, portfolio revaluations.

The aim of the variance reduction is to produce an alternative estimator to MC with much smaller standard deviation of the error. Then, achieving the desired accuracy in resolving the problem requires smaller number of simulations, and, therefore, less time. One should, however, carefully develop such alternative estimators in the sense, that an additional effort in setting up and performing a variance reduction approach should be less significant than the corresponding reduction in variance. More precisely, one should always question himself, whether it is worth using a variance reduction in a particular application or not.

One of the most effective and widely applicable technique for variance reduction of MC simulation is control variates method. It takes advantage of correlation with an auxiliary variable, for which the quantity of interest is known or can be easily estimated. In the context of the probability of high portfolio losses estimation an approximation of the loss function can play the role of such an auxiliary variable. The corresponding variance reduction based on the Delta-Gamma approximation and control variates is given by Glasserman (2004). We have tried to apply the control variate in the context of our new methods, also in combination with other variance reduction approaches. However, similar to control variates based method from Glasserman (2004), we achieved variance reduction rate significantly smaller then of ISS or our methods. Therefore, we

do not present methods based on control variates here. Besides, this method showed a deterioration of performance with the decrease of the VaR level.

Another method, which is typically used in rare event simulations, namely importance sampling, reduces variance by changing the probability measure of the generated quantity in such a way, that the event of interest occurs more often. The correct estimator is then achieved with the help of the so called likelihood ratio factors.

Finally, stratification reduces the variance by sampling and estimating the required quantity from different parts of some partition of the sample space. The sample can then be reallocated, to reduce the variance in the parts where it is largest.

In this chapter we give a general overview of importance sampling and stratification methods. Their combination showed a great variance reduction in a number of applications and, as numerical study in Chapter 6 shows, also for market risk estimation. Here we bring together the material from the books of Asmussen and Glynn (2007), Chan and Kroese (2012), Chan et al. (2011), Glasserman (2004), Korn et al. (2010) and Rubinstein and Kroese (2004, 2008). Since the corresponding theory is a standard material of most of these books, we will often not explicitly refer to the literature.

Before coming directly to variance reduction procedures, we start here with a formulation of the MC method for estimating the mean of some measurable real-valued function of a random vector. The following applies to the whole thesis:

Assumption 2.1. A parametric absolutely continuous case is considered. More precisely, the random vector of interest has the joint density from some parametric family of densities.

In order to determine a set of consistent notations, during this chapter by $\mathbf{X}$ we denote a random vector from Assumption 2.1 and by $\mathcal{G} = \{f(\mathbf{x}; \mathbf{v}), \mathbf{v} \in \mathcal{V}\}$ its parametric family of densities. The joint density of $\mathbf{X}$ is denoted as $f(\mathbf{x}; \mathbf{u})$ by fixing the original vector parameter $\mathbf{u}$.

We are interested in estimating the following mean:

$$l = \mathbb{E}[H(\mathbf{X})] = \int H(\mathbf{x}) f(\mathbf{x}; \mathbf{u}) d\mathbf{x}, \tag{2.1}$$

for some measurable real-valued function H.

Definition 2.2. By generating a sample $\mathbf{X}_1, \ldots, \mathbf{X}_N$ of independent identically distributed random vectors (often simply called sample) with joint density $f(\mathbf{x}; \mathbf{u})$ and

some natural number N, an unbiased MC estimator of l is defined as follows:

$$\hat{l} = \frac{1}{N} \sum_{i=1}^{N} H(\mathbf{X}_i). \tag{2.2}$$

For further reading regarding MC estimator we refer to Korn et al. (2010).

2.1 Importance Sampling

Importance sampling attempts to reduce the variance by changing the probability measure under which the underlying random vector is generated. The main idea is to change the corresponding distribution in such a way, that the values which contribute the most (and are, therefore, most 'important') to the quantity of interest have higher probability.

For every probability density $g(\mathbf{x})$ such that the following holds true:

$$g(\mathbf{x}) = 0 \Rightarrow H(\mathbf{x}) f(\mathbf{x}; \mathbf{u}) = 0, \tag{2.3}$$

we have the following representation of the mean of interest:

$$\begin{aligned} l &= \mathbb{E}[H(\mathbf{X})] = \int H(\mathbf{x}) f(\mathbf{x}; \mathbf{u}) d\mathbf{x} = \\ &= \int H(\mathbf{x}) \frac{f(\mathbf{x}; \mathbf{u})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x} = \mathbb{E}_g \left[H(\mathbf{X}) \frac{f(\mathbf{X}; \mathbf{u})}{g(\mathbf{X})} \right], \end{aligned}$$

where $\mathbb{E}_g$ denotes the expectation w.r.t. the density $g(\mathbf{x})$.

Definition 2.3. By generating a sample $\mathbf{X}_1, \ldots, \mathbf{X}_N$ of independent identically distributed random vectors with joint density $g(\mathbf{x})$, which is called importance sampling density, and some natural number N, an unbiased importance sampling estimator of l is defined as follows:

$$\hat{l} = \frac{1}{N} \sum_{i=1}^{N} H(\mathbf{X}_i) \frac{f(\mathbf{X}_i; \mathbf{u})}{g(\mathbf{X}_i)}. \tag{2.4}$$

The ration of densities $\frac{f(\mathbf{x};\mathbf{u})}{g(\mathbf{x})}$ is called likelihood ratio.

Clearly, the above procedure leaves us a huge amount of freedom, i.e. we can choose any density function $g(\mathbf{x})$ which satisfies (2.3). Therefore, the next question arises, how to chose the important sampling density so that the importance sampling estimator has smaller variance than the MC estimator? Since both estimators are unbiased, instead of comparing their variances it suffices to compare second moments. The second moment

of the importance sampling estimator reads as follows:

$$\mathbb{E}_g\left[\left(H(\mathbf{X})\frac{f(\mathbf{X};\mathbf{u})}{g(\mathbf{X})}\right)^2\right] = \mathbb{E}\left[H^2(\mathbf{X})\frac{f(\mathbf{X};\mathbf{u})}{g(\mathbf{X})}\right]. \tag{2.5}$$

It is clear, that depending on the choice of $g(\mathbf{x})$, it might be larger or smaller than the second moment $\mathbb{E}\left[H^2(\mathbf{X})\right]$ of the MC estimator. Therefore, successful variance reduction with the help of importance sampling procedure lies in the art of selecting an effective importance sampling density. The following theorem gives the best possible change of measure in a special case under which the variance of an importance sampling estimator is minimized.

Theorem 2.1. *Consider the special case in which the function $H(\mathbf{x})$ is nonnegative. Define the density $g^*(\mathbf{x})$ in the following way*

$$g^*(\mathbf{x}) = \frac{H(\mathbf{x})f(\mathbf{x};\mathbf{u})}{l}. \tag{2.6}$$

Then the importance sampling estimator (2.4) under the choice of the importance sampling density $g(\mathbf{x}) = g^*(\mathbf{x})$ provides a zero-variance (smallest possible variance) estimator of l. The density $g^*(\mathbf{x})$ is, hence, called optimal importance sampling density.

Proof. Note, that in the case of a nonnegative function $H(\mathbf{x})$ the function $g^*(\mathbf{x})$ is indeed density with normalizing constant l according to (2.1).

Since $H(\mathbf{x})\frac{f(\mathbf{x};\mathbf{u})}{g^*(\mathbf{x})} \equiv l$ is a constant regardless of $\mathbf{x}$, the importance sampling estimator (2.4) is also a constant, namely $\hat{l} = l$. Such an estimator obviously has zero-variance. □

In general, the implementation of the optimal importance sampling density $g^*(\mathbf{x})$ as per (2.6) is problematic, because we do not know its normalizing constant l. This is the value we are actually trying to estimate. However, in a special case, there are some methods for sampling from such a density without knowledge of the normalizing constant. These methods are time-consuming and not highly precise, but they help in the derivation of our methods. Such a special case is described in the Remark 2.2. Corresponding sampling methods are called Markov Chain Monte Carlo (MCMC) methods and are briefly presented in Appendix A. For more information on these methods we refer an interested reader to, e.g., Gilks et al. (1995).

Remark 2.2. *If $H(\mathbf{x})$ is an indicator function of some set, then the optimal importance sampling density is the original density conditioned on the set. More general, when $H(\mathbf{x}) = I_A(G(\mathbf{x}))$ for some set A from the range (set of outputs) of some function $G(\mathbf{x})$, then the optimal importance sampling density is precisely the conditional density of $\mathbf{X}$ given $G(\mathbf{X}) \in A$ (assuming that $\mathbb{P}(G(\mathbf{X}) \in A) > 0$).*

An immediate consequence of the Remark 2.2, is that the importance sampling density should approximate the corresponding conditional density. This, in particular, means that, under the importance sampling density, the event $\{G(\mathbf{X}) \in A\}$ should be more likely than under the original density.

We, further, present some easy and popular methods on how to obtain an importance sampling density, that became useful in deriving the variance reduction procedures for market risk estimation discussed in this thesis. We use (and, therefore, present) all these methods only for estimating the special case of an expectation according to the following assumption.

Assumption 2.4. Further in this chapter we are interested in estimating the mean (2.1) in the special case of the function H,

$$H(\mathbf{x}) = I_{(\gamma,+\infty)}(G(\mathbf{x})),$$

for some real level γ and real-valued function G. In this case the problem of estimating the mean (2.1) simplifies to the following problem of estimating the tail probability:

$$l = \mathbb{P}(G(\mathbf{X}) > \gamma). \tag{2.7}$$

2.1.1 Exponential Twisting method

The main idea of the following procedure is to find the importance sampling change of measure among the exponential changes of measure. This explains the name of the method. In order to define this change of measure, we first present some extra material. We indicate Glasserman (2004), Korn et al. (2010) as the most descriptive references on the exponential twisting.

Definition 2.5. For any vector $\boldsymbol{\theta}$ of the same dimension as a random vector $\mathbf{X}$ the function

$$\phi(\boldsymbol{\theta}) = \mathbb{E}[e^{\boldsymbol{\theta}^T\mathbf{X}}] = \int e^{\boldsymbol{\theta}^T\mathbf{x}} f(\mathbf{x};\mathbf{u})d\mathbf{x}$$

is called moment generating function.

Definition 2.6. For any vector $\boldsymbol{\theta}$ of the same dimension as a random vector $\mathbf{X}$ the function

$$\hat{\phi}(\boldsymbol{\theta}) = \mathbb{E}[e^{i\boldsymbol{\theta}^T\mathbf{X}}] = \phi(i\boldsymbol{\theta})$$

is called characteristic function.

Definition 2.7. For any vector $\boldsymbol{\theta}$ of the same dimension as a random vector $\mathbf{X}$ the function

$$\psi(\boldsymbol{\theta}) = \log \mathbb{E}[e^{\boldsymbol{\theta}^T\mathbf{X}}] = \log \phi(\boldsymbol{\theta})$$

is called cumulant generating function.

Let $\Theta = \{\boldsymbol{\theta} : \psi(\boldsymbol{\theta}) < \infty\}$ and suppose that Θ is nonempty. For each $\boldsymbol{\theta} \in \Theta$, set

$$f_{\boldsymbol{\theta}}(\mathbf{x}; \mathbf{u}) = e^{\boldsymbol{\theta}^T \mathbf{x} - \psi(\boldsymbol{\theta})} f(\mathbf{x}; \mathbf{u}). \tag{2.8}$$

Definition 2.8. Each $f_{\boldsymbol{\theta}}(\mathbf{x}; \mathbf{u})$ from (2.8) is a probability density, and the family $\{f_{\boldsymbol{\theta}}(\mathbf{x}; \mathbf{u}), \boldsymbol{\theta} \in \Theta\}$ is called exponential family of densities. The transformation from $f(\mathbf{x}; \mathbf{u})$ to $f_{\boldsymbol{\theta}}(\mathbf{x}; \mathbf{u})$ is called exponential twisting or exponential tilting or simply exponential change of measure.

The cumulant generating function $\psi(\boldsymbol{\theta})$ carries important information about the exponential change of measure $f_{\boldsymbol{\theta}}(\mathbf{x}; \mathbf{u})$. In particular, $\nabla\psi(\boldsymbol{\theta})$ is the mean of $f_{\boldsymbol{\theta}}(\mathbf{x}; \mathbf{u})$. Indeed:

$$\nabla\psi(\boldsymbol{\theta}) = \frac{\nabla\phi(\boldsymbol{\theta})}{\phi(\boldsymbol{\theta})} = \frac{\mathbb{E}[\mathbf{X} e^{\boldsymbol{\theta}^T \mathbf{X}}]}{\mathbb{E}[e^{\boldsymbol{\theta}^T \mathbf{X}}]} = \mathbb{E}[\mathbf{X} e^{\boldsymbol{\theta}^T \mathbf{X} - \psi(\boldsymbol{\theta})}] = \mathbb{E}_{\boldsymbol{\theta}}[\mathbf{X}],$$

where $\mathbb{E}_{\boldsymbol{\theta}}$ is the expectation w.r.t. the density $f_{\boldsymbol{\theta}}(\mathbf{x}; \mathbf{u})$. Therefore, exponential twisting with parameter $\boldsymbol{\theta}$ shifts the mean of $\mathbf{X}$ to the vector $\nabla\psi(\boldsymbol{\theta})$. When such a vector can easily be calculated, an appropriate mean shift can guide our choice of the twisting parameter $\boldsymbol{\theta}$, and therefore importance sampling density, to achieve a successful variance reduction.

In the special case of our interest (under Assumption 2.4) it is possible to derive an appropriate choice of the twisting parameter. The corresponding parameter is chosen such that the variance of the exponential twisting estimator (the importance sampling estimator (2.4) with exponential change of measure (2.8)) is minimized. It is equivalent to the minimization of the second moment of an estimator due to unbiasedness. But fist we have to adapt the notations to the problem (2.7).

Now, for real parameter θ, by $\phi(\theta), \hat{\phi}(\theta)$ and $\psi(\theta)$ we denote the moment generating, characteristic and cumulant generating functions of the random variable $G(\mathbf{X})$ in a similar way to Definitions 2.5, 2.6 and 2.7. Similarly to (2.8), the exponential change of measure in this case is defined as

$$f_{\theta}(\mathbf{x}; \mathbf{u}) = e^{\theta G(\mathbf{x}) - \psi(\theta)} f(\mathbf{x}; \mathbf{u}). \tag{2.9}$$

for each $\theta \in \Theta = \{\theta : \psi(\theta) < \infty\}$.

Instead of minimizing directly the second moment of an exponential twisting estimator in this case, it is easier to minimize its upper bound under the assumption of a positive twisting parameter θ (twisting parameter is under our choice, therefore, we can assume

that it is positive),

$$\mathbb{E}_\theta \left[e^{-2\theta G(\mathbf{X})+2\psi(\theta)} I_{(\gamma,+\infty)}(G(\mathbf{X})) \right] = \mathbb{E} \left[e^{-\theta G(\mathbf{X})+\psi(\theta)} I_{(\gamma,+\infty)}(G(\mathbf{X})) \right] \leq e^{-\theta\gamma+\psi(\theta)}.$$

The minimum of the upper bound $e^{-\theta\gamma+\psi(\theta)}$ is achieved by minimizing $\psi(\theta) - \theta\gamma$ w.r.t. θ. Since the function $\psi(\theta)$ is convex (general property of cumulant generating function), the minimum is a root of the following equation:

$$\psi'(\theta) = \gamma,$$

which corresponds to the shift of the mean to the level γ.

In order to implement the exponential twisting algorithm, we have to sample from the exponential change of measure (2.9). The choice of the corresponding sampling procedure depends on the function G, which is given by the particular application.

2.1.2 Cross - Entropy and Variance Minimization method

In this section we discuss the main ideas behind one of the most recent importance sampling procedures, namely CE algorithm. This method derives its name from the Kullback-Leibler cross-entropy, which is used as a measure of 'distance' between two densities. Such a measure is not a distance in a formal way, because, for example, it is not symmetric, as one can easily see from (2.10). The main advantage of using such a distance, as we will see, is that the corresponding importance sampling algorithm often has simple and explicit solutions for the involved optimization problems. Many other potential measures of distance are being applied in a similar setting by Pupashenko (2011). A comparison of the methods, however, showed obvious advantages of the Kullback-Leibler cross-entropy in the derivation of the CE importance sampling algorithm. Among the literature on CE algorithm we would like to emphasize on Rubinstein and Kroese (2004, 2008) where authors give a descriptive derivation of the method, as well as examples of different applications in which it has been successfully applied.

Definition 2.9. The Kullback-Leibler cross-entropy between two densities $g(\mathbf{x})$ and $h(\mathbf{x})$ is defined as follows:

$$\mathcal{D}(g,h) = \int g(\mathbf{x}) \log \frac{g(\mathbf{x})}{h(\mathbf{x})} d\mathbf{x} = \int g(\mathbf{x}) \log g(\mathbf{x}) d\mathbf{x} - \int g(\mathbf{x}) \log h(\mathbf{x}) d\mathbf{x}. \tag{2.10}$$

By Jensen's inequality for the convex function '$-\log(\cdot)$' the following holds true,

$$\mathcal{D}(g,h) = \mathbb{E}_g \left[-\log \frac{h(\mathbf{Y})}{g(\mathbf{Y})} \right] \geq -\log \mathbb{E}_g \left[\frac{h(\mathbf{Y})}{g(\mathbf{Y})} \right] = -\log 1 = 0,$$

where the random vector $\mathbf{Y}$ is assumed to have density $g(\mathbf{x})$ and $\mathbb{E}_g$ denotes the expectation w.r.t. this density. Therefore, $\mathcal{D}(g,h) \geq 0$, with equality if and only if $g(\mathbf{x}) = h(\mathbf{x})$ for all $\mathbf{x}$. For more details see, e.g. Rubinstein and Kroese (2008).

The main idea of the CE algorithm is to choose the importance sampling density with the minimum Kullback-Leibler cross-entropy to the optimal importance sampling density $g^*(\mathbf{x})$ from (2.6). If the optimization is performed over all possible densities, then the solution is $g^*(\mathbf{x})$ itself, which is useless from practical point of view as described in Section 2.1. The idea of the CE algorithm is to restrict the choice of the importance sampling density by optimizing only over the parametric family $\mathcal{G} = \{f(\mathbf{x};\mathbf{v}), \mathbf{v} \in \mathcal{V}\}$ of the original density $f(\mathbf{x};\mathbf{u})$.

We remind the reader, that we deal with the special case of estimating the tail probability according to Assumption 2.4. In such a case there exists a special modification of the CE algorithm, in literature also called adaptive, multi-level or CE method for rare event simulation (see Rubinstein and Kroese (2004)). In this chapter we review particularly this modification and, for simplicity, call it CE algorithm. The corresponding minimization problem of interest reads in such a case as follows:

$$\min_{\mathbf{v}\in\mathcal{V}} \mathcal{D}(g^*(\cdot), f(\cdot;\mathbf{v})),$$

where under Assumption 2.4 we have:

$$g^*(\mathbf{x}) = \frac{I_{(\gamma,+\infty)}(G(\mathbf{x}))f(\mathbf{x};\mathbf{u})}{l}. \tag{2.11}$$

Definition 2.10. In the setting of CE algorithm, the vector parameter $\mathbf{v}$ is called reference parameter, and the solution to the above stated minimization problem, denoted as $\mathbf{v}^*$, is called an optimal reference parameter.

If we assume that the optimal reference parameter $\mathbf{v}^*$ is known, the corresponding CE estimator is then an importance sampling estimator (2.4) with importance sampling density being $f(\mathbf{x};\mathbf{v}^*)$. In order to define the CE algorithm, it is left to describe how to find an optimal reference parameter.

Using the definition of Kullback-Leibler cross-entropy (2.10) and the definition of the optimal importance sampling density $g^*(\mathbf{x})$ from (2.11), the above minimization problem:

$$\min_{\mathbf{v}\in\mathcal{V}} \mathcal{D}(g^*(\cdot), f(\cdot;\mathbf{v})) = \min_{\mathbf{v}\in\mathcal{V}} \left(\int g^*(\mathbf{x}) \log g^*(\mathbf{x}) d\mathbf{x} - \int g^*(\mathbf{x}) \log f(\mathbf{x};\mathbf{v}) d\mathbf{x} \right),$$

can be easily simplified to the following maximization problem:

$$\begin{aligned} & \max_{\mathbf{v}\in\mathcal{V}} \int I_{(\gamma,+\infty)}(G(\mathbf{x}))f(\mathbf{x};\mathbf{u})\log f(\mathbf{x};\mathbf{v})d\mathbf{x} = \\ = \quad & \max_{\mathbf{v}\in\mathcal{V}} \mathbb{E}_{\mathbf{u}}\left[I_{(\gamma,+\infty)}(G(\mathbf{X}))\log f(\mathbf{X};\mathbf{v})\right]. \end{aligned}$$

In order to define an adaptive updating of the reference parameter it is useful to even further modify the optimization problem in the following way:

$$\max_{\mathbf{v}\in\mathcal{V}} \mathbb{E}_{\mathbf{w}}\left[I_{(\gamma,+\infty)}(G(\mathbf{X}))W(\mathbf{X};\mathbf{u},\mathbf{w})\log f(\mathbf{X};\mathbf{v})\right],$$

where

$$W(\mathbf{x};\mathbf{u},\mathbf{w}) = \frac{f(\mathbf{x};\mathbf{u})}{f(\mathbf{x};\mathbf{w})}$$

is a likelihood ration of densities $f(\mathbf{x};\mathbf{u})$ and $f(\mathbf{x};\mathbf{w})$.

In most situations, the above optimization problem is hard to solve analytically as it might be hard or impossible to find the corresponding expectation. However, in order to estimate the optimal reference parameter one can use the following optimization stochastic counterpart for some natural number M:

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{w})\log f(\mathbf{X_i};\mathbf{v}), \tag{2.12}$$

where $\mathbf{X}_1,\ldots,\mathbf{X}_M$ is a sample from density $f(\mathbf{x};\mathbf{w})$. According to Rubinstein and Kroese (2004), in typical applications the function under maximization (w.r.t. the argument $\mathbf{v}$) in (2.12) is convex and differentiable w.r.t. $\mathbf{v}$, in which case the solution to (2.12) can be readily obtained by solving (w.r.t. $\mathbf{v}$) the following system of equations:

$$\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{w})\nabla\log f(\mathbf{X}_i;\mathbf{v}) = 0. \tag{2.13}$$

The advantage of such an approach is that the solution to (2.13) can be often obtained analytically. In particular, this happens for the so called exponential family of densities (see Rubinstein and Kroese (2004, 2008)). It is not important to understand the definition of this family, we rather mention here, that exponential families include many common distributions, such as the normal, exponential, gamma, chi-squared, beta, Poisson and many others.

An estimate, achieved from (2.12) or (2.13) gives us a procedure of a reference parameter update in order to estimate the optimal reference parameter. More precisely, we start from $\mathbf{w} = \mathbf{u}$, then estimate $\hat{\mathbf{v}}$ via (2.12) or (2.13) and plug it in the same estimators again

as $\mathbf{w}$. This procedure continues until some stopping criterion is reached (for details see Rubinstein and Kroese (2004, 2008)).

Remark 2.3. *It is important to understand that the quality of the optimal reference parameter estimator via (2.12) or (2.13) depends on the sample size M, which from the time efficiency point of view should be kept as small as possible. Therefore, if the probability of an event $\{G(\mathbf{X}) > \gamma\}$ is small under the density $f(\mathbf{x}; \mathbf{w})$ (this is usually the case especially for the starting parameter $\mathbf{w} = \mathbf{u}$), most of summands in (2.12) or (2.13) are zeros, leading to a high variance of the reference parameter estimate. In order to keep sample size M small, an adaptive updating of the level in parallel with updating the reference parameter is introduced by Rubinstein and Kroese (2004, 2008).*

The CE algorithm is initialized by the choice of a not very small rarity parameter ρ (typically between 0.01 and 0.1) and by setting $\mathbf{w} = \mathbf{u}$. Next, an intermediate level smaller than the original γ is defined, such that under the density $f(\mathbf{x}; \mathbf{w}) = f(\mathbf{x}; \mathbf{u})$, the probability of $G(\mathbf{X})$ greater then this level is at least ρ. Under such an intermediate level a reference parameter is estimated. The estimated reference parameter is then taken as $\mathbf{w}$ and the procedure is repeated until the intermediate level reaches the original level γ. In other words, in each iteration step, the CE algorithm consists first of updating the level and then updating the reference parameter. For more details on both phases we refer the reader to Rubinstein and Kroese (2004, 2008). We further summarize all said above in the following CE algorithm:

Algorithm 2.1 (CE)**.**

1. *Define $\hat{\mathbf{v}}_0 = \mathbf{u}$ and set $j = 1$. Fix input parameters ρ, M and N.*

2. *Generate a sample $\mathbf{X}_1, ..., \mathbf{X}_M$ from the density $f(\mathbf{x}; \hat{\mathbf{v}}_{j-1})$ and compute the sample $(1-\rho)-$ quantile $\hat{\gamma}_j$ of a sample $G(\mathbf{X}_1), ..., G(\mathbf{X}_M)$. If $\hat{\gamma}_j > \gamma$, set $\hat{\gamma}_j = \gamma$.*

3. *Using the same sample $\mathbf{X}_1, ..., \mathbf{X}_M$, find an estimate from the adaptive optimization problem in the form of (2.12) or (2.13),*

$$\max_{\mathbf{v} \in \mathcal{V}} \frac{1}{M} \sum_{i=1}^{M} I_{(\hat{\gamma}_j, \infty)}(G(\mathbf{X}_i)) W(\mathbf{X}_i; \mathbf{u}, \hat{\mathbf{v}}_{j-1}) \log f(\mathbf{X}_i; \mathbf{v}), \tag{2.14}$$

   and denote the solution as $\hat{\mathbf{v}}_j$.

4. *If $\hat{\gamma}_j < \gamma$, set $j = j + 1$ and reiterate from step 2, else set an optimal reference parameter estimate $\hat{\mathbf{v}}^* = \hat{\mathbf{v}}_j$ and proceed with step 5.*

5. *Estimate the probability l using the importance sampling estimator (2.4) in the form*

$$\hat{l} = \frac{1}{N} \sum_{i=1}^{N} I_{(\gamma,\infty)}(G(\mathbf{X}_i)) W(\mathbf{X}_i; \mathbf{u}, \hat{\mathbf{v}}^*), \tag{2.15}$$

where $\mathbf{X}_1, ..., \mathbf{X}_N$ is a sample generated from the density $f(\mathbf{x}; \hat{\mathbf{v}}^*)$.

Remark 2.4. *As we mentioned at the beginning of this chapter, the additional effort for setting up and performing a variance reduction approach (in case of Algorithm 2.1 these are steps 1-4) should be kept as small as possible. The corresponding efficiency of the CE algorithm depends, therefore, on the choice of the rarity parameter ρ and the sample size M for the reference parameter update. The rarity parameter influences the number of iterations of steps 2-4. If ρ is chosen to be large, then the intermediate level $\hat{\gamma}_j$ converges slowly towards the true level γ, leading to a higher number of iterations. If it is chosen to be small, the optimization problem (2.14) has more zero summands, forcing us to increase the sample size M. Therefore, ρ is usually chosen between* 0.01 *and* 0.1*. Moreover, if the probability under estimation l is not very small (larger than* 0.001*), then the value of* 0.1 *is preferred.*

The sample size N is responsible directly for the accuracy of the estimator $\hat{l}$, whereas the sample size M is responsible for the accuracy of the optimal reference parameter estimator. Therefore, usually a much smaller sample size M should be chosen than the final sample size N. Our choices on the corresponding parameters in particular examples are discussed in Chapter 6.

As we already mentioned before, the main advantage of CE algorithm is that the optimization problem (2.14) can be often solved analytically. In some applications, where this is not the case, the so called Variance Minimization algorithm (VM) (see e.g. Rubinstein and Kroese (2004, 2008)) can become more handy. VM algorithm is very similar to CE algorithm. It only differs in the step of the reference parameter update. The optimization problem for the CE reference parameter update is based on the minimization of the Kullback-Leibler cross-entropy to the optimal importance sampling density. VM algorithm minimizes the variance of the importance sampling estimator directly. More precisely, according to (2.5) the parametric variance minimization problem reads as follows:

$$\min_{\mathbf{v} \in \mathcal{V}} \mathbb{E}_{\mathbf{u}} \left[I_{(\gamma,+\infty)}(G(\mathbf{X})) W(\mathbf{X}; \mathbf{u}, \mathbf{v}) \right].$$

Similarly to CE algorithm, in order to define an adaptive updating of the reference parameter the following modification is helpful:

$$\min_{\mathbf{v} \in \mathcal{V}} \mathbb{E}_{\mathbf{w}} \left[I_{(\gamma,+\infty)}(G(\mathbf{X})) W(\mathbf{X}; \mathbf{u}, \mathbf{v}) W(\mathbf{X}; \mathbf{u}, \mathbf{w}) \right].$$

For some natural number M, the corresponding optimization stochastic counterpart can be written as:

$$\min_{\mathbf{v}\in\mathcal{V}} \frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{v})W(\mathbf{X}_i;\mathbf{u},\mathbf{w}), \tag{2.16}$$

where $\mathbf{X}_1,\ldots,\mathbf{X}_M$ is a sample from density $f(\mathbf{x};\mathbf{w})$. The analogue to (2.13) is, therefore, of the form:

$$\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))\nabla W(\mathbf{X}_i;\mathbf{u},\mathbf{v})W(\mathbf{X}_i;\mathbf{u},\mathbf{w}) = 0. \tag{2.17}$$

This all summarizes in VM algorithm, which can be achieved from Algorithm 2.1 by substituting the corresponding reference parameter update rules (2.12) and (2.13) with (2.16) and (2.17) respectively. In more details,

Algorithm 2.2 (VM)**.**

1. *Define* $\hat{\mathbf{v}}_0 = \mathbf{u}$ *and set* $j = 1$*. Fix input parameters* ρ*,* M *and* N*.*

2. *Generate a sample* $\mathbf{X}_1,...,\mathbf{X}_M$ *from the density* $f(\mathbf{x};\hat{\mathbf{v}}_{j-1})$ *and compute the sample* $(1-\rho)-$ *quantile* $\hat{\gamma}_j$ *of a sample* $G(\mathbf{X}_1),...,G(\mathbf{X}_M)$*. If* $\hat{\gamma}_j > \gamma$*, set* $\hat{\gamma}_j = \gamma$*.*

3. *Using the same sample* $\mathbf{X}_1,...,\mathbf{X}_M$*, find an estimate from the adaptive optimization problem in the form of (2.16) or (2.17),*

$$\min_{\mathbf{v}\in\mathcal{V}} \frac{1}{M}\sum_{i=1}^{M} I_{(\hat{\gamma}_j,\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{v})W(\mathbf{X}_i;\mathbf{u},\hat{\mathbf{v}}_{j-1}), \tag{2.18}$$

   and denote the solution as $\hat{\mathbf{v}}_j$*.*

4. *If* $\hat{\gamma}_j < \gamma$*, set* $j = j+1$ *and reiterate from step 2, else set an optimal reference parameter estimate* $\hat{\mathbf{v}}^* = \hat{\mathbf{v}}_j$ *and proceed with step 5.*

5. *Estimate the probability* l *using the importance sampling estimator (2.4) in the form*

$$\hat{l} = \frac{1}{N}\sum_{i=1}^{N} I_{(\gamma,\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\hat{\mathbf{v}}^*), \tag{2.19}$$

   where $\mathbf{X}_1,...,\mathbf{X}_N$ *is a sample generated from the density* $f(\mathbf{x};\hat{\mathbf{v}}^*)$*.*

Remark 2.4 applies also to Algoritm 2.2. For a comparison of the CE and VM methods we refer to Chan et al. (2011), where authors conclude that both methods prescribe the same importance sampling parameters, at least asymptotically.

2.1.3 Improved Cross - Entropy method

CE algorithm is a flexible adaptive importance sampling procedure that has been successfully applied to a wide variety of applications. However, Chan and Kroese (2012) mentioned, that in some high dimensional settings, CE algorithm may give unreliable results. This is due to the so called degeneracy of the likelihood ratio estimator problem, which appears in the reference parameter update step (2.12), (2.13), (2.16) or (2.17). This problem is a tendency of the likelihood ratio to attain a very skewed distribution with a concentration in values close to zero, but also attaining very large values with a small, but non-zero probability (see Rubinstein and Kroese (2008)). It can be solved by increasing the sample size M necessary for the reference parameter update step, however, leading to a substantial increase of the computational time, especially in high dimensions. A first attempt on how to deal with the curse of dimensionality of the likelihood ratios in the CE algorithm is made by Rubinstein and Glynn (2009). We, however, are more interested in simpler and more efficient variations of the CE algorithm, whose performance does not suffer from the increase in the dimension of the problem, namely the Improved Cross-Entropy (ICE) algorithm, that is given by Chan and Kroese (2012). In this section we briefly present this method.

The ICE method is motivated by the investigation, that an estimator $\hat{\mathbf{v}}^*$ obtained from CE algorithm is not a good estimator of $\mathbf{v}^*$ in some high dimensional settings. Therefore, instead of solving the reference parameter update step (2.12), (2.13),(2.16) or (2.17) sequentially to obtain $\hat{\mathbf{v}}^*$, an alternative estimator is considered. The proposed estimator does not involve any likelihood ratio (hence can not suffer from the degeneracy of the likelihood ratio estimator problem), and can be obtained in one step.

Recall that the optimal reference parameter $\mathbf{v}^*$ is originally a solution to the following problem,

$$\max_{\mathbf{v}\in\mathcal{V}} \mathbb{E}_{\mathbf{u}} \left[I_{(\gamma,+\infty)}(G(\mathbf{X})) \log f(\mathbf{X};\mathbf{v}) \right].$$

The corresponding estimator $\hat{\mathbf{v}}^*$ of the optimal reference parameter $\mathbf{v}^*$ can be directly achieved from the corresponding sample counterpart:

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M} \sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i)) \log f(\mathbf{X}_i;\mathbf{v}),$$

for some natural number M and sample $\mathbf{X}_1,\ldots,\mathbf{X}_M$ from the density $f(\mathbf{x};\mathbf{u})$. The reason why this is not practical and why an adaptive procedure in the CE algorithm is proposed, is that this estimator has high variance due to the fact that most of the indicators in the sum are zeros. However, one can overcome this difficulty by drawing a sample from an optimal importance sampling density $g^*(\mathbf{x})$. Assuming that this is

possible, the corresponding maximization problem of the estimator takes the form:

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))\frac{f(\mathbf{X}_i;\mathbf{u})}{g^*(\mathbf{X}_i)}\log f(\mathbf{X}_i;\mathbf{v}),$$

and can be further simplified as follows:

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M}\sum_{i=1}^{M} \log f(\mathbf{X}_i;\mathbf{v}),$$

where $\mathbf{X}_1,\ldots,\mathbf{X}_M$ is a sample from the density $g^*(\mathbf{x})$. The main advantage of this optimization problem is that it does not involve any indicator function or likelihood ratio. Hence, beside a substantial computational saving in high-dimensional settings, it gives a more robust and numerically stable solution. Generating draws from optimal importance sampling density $g^*(\mathbf{x})$ (conditional density) requires additional effort in setting the algorithm, as well as extra computational time, but with the advent of MCMC methods this problem is well studied. For a review of useful MCMC methods for sampling from the conditional distribution we refer the reader to Appendix A.

All said above summarizes to the following ICE algorithm:

Algorithm 2.3 (ICE)**.**

1. *Fix input parameters M and N.*
2. *Generate a sample $\mathbf{X}_1,...,\mathbf{X}_M$ from the density $g^*(\mathbf{x})$ given by (2.11) by means of some chosen MCMC method.*
3. *Using this sample, find an estimate from the optimization problem,*

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M}\sum_{i=1}^{M} \log f(\mathbf{X}_i;\mathbf{v}), \tag{2.20}$$

   and denote the solution as $\hat{\mathbf{v}}^*$.
4. *Estimate the probability l using the importance sampling estimator (2.4) in the form*

$$\hat{l} = \frac{1}{N}\sum_{i=1}^{N} I_{(\gamma,\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\hat{\mathbf{v}}^*), \tag{2.21}$$

   where $\mathbf{X}_1,...,\mathbf{X}_N$ is a sample generated from the density $f(\mathbf{x};\hat{\mathbf{v}}^*)$.

Authors in Chan and Kroese (2012) suggest the Gibbs sampler (see e.g. Gilks et al. (1995)) as MCMC method for sampling from the conditional density. This method needs to be adjusted for different models of the changes in risk factors. For the case of

Student t-distributed changes in risk factors it is done in Chan and Kroese (2012). In order to achieve the desired flexibility for our methods, we have slightly modified step 2 in Algorithm 2.3. Some more flexible MCMC methods are proposed in Appendix A.

2.2 Stratification

The main principle underlying stratification (in literature also called stratified sampling) is to subdivide the set of possible outcomes of $H(\mathbf{X})$ in smaller subsets that mirror the properties of $H(\mathbf{X})$ as much as possible (see Glasserman (2004), Korn et al. (2010)). The indicator of such properties does not have to be the same random variable $H(\mathbf{X})$. We denote it as some other random vector $\mathbf{Y}$, which has to satisfy the Assumption 2.11.

Assumption 2.11. The random vector $\mathbf{Y}$ satisfies:

- the probability distribution of $\mathbf{Y}$ is known and easy to calculate,
- it is easy to sample $H(\mathbf{X})$ conditional on $\mathbf{Y}$.

Definition 2.12. Such a random vector $\mathbf{Y}$, that satisfies Assumption 2.11, is called stratification variable.

In order to formulate the stratification procedure one has to define for some natural number K the strata $A_1, \ldots, A_K$ to be disjoint subsets of all possible outcomes of the random vector $\mathbf{Y}$, such that $\mathbb{P}(\mathbf{Y} \in \bigcup_{i=1}^{K} A_i) = 1$. Then the mean under estimation (2.1) can be written as follows:

$$l = \mathbb{E}[H(\mathbf{X})] = \sum_{i=1}^{K} \mathbb{P}(\mathbf{Y} \in A_i)\mathbb{E}[H(\mathbf{X})|\mathbf{Y} \in A_i] = \sum_{i=1}^{K} p_i \mathbb{E}[H(\mathbf{X})|\mathbf{Y} \in A_i], \tag{2.22}$$

where $p_i = \mathbb{P}(\mathbf{Y} \in A_i)$ are known by Assumption 2.11.

In order to define the stratified estimator via (2.22) it is left to estimate K different conditional expectations $\mathbb{E}[H(\mathbf{X})|\mathbf{Y} \in A_i]$ by the MC method with sample sizes $N_1, \ldots, N_K$, correspondingly. More precisely the estimator reads as follows:

$$\hat{l} = \sum_{i=1}^{K} p_i \frac{1}{N_i} \sum_{j=1}^{N_i} H(\mathbf{X}_{ij}), \tag{2.23}$$

where $(\mathbf{Y}_{ij}, \mathbf{X}_{ij}), j = 1, \ldots, N_i$ is a sample generated from the conditional distribution of $(\mathbf{Y}, \mathbf{X})$ given $\mathbf{Y} \in A_i$ for $i = 1, \ldots, K$. For the ease of comparison with the MC estimator (2.2) we assume that $N = N_1 + \ldots + N_K$.

Definition 2.13. The corresponding set of sample sizes $N_1, \ldots, N_K$ for each stratum $A_1, \ldots, A_K$, such that $N = N_1 + \ldots + N_K$, is called allocation.

It should be clear now, that the use of the stratification (2.23) involves the following issues:

- choosing the stratification variable $\mathbf{Y}$, the number of strata K, strata $A_1, \ldots, A_K$, and the allocation $N_1, \ldots, N_K$,
- sampling from the conditional distribution of $(\mathbf{Y}, \mathbf{X})$ given $\mathbf{Y} \in A_i$ for $i = 1, \ldots, K$.

The choice of the stratification variable, the number of strata and strata itself depends highly on the particular application and is later discussed in this thesis in certain algorithms. Regarding sampling from a conditional distribution we refer the reader to Appendix A. Further in this section we give some examples of an allocation, such that a usually substantial variance reduction rate is achieved.

First, we need to introduce some more notations across this section. For $\mathbf{X}_{ij}$ having the distribution of $\mathbf{X}$ given $\mathbf{Y} \in A_i$ for $i = 1, \ldots, K$, let

$$\mu_i = \mathbb{E}[H(\mathbf{X}_{ij})] = \mathbb{E}[H(\mathbf{X})|\mathbf{Y} \in A_i],$$

$$\sigma_i^2 = \mathbb{V}\mathbf{ar}[H(\mathbf{X}_{ij})] = \mathbb{V}\mathbf{ar}[H(\mathbf{X})|\mathbf{Y} \in A_i].$$

We first show, that the estimator (2.23) is unbiased, because

$$\mathbb{E}[\hat{l}] = \sum_{i=1}^{K} p_i \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbb{E}[H(\mathbf{X}_{ij})] = \sum_{i=1}^{K} p_i \mu_i = \mathbb{E}[H(\mathbf{X})] = l,$$

where we have used the Bayes' formula for expectations for a collection of mutually exclusive and exhaustive events $A_1, \ldots, A_K$.

Due to the conditional independence of subestimators, the variance of the stratified estimator (2.23) is

$$\mathbb{V}\mathbf{ar}[\hat{l}] = \sum_{i=1}^{K} p_i^2 \frac{1}{N_i^2} \sum_{j=1}^{N_i} \mathbb{V}\mathbf{ar}[H(\mathbf{X}_{ij})] = \sum_{i=1}^{K} p_i^2 \frac{\sigma_i^2}{N_i} = \sum_{i=1}^{K} \frac{p_i}{N_i} p_i \sigma_i^2.$$

The variance of the MC estimator (2.2) with the same total sample size $N = N_1 + \ldots + N_K$ equals $\mathbb{V}\mathbf{ar}[H(\mathbf{X})]/N$ and, due to the Bayes' formula for variances of the collection of mutually exclusive and exhaustive events $A_1, \ldots, A_K$, satisfies the following:

$$\frac{\mathbb{V}\mathbf{ar}[H(\mathbf{X})]}{N} = \frac{\mathbb{E}[\mathbb{V}\mathbf{ar}[H(\mathbf{X})|\mathbf{Y}]] + \mathbb{V}\mathbf{ar}[\mathbb{E}[H(\mathbf{X})|\mathbf{Y}]]}{N} \geq \frac{\mathbb{E}[\mathbb{V}\mathbf{ar}[H(\mathbf{X})|\mathbf{Y}]]}{N} = \sum_{i=1}^{K} \frac{1}{N} p_i \sigma_i^2,$$

with a strict inequality if $\mathbb{E}[\mathbb{V}\mathbf{ar}[H(\mathbf{X})|\mathbf{Y}]]$ is not almost surely constant. In such a case, clearly an allocation of the form $N_i = Np_i$, under the assumption that all N_i are natural numbers, corresponds to a stratification with variance strictly smaller than the one of the MC method. Such an allocation is called proportional allocation. For more details on the stratification we again refer to Glasserman (2004), Korn et al. (2010).

Chapter 3

Approximations of the Portfolio Loss function

As mentioned in Section 1.1, the MC algorithm for market risk estimation is often a method of choice for complex non-linear portfolios, that usually contain a lot of different options. These portfolios are, therefore, of particular interest for the algorithms proposed in Chapter 4. The main problem of the MC method in such a case is that for high precision it involves a great number of time-costly portfolio revaluations. Therefore, the variance reduction approaches presented in Chapter 2 are mandatory. Although they may be applied directly to the problem, the additional effort in setting up an estimator may be huge because of the same necessity of a large number of time-costly portfolio revaluations. Therefore, as we already mentioned in Section 1.2, instead of using the variance reduction methods directly, an approximation of the portfolio loss function is used in order to guide the selection of an importance sampling density. In this chapter we further present some approximations of the portfolio losses, which became popular, especially for option portfolios.

We start by introducing some notations, which apply also to the following chapters. First, we fix the starting time t. By $\mathbf{S} = \mathbf{S}(t)$ we denote a d-dimensional vector of market prices and rates (risk factors) at time t. We keep the simple notation $\mathbf{S}$, omitting t, for starting prices, because the starting time t is fixed. Since we are interested in portfolio losses over some time, by Δt we denote the risk-measurement horizon and by $\Delta\mathbf{S}$ the corresponding change in market prices over this horizon, i.e. from time t to time $t + \Delta t$. By $V(\mathbf{S}, t)$ we denote the portfolio value at time t and market price $\mathbf{S}$ at the same time t. Since we are interested in the losses over the time horizon Δt only, the intermediate values of the portfolio are not important, and the final value of the portfolio at time $t + \Delta t$ is hence denoted by $V(\mathbf{S} + \Delta\mathbf{S}, t + \Delta t)$. The corresponding portfolio losses over

the time horizon Δt are written via changes in portfolio value ΔV (simplified notation of $\Delta V(\Delta \mathbf{S}, \Delta t)$) and denoted by

$$L(\Delta \mathbf{S}, \Delta t) = -\Delta V(\Delta \mathbf{S}, \Delta t) = V(\mathbf{S}, t) - V(\mathbf{S} + \Delta \mathbf{S}, t + \Delta t).$$

In general, the portfolio loss is a function of two parameters: the vector $\Delta \mathbf{S}$ and the time horizon Δt. However, in the case of market risk estimation, the time horizon is usually fixed, hence the simplified notation $L(\Delta \mathbf{S})$ is often used when the dependency on the time horizon does not have to be shown. We should also note that $\Delta \mathbf{S}$ is a random vector and, therefore, $L(\Delta \mathbf{S})$ is random.

A very simple idea is used behind all approximations mentioned in this chapter. A common Taylor series expansion is a good approximation to any function, so also for the changes in portfolio value ΔV. The portfolio loss function simply represents the negative changes in the portfolio value ΔV. All approximations given here are Taylor series expansions until the specific order. The Taylor series expansion for the portfolio value $V(\mathbf{S} + \Delta \mathbf{S}, t + \Delta t)$ around the starting vector $(\mathbf{S}, t)$ can be written in the following form:

$$\sum_{m_0=0}^{\infty} \sum_{m_1=0}^{\infty} \dots \sum_{m_d=0}^{\infty} \frac{\Delta t^{m_0} \Delta \mathbf{S}_1^{m_1} \dots \Delta \mathbf{S}_d^{m_d}}{m_0! m_1! \dots m_d!} \frac{\partial^{m_0+m_1+\dots+m_d} V(\mathbf{S}, t)}{\partial t^{m_0} \partial \mathbf{S}_1^{m_1} \dots \partial \mathbf{S}_d^{m_d}}. \tag{3.1}$$

One should note, that the term with all $m_i = 0$ is simply $V(\mathbf{S}, t)$. Therefore, it is easy to get a series expression for ΔV or $L(\Delta \mathbf{S}, \Delta t)$ out of (3.1).

In the following sections we present some approximations commonly used in applications of market risk estimation. The Greek letters in their names denote the fact that these approximations are based on the derivatives of the portfolio values. These derivatives are usually termed by Greek letters and called, therefore, Greeks. There is a wide variety of literature on Greeks and their explicit formulas for some options. Because, with the help of mathematical programming languages, it is easy to implement such derivatives, we do not present or use these formulas in this thesis. For more information about these approximations and Greeks we refer an interested reader to Glasserman (2004), Hull (2012a,b).

3.1 Classical approximations

3.1.1 Delta approximation

This approximation is based on the first derivative w.r.t. time and first derivatives w.r.t. the risk factors, more precisely

$$\Delta V \approx \frac{\partial V(\mathbf{S},t)}{\partial t}\Delta t + \sum_{i=1}^{d} \frac{\partial V(\mathbf{S},t)}{\partial \mathbf{S}_i}\Delta \mathbf{S}_i. \tag{3.2}$$

First derivative of the portfolio w.r.t. changes in risk factors is called Delta. Second summand in (3.2) is a sum of Deltas w.r.t. each underlying risk factor. This explains the name Delta for this approximation.

Although this approximation is simple and easy to compute, there is, however, an obvious drawback, especially for option portfolios, which are of a particular interest in this thesis. Delta approximation is a linear function of the changes in risk factors, therefore it is not a good approximation for the losses of portfolios containing options that are highly nonlinear. Besides, the numerical study in Glasserman et al. (2000a,b) showed inefficiency of using such an approximation. One has, therefore, to include some higher derivatives w.r.t. underlying risk factors. This is the matter of the next section.

3.1.2 Delta - Gamma approximation

In this section we give maybe the most popular approximation used for portfolios with options. The following approximation is called Delta-Gamma approximation, because it extends the Delta approximation (3.2) by adding second order derivatives w.r.t. changes in risk factors. Such derivatives are called Gamma. The corresponding approximation reads as follows:

$$\Delta V \approx \frac{\partial V(\mathbf{S},t)}{\partial t}\Delta t + \sum_{i=1}^{d} \frac{\partial V(\mathbf{S},t)}{\partial \mathbf{S}_i}\Delta \mathbf{S}_i + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 V(\mathbf{S},t)}{\partial \mathbf{S}_i \partial \mathbf{S}_j}\Delta \mathbf{S}_i \Delta \mathbf{S}_j. \tag{3.3}$$

This approximation has a particular value in this thesis. It is a key approximation for variance reduction methods introduced by Glasserman (2004), Glasserman et al. (2000a,b, 2002). We review these methods in Sections 4.1.1, 4.1.2, 4.2.1 and 4.2.2. In contrast to the methods introduced by us, the corresponding methods from Glasserman (2004), Glasserman et al. (2000a,b, 2002) involve some additional investigation of the

Delta-Gamma approximation in particular cases. More precisely, under some assumptions on the distribution of the underlying risk factors, the distribution of the Delta-Gamma approximation is derived explicitly. We present the corresponding results in Section 3.3 (see e.g. Glasserman (2004)).

Since the algorithms developed in this thesis are flexible in the sense that any approximation of the portfolio loss function can be used, we further give some more approximations, which we later use in the tests in Chapter 6. In some examples, a significant improvement compared to a classical Delta-Gamma approximation is achieved.

3.1.3 Charm - Speed approximation

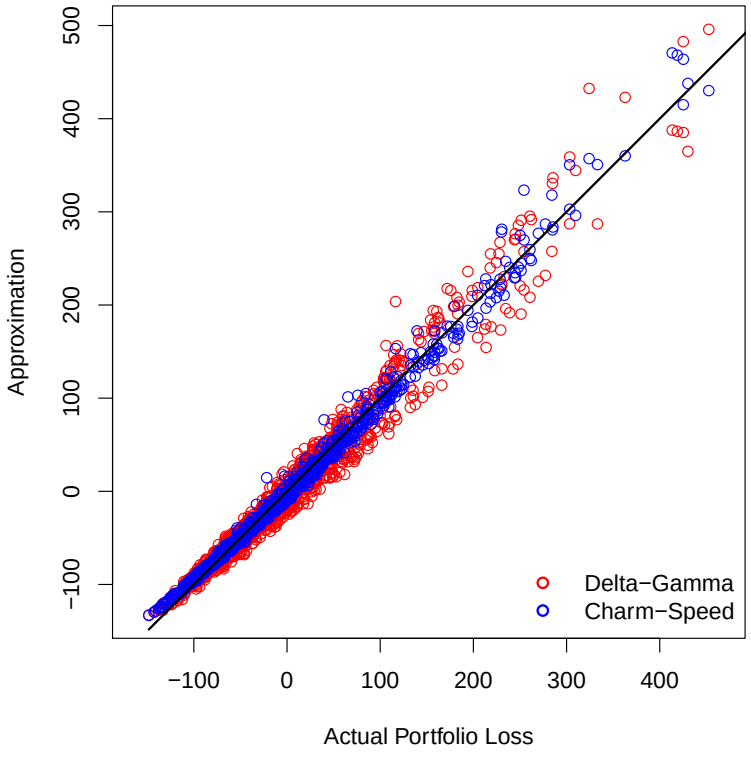


FIGURE 3.1: Comparison of Delta-Gamma and Charm-Speed approximations against actual portfolio losses in 1000 randomly generated scenarios for a portfolio of short 10 ATM calls, 10 ATM puts, 10 assets with correlation 0.3, 0.1 years maturity, Gaussian changes in risk factors (for details on the portfolio see Chapter 6).

A simple way to improve the Delta-Gamma approximation is to add some higher oder derivatives of the Taylor series expansion. Ederington and Guan (2007) investigated, whether any third order derivatives are important in explaining changes in prices of complex nonlinear portfolios. The authors conclude that, while Gamma is usually the

most important, several other higher order derivatives can considerably improve the approximation. As one can conclude from the numerical study in Chapter 6, the Charm-Speed approximation introduced in the following can lead to a significant improvement compared to the Delta-Gamma approximation, without significant extra costs. A comparison of these two approximations for a portfolio of short 10 ATM calls and 10 ATM puts with 0.1 years maturity on 10 assets with correlation 0.3 (for details on portfolio see Chapter 6) is plotted against actual portfolio losses in the Figure 3.1.

Our investigations on a wide variety of portfolios showed that the best improvement of the Delta-Gamma approximation is achieved by adding second cross derivatives w.r.t. time and third derivatives w.r.t. the changes in risk factors. Such derivatives are called Charm and Speed correspondingly, explaining the name of the approximation. The corresponding approximation reads as follows:

$$\begin{aligned} \Delta V &\approx \frac{\partial V(\mathbf{S},t)}{\partial t}\Delta t + \sum_{i=1}^{d}\frac{\partial V(\mathbf{S},t)}{\partial \mathbf{S}_i}\Delta \mathbf{S}_i + \sum_{i=1}^{d}\frac{\partial^2 V(\mathbf{S},t)}{\partial \mathbf{S}_i \partial t}\Delta \mathbf{S}_i \Delta t + \\ &+ \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 V(\mathbf{S},t)}{\partial \mathbf{S}_i \partial \mathbf{S}_j}\Delta \mathbf{S}_i \Delta \mathbf{S}_j + \frac{1}{6}\sum_{i,j,k=1}^{d}\frac{\partial^3 V(\mathbf{S},t)}{\partial \mathbf{S}_i \partial \mathbf{S}_j \partial \mathbf{S}_k}\Delta \mathbf{S}_i \Delta \mathbf{S}_j \Delta \mathbf{S}_k. \end{aligned} \tag{3.4}$$

3.2 Modified approximations

In the next two sections we are going to present the improved Delta-Gamma and Charm-Speed approximations based on the modification given by Rakotondratsimba (2009). Such a modification is valid for portfolios with valuation formula, whose explicit dependency on the time can be expressed via the time left to maturity (final payment time, end of the contract). Among such portfolios are options based portfolios, in particular all portfolios from examples in Chapter 6. In order to make it more rigorous, we need some more notations.

For simplicity, we assume that the portfolio consists of options with the same maturity time T. Then, the portfolio value at time t and market price $\mathbf{S}$ at the same time t can be denoted as:

$$V(\mathbf{S},t) \equiv V(\mathbf{S};T-t) \equiv V(\mathbf{S},t,T),$$

where $V(\mathbf{S},t)$ is the notation from Section 3.1, $V(\mathbf{S};T-t)$ is the notation that shows the explicit dependency on the time left to maturity $T-t$, and $V(\mathbf{S},t,T)$ is the notation to emphasize the maturity T. According to these notations, the final value of the portfolio

at time $t + \Delta t$ is hence denoted by:

$$\begin{aligned} & V(\mathbf{S} + \Delta\mathbf{S}, t + \Delta t) \equiv V(\mathbf{S} + \Delta\mathbf{S}; T - \Delta t - t) \equiv \\ \equiv \quad & V(\mathbf{S} + \Delta\mathbf{S}, t + \Delta t, T) \equiv V(\mathbf{S} + \Delta\mathbf{S}, t, T - \Delta t), \end{aligned}$$

where the last equality shows the main advantage of the dependency on time via the time left to maturity. $V(\mathbf{S} + \Delta\mathbf{S}, t, T - \Delta t)$ denotes the portfolio value at time t with asset price $\mathbf{S} + \Delta\mathbf{S}$ and maturity $T - \Delta t$.

Using these notations, the corresponding changes in the portfolio value ΔV can be written as follows,

$$\begin{aligned} & \Delta V(\Delta\mathbf{S}, \Delta t) = V(\mathbf{S} + \Delta\mathbf{S}; T - \Delta t - t) - V(\mathbf{S}; T - t) = \\ = \quad & V(\mathbf{S}; T - \Delta t - t) - V(\mathbf{S}; T - t) + V(\mathbf{S} + \Delta\mathbf{S}; T - \Delta t - t) - V(\mathbf{S}; T - \Delta t - t) = \\ = \quad & V(\mathbf{S}, t, T - \Delta t) - V(\mathbf{S}, t, T) + V(\mathbf{S} + \Delta\mathbf{S}, t + \Delta t) - V(\mathbf{S}, t + \Delta t) = \\ = \quad & R(\mathbf{S}, t, T, \Delta t) + [V(\mathbf{S} + \Delta\mathbf{S}, t + \Delta t) - V(\mathbf{S}, t + \Delta t)], \end{aligned}$$

where $R(\mathbf{S}, t, T, \Delta t) = V(\mathbf{S}, t, T - \Delta t) - V(\mathbf{S}, t, T)$ is the difference of prices of two options at time t and market price $\mathbf{S}$ at the same time t with maturities $T - \Delta t$ and T respectively. This term is deterministic and easy to compute.

In order to achieve an approximation, the second term $V(\mathbf{S} + \Delta\mathbf{S}, t + \Delta t) - V(\mathbf{S}, t + \Delta t)$ now can be written in the Taylor series expansion w.r.t. the vector parameter of changes in risk factors and not time, as it does not change. This explains an advantage of the modified approximations compared to the classical ones: modified approximation is exact w.r.t. time. One can understand it as applying Taylor series expansion with all partial derivatives w.r.t. t up to infinite order. It should be also mentioned, that the form of the modified approximation is very close to the classical one, therefore, it is easy to adopt the implemented classical approximations for the purpose of computing the modified one. Besides, usually smaller time costs are necessary in order to compute such an approximation compared to the classical one.

3.2.1 Modified Delta - Gamma approximation

In this section we present a modified Delta-Gamma approximation, which is achieved with the help of a second order Taylor series expansion w.r.t. the changes in risk factors of $V(\mathbf{S} + \Delta\mathbf{S}, t + \Delta t)$ around the starting vector $\mathbf{S}$. The argument $t + \Delta t$ is fixed here.

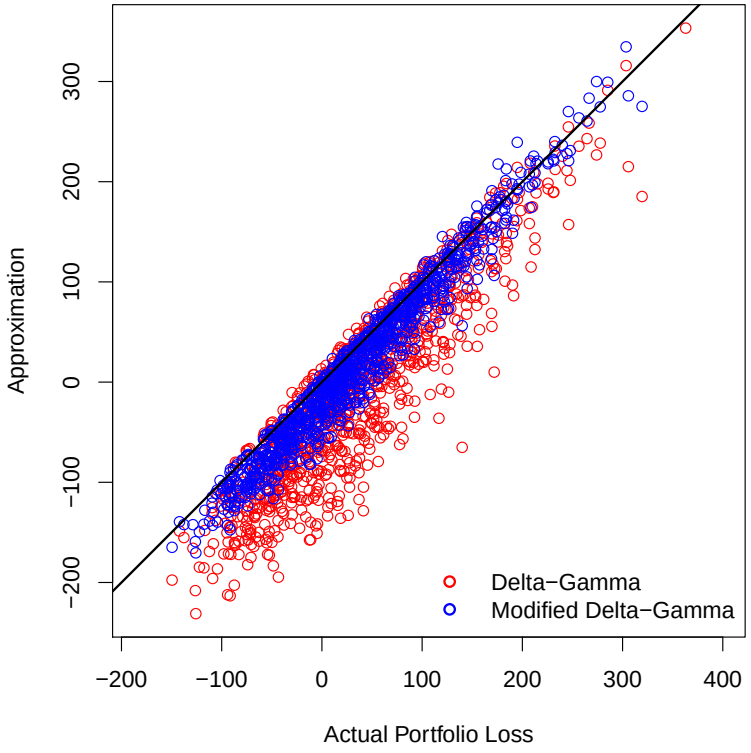


FIGURE 3.2: Comparison of Delta-Gamma and Modified Delta-Gamma approximations against actual portfolio losses in 1000 randomly generated scenarios for a portfolio of short 10 ATM down-and-out calls, 5 ATM puts, 10 uncorrelated assets, 0.1 years maturity, Gaussian changes in risk factors (for details on the portfolio see Chapter 6).

More precisely, the approximation reads as follows:

$$\Delta V \approx R(\mathbf{S}, t, T, \Delta t) + \sum_{i=1}^{d} \frac{\partial V(\mathbf{S}, t+\Delta t)}{\partial \mathbf{S}_i} \Delta \mathbf{S}_i + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 V(\mathbf{S}, t+\Delta t)}{\partial \mathbf{S}_i \partial \mathbf{S}_j} \Delta \mathbf{S}_i \Delta \mathbf{S}_j, \quad (3.5)$$

where $R(\mathbf{S}, t, T, \Delta t) = V(\mathbf{S}, t, T-\Delta t) - V(\mathbf{S}, t, T)$ under the notations from Section 3.2.

A comparison of the Delta-Gamma and this approximation for a portfolio of short 10 ATM down-and-out calls and 5 ATM puts with 0.1 years maturity on 10 uncorrelated assets (for details on portfolio see Chapter 6) is plotted against actual portfolio losses in the Figure 3.2.

3.2.2 Modified Charm - Speed approximation

Similarly to Section 3.2.1, a modified Charm-Speed approximation is achieved with the help of a third order Taylor series expansion w.r.t. the changes in risk factors of $V(\mathbf{S} + \Delta\mathbf{S}, t + \Delta t)$ around the starting vector $\mathbf{S}$. The argument $t + \Delta t$ is fixed here.

More precisely, the approximation reads as follows:

$$\begin{aligned}\Delta V &\approx R(\mathbf{S},t,T,\Delta t)+\sum_{i=1}^{d}\frac{\partial V(\mathbf{S},t+\Delta t)}{\partial \mathbf{S}_i}\Delta\mathbf{S}_i+\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 V(\mathbf{S},t+\Delta t)}{\partial \mathbf{S}_i\partial\mathbf{S}_j}\Delta\mathbf{S}_i\Delta\mathbf{S}_j+\\ &+\frac{1}{6}\sum_{i,j,k=1}^{d}\frac{\partial^3 V(\mathbf{S},t+\Delta t)}{\partial \mathbf{S}_i\partial\mathbf{S}_j\partial\mathbf{S}_k}\Delta\mathbf{S}_i\Delta\mathbf{S}_j\Delta\mathbf{S}_k, \qquad (3.6)\end{aligned}$$

where $R(\mathbf{S},t,T,\Delta t)=V(\mathbf{S},t,T-\Delta t)-V(\mathbf{S},t,T)$ under the notations from Section 3.2.

3.3 Explicit distribution of Delta-Gamma approximation

In this section we present the derivation of the explicit distribution of the Delta-Gamma approximation in two cases - multivariate normal and Student t-distributed changes in risk factors. Both derivation can be found in Glasserman (2004). They are crucial for the IS and ISS algorithms in both cases presented in Chapter 4.

3.3.1 Multivariate Gaussian case

We start here form the following assumption, under which the distribution of the Delta-Gamma approximation (3.3) is found in, e.g., Glasserman (2004):

Assumption 3.1. The underlying change in the risk factors $\Delta\mathbf{S}$ has a d-dimensional multivariate normal distribution (see e.g. Anderson (2003)) with zero mean vector, i.e. $\Delta\mathbf{S}\sim N_d(\mathbf{0},\mathbf{\Sigma}_S)$.

For the following discussion it is useful to rewrite approximation (3.3) in matrix form:

$$\Delta V\approx\frac{\partial V(\mathbf{S},t)}{\partial t}\Delta t+\boldsymbol{\delta}^T\Delta\mathbf{S}+\frac{1}{2}\Delta\mathbf{S}^T\mathbf{\Gamma}\Delta\mathbf{S}, \qquad (3.7)$$

where

$$\boldsymbol{\delta}_i=\frac{\partial V(\mathbf{S},t)}{\partial\mathbf{S}_i},\quad \mathbf{\Gamma}_{ij}=\frac{\partial^2 V(\mathbf{S},t)}{\partial\mathbf{S}_i\partial\mathbf{S}_j},$$

are first and second derivatives of V evaluated at $(\mathbf{S},t)$.

In order to simplify the derivation of the distribution of the Delta-Gamma approximation, some more modification of its expression are done. First, under Assumption 3.1, it is helpful to express the correlated normals $\Delta\mathbf{S}\sim N_d(\mathbf{0},\mathbf{\Sigma}_S)$ via independent normals $\mathbf{Z}\sim N_d(\mathbf{0},\mathbf{I})$ by setting,

$$\Delta\mathbf{S}=\mathbf{C}\mathbf{Z},\quad \text{with}\quad \mathbf{C}\mathbf{C}^T=\mathbf{\Sigma}_S.$$

Under a special choice of the matrix $\mathbf{C}$, the Delta-Gamma approximation for losses $L(\Delta\mathbf{S}, \Delta t) = -\Delta V$ in terms of $\mathbf{Z}$ becomes (see Glasserman (2004)):

$$L(\mathbf{CZ}) \approx a + \mathbf{b}^T\mathbf{Z} + \mathbf{Z}^T\mathbf{\Lambda Z} = a + \sum_{i=1}^{d}(\mathbf{b}_i\mathbf{Z}_i + \lambda_i\mathbf{Z}_i^2) = Q(\mathbf{Z}),$$

with $a = -\frac{\partial V(\mathbf{S},t)}{\partial t}\Delta t$, $b = -\mathbf{C}^T\boldsymbol{\delta}$ and $\mathbf{\Lambda} = \text{diag}(\lambda_i) = -\frac{1}{2}\mathbf{C}^T\mathbf{\Gamma C}$. The special choice of the matrix $\mathbf{C}$ is necessary in order to achieve the diagonalized matrix Λ. Since $\mathbf{\Sigma}_S$ is a covariance matrix, it can be expressed via $\mathbf{\Sigma}_S = \tilde{\mathbf{C}}\tilde{\mathbf{C}}^T$ by the Cholesky decomposition. Then, the matrix $-\frac{1}{2}\tilde{\mathbf{C}}^T\mathbf{\Gamma}\tilde{\mathbf{C}}$ is symmetric and can, therefore, be represented as:

$$-\frac{1}{2}\tilde{\mathbf{C}}^T\mathbf{\Gamma}\tilde{\mathbf{C}} = \mathbf{U\Lambda U}^T,$$

for some diagonal matrix $\mathbf{\Lambda}$ of eigenvalues and orthogonal matrix $\mathbf{U}$ ($\mathbf{UU}^T$ is the identity matrix) of eigenvectors. Finally, $\mathbf{C} = \tilde{\mathbf{C}}\mathbf{U}$ (see Glasserman (2004)).

With the help of the moment generating and characteristic functions (see definitions in Section 2.1.1), the distribution of the Delta-Gamma approximation $Q = Q(\mathbf{Z})$ is expressed via the inversion integral of the form (see Glasserman (2004)):

$$\mathbb{P}(Q \le x) - \mathbb{P}(Q \le x - y) = \frac{1}{\pi}\int_0^\infty Re\left(\hat{\phi}(u)\left[\frac{e^{iuy} - 1}{iu}\right]e^{-iux}\right)du, \tag{3.8}$$

where the cumulant generating function of Q is of the following form:

$$\psi(\theta) = a\theta + \frac{1}{2}\sum_{i=1}^{d}\left(\frac{\theta^2\mathbf{b}_i^2}{1 - 2\theta\lambda_i} - \log(1 - 2\theta\lambda_i)\right), \tag{3.9}$$

for all θ satisfying $\max_i \theta\lambda_i < \frac{1}{2}$. To find $\mathbb{P}(Q \le x)$ one should choose y large enough so that $\mathbb{P}(Q \le x - y) \approx 0$, see Glasserman (2004), Glasserman et al. (2000a,b).

3.3.2 Multivariate Student t-case

Similar to Section 3.3.1, we present here the exact distribution of the Delta-Gamma approximation (3.3) under the the following assumption (see e.g. Glasserman (2004)):

Assumption 3.2. The underlying change in the risk factors $\Delta\mathbf{S}$ has a d-dimensional multivariate Student t-distribution (see e.g. Anderson (2003)) with zero mean vector, i.e. $\Delta\mathbf{S} \sim t_d(\nu, \mathbf{0}, \mathbf{\Sigma}_S)$.

The matrix form of the Delta-Gamma approximation (3.7) does not depend on the distribution of the changes in risk factors, therefore, also applies here. Now, in order to

derive the distribution of the Delta-Gamma approximation under Assumption 3.2, it is useful to express the Student t random vector $\Delta\mathbf{S} \sim t_d(\nu, \mathbf{0}, \mathbf{\Sigma}_S)$ via standard Normal random vector $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$:

$$\Delta\mathbf{S} = \frac{\mathbf{CZ}}{\sqrt{Y/\nu}},$$

where $\mathbf{CC}^T = \mathbf{\Sigma}_S$ and $Y \sim \chi^2_\nu$ is a chi-squared distributed random variable with ν degrees of freedom and independent of the vector $\mathbf{Z}$. For the same purposes as in Section 3.3.1, the matrix $\mathbf{C}$ is chosen in the same special way as described in Section 3.3.1. Then, the Delta-Gamma approximation for losses $L(\Delta\mathbf{S}, \Delta t) = -\Delta V$ in terms of $\mathbf{Z}$ becomes (see Glasserman (2004)):

$$L\left(\frac{\mathbf{CZ}}{\sqrt{Y/\nu}}\right) \approx a + \frac{\mathbf{b}^T\mathbf{Z}}{\sqrt{Y/\nu}} + \frac{\mathbf{Z}^T\mathbf{\Lambda}\mathbf{Z}}{Y/\nu} = a + \sum_{i=1}^{d}\left(\mathbf{b}_i\frac{\mathbf{Z}_i}{\sqrt{Y/\nu}} + \lambda_i\frac{\mathbf{Z}_i^2}{Y/\nu}\right) = Q(\mathbf{Z}, Y),$$

with the same notations as in Section 3.3.1. Here we also have to write the dependence of the approximation on the random variable Y.

Glasserman (2004) suggests to use another representation of the Delta-Gamma approximation in this case, that he called indirect Delta-Gamma approximation. More precisely, for the fixed real value x, the following function is defined:

$$Q_x(\mathbf{Z}, Y) = (Q(\mathbf{Z}, Y) - x)(Y/\nu) = (a - x)(Y/\nu) + \sum_{i=1}^{d}\left(\mathbf{b}_i\sqrt{Y/\nu}\mathbf{Z}_i + \lambda_i\mathbf{Z}_i^2\right).$$

Then the distribution of the Delta-Gamma approximation under Assumption 3.2 can be expressed as:

$$\mathbb{P}(Q(\mathbf{Z}, Y) < x) = \mathbb{P}(Q_x(\mathbf{Z}, Y) < 0).$$

As mentioned in Glasserman (2004), the random variable $Q_x(\mathbf{Z}, Y)$ is easier to work with.

Conditional on Y, the approximation $Q_x(\mathbf{Z}, Y)$ has a similar form as the Delta-Gamma approximation for the Gaussian case, but with other coefficients (see Section 3.3.1), therefore, using the expression for the cumulant generating function in the case (3.9) the following holds true:

$$\mathbb{E}[e^{\theta Q_x(\mathbf{Z},Y)}|Y] = \exp\left\{\left((a - x)\theta/\nu + \frac{1}{2}\sum_{i=1}^{d}\frac{\theta^2\mathbf{b}_i^2/\nu}{1 - 2\theta\lambda_i}\right)Y\right\}\prod_{i=1}^{d}\frac{1}{\sqrt{1 - 2\theta\lambda_i}},$$

for all θ satisfying $\max_i \theta\lambda_i < \frac{1}{2}$.

If the coefficient of Y in the above expression is denoted by $a(\theta)$, more precisely:

$$a(\theta) = (a - x)\theta/\nu + \frac{1}{2}\sum_{i=1}^{d} \frac{\theta^2 \mathbf{b}_i^2/\nu}{1 - 2\theta\lambda_i},$$

then the above conditional expectation can be written in the form:

$$\mathbb{E}[e^{\theta Q_x(\mathbf{Z},Y)}|Y] = e^{a(\theta)Y} \prod_{i=1}^{d} \frac{1}{\sqrt{1 - 2\theta\lambda_i}}.$$

This expression helps in the derivation of the moment generating function of the random variable $Q_x(\mathbf{Z}, Y)$ (see Glasserman (2004)):

$$\begin{aligned} \phi_x(\theta) &= \mathbb{E}[e^{\theta Q_x(\mathbf{Z},Y)}] = \mathbb{E}[\mathbb{E}[e^{\theta Q_x(\mathbf{Z},Y)}|Y]] = \\ &= \mathbb{E}[e^{a(\theta)Y}] \prod_{i=1}^{d} \frac{1}{\sqrt{1 - 2\theta\lambda_i}} = (1 - 2a(\theta))^{-\nu/2} \prod_{i=1}^{d} \frac{1}{\sqrt{1 - 2\theta\lambda_i}}, \end{aligned} \tag{3.10}$$

for all θ satisfying $\max_i \theta\lambda_i < \frac{1}{2}$ and $a(\theta) < 0.5$, where in the last equality we have used the moment generating function of the chi-squared distributed random variable $Y \sim \chi^2_\nu$ with ν degrees of freedom. One should note, that the dependence on the value x is made here through $a(\theta)$.

According to Glasserman (2004), the expression (3.10) of the moment generating function of the random variable $Q_x(\mathbf{Z}, Y)$ is used to define the corresponding characteristic function according to the Definition 2.6. With the help of such a characteristic function, the probability $\mathbb{P}(Q_x(\mathbf{Z}, Y) < 0)$ is defined via inversion integral (3.8), thus giving the value of the probability $\mathbb{P}(Q(\mathbf{Z}, Y) < x)$.

As Glasserman (2004) mentions, the calculation of the distribution of the Delta-Gamma approximation in this case is less efficient than in the Gaussian case of the Section 3.3.1. This is because, calculating the $\mathbb{P}(Q(\mathbf{Z}, Y) < x)$ at each point x involves a separate inversion integral (3.8), since it depends on the value x. This might be an issue, if we need the distribution of the Delta-Gamma approximation in many different points.

Chapter 4

Variance Reduction of the Probability of High Portfolio Losses

The purpose of the thesis is to develop a new flexible and easy tool for an efficient estimation of the market risk measures. Following the research of Glasserman et al., in this chapter we, however, start from a slightly more general problem of variance reduction of the probability of high portfolio losses estimation. Such a quantity can be expressed as a mean, and is, therefore, easier to work with. Moreover, it can be also used in other applications of finance. Later, in Chapter 5, we discuss the corresponding transformation from the probability of high portfolio losses to VaR and CVaR, therefore developing the variance reduction procedures for the estimation of these market risk measures.

In order to describe the problem in more details, we first introduce notations consistent with those given in Chapter 3. First, we fix the time t at which the market risk measures are estimated. This time is also used as a starting time for the portfolio valuation and for the sake of simplicity is often omitted. Market risk measures are usually calculated over some time horizon, which is fixed and denoted by Δt. Again, for simplicity of notations we often omit it, when the explicit dependency is not important. Market risk is associated with unforeseen changes in risk factors over the corresponding time horizon Δt, which we denote by the d-dimensional random vector $\Delta\mathbf{S}$. To define the losses over this time horizon we also need do define starting values of the risk factors, which are given and fixed, and denoted by a d-dimensional vector $\mathbf{S} = \mathbf{S}(t)$. By $V(\mathbf{S}, t)$ and $V(\mathbf{S}+\Delta\mathbf{S}, t+\Delta t)$ we denote the portfolio value at time t and $t+\Delta t$ respectively. The random portfolio loss over the time horizon Δt is denoted by $L = L(\Delta\mathbf{S}) = L(\Delta\mathbf{S}, \Delta t)$

and defined as:

$$L(\Delta\mathbf{S}) = -\Delta V(\Delta\mathbf{S}, \Delta t) = V(\mathbf{S}, t) - V(\mathbf{S} + \Delta\mathbf{S}, t + \Delta t).$$

Any approximation of the portfolio losses is also a function of the random vector $\Delta\mathbf{S}$ and is denoted by $Q = Q(\Delta\mathbf{S}) = Q(\Delta\mathbf{S}, \Delta t)$. Also, in this notation, we often omit the fixed time horizon Δt.

The following assumption is crucial in the development of the variance reduction procedures for high portfolio losses estimation:

Assumption 4.1. Revaluation time of the approximation Q is negligible in comparison to the time necessary to revalue the portfolio L.

Assumption 4.1 cannot be made more rigorous, as it highly depends on the implementation of the application, the dimension and the complexity of the portfolio. According to numerical tests in Chapter 6, we expect, that an approximation should be calculated at least $100 - 1000$ times faster than the portfolio loss. This is the case for portfolios considered in our numerical study in Chapter 6. Since those portfolios are quite simple, we expect the speed-up rate in real applications to be even higher. We would like to mention, that under optimal implementation, the time consumption of the revaluation of an approximation Q is fixed, as it is usually a polynomial. The portfolio loss revaluation depends on the corresponding products and, potentially, can be extremely heavy in calculation. The underlying dimension is the same in both cases (for Q and L), but for an approximation Q it can even be reduced (see e.g. Albanese et al. (2002)).

Under Assumption 4.1, an approximation Q can significantly reduce the additional effort in setting up a variance reduction procedure without loss of efficiency. Hence, this makes the corresponding alternative estimators useful in practice. Even if such an approximation is not accurate, numerical results in Chapter 6 show only a little deterioration of the performance compared to the original loss function. The approximation Q, therefore, is used to guide the choice of the importance sampling change of measure for variance reduction procedures.

We now define the quantity, for which the variance reduction procedures are developed:

$$l = \mathbb{P}(L(\Delta\mathbf{S}) > x), \tag{4.1}$$

for some fixed and known level x.

We remind the reader that, under Assumption 2.1, the density function of the changes in risk factors $\Delta\mathbf{S}$ is denoted by $f(\mathbf{x}; \mathbf{u})$, with fixed original vector parameter $\mathbf{u}$, out of the family of densities $\mathcal{G} = \{f(\mathbf{x}; \mathbf{v}), \mathbf{v} \in \mathcal{V}\}$ parametrized by $\mathbf{v}$.

The MC estimator (2.2) under these notations reads as follows:

$$\hat{l} = \frac{1}{N} \sum_{i=1}^{N} I_{(x,+\infty)}(L(\Delta \mathbf{S}_i)),$$

where $\Delta \mathbf{S}_1, \ldots, \Delta \mathbf{S}_N$ is a sample of independent identically distributed random vectors with joint density $f(\mathbf{x}; \mathbf{u})$.

Such a probability l fits into Assumption 2.4, therefore, the variance reduction procedures presented in Chapter 2 can be applied. Similar to the research of Glasserman et al., in Section 4.1 we start from applying the importance sampling procedures for the approximation Q in order to find an appropriate change of measure, which is further used in the importance sampling estimator for the original losses. By applying additionally an original stratification in Section 4.2 we improve the variance reduction even further. Sections 4.1.1, 4.1.2, 4.2.1 and 4.2.2 present algorithms from Glasserman (2004), Glasserman et al. (2000a,b, 2002). The light-tailed case of multivariate Gaussian distributed changes in risk factors is considered in Sections 4.1.1 and 4.2.1, whereas the heavy-tailed case of a multivariate Student t-distribution is given in Sections 4.1.2 and 4.2.2. Any other model of changes in risk factors in the context of market risk estimation is not covered by the literature. Ideas of other methods introduced in the rest of the Sections 4.1 and 4.2 are presented in the publication of mine Pupashenko (2014). Some improvements and detailed explanation of these methods are, however, first given here.

For the desired flexibility of the algorithms an easy and efficient special stratification, first introduced in my joint work with R. Korn in Korn and Pupashenko (2014), is presented in details in Section 4.3.1. The corresponding modification of the algorithms is further developed in the remainder of the Section 4.3.

An extra improvement of the efficiency of the new algorithms is achieved by improving the importance sampling part in Section 4.4. Here we develop an improved version of the algorithms, which are first presented in my joint paper with R. Korn in Korn and Pupashenko (2014). Section 4.5 summarizes and compares all algorithms from Chapter 4 with regard to the time consumption.

4.1 Importance Sampling based algorithms

In this section we develop variance reduction algorithms based only on different importance sampling procedures. They are important in developing further methods by applying additional stratification. Besides, in the numerical study in Chapter 6, the difference in performance is showed.

Numerical study in Chapter 6 shows approximately the same performance for all methods given in this section. The algorithms presented in Sections 4.1.1 and 4.1.2 are developed in Glasserman (2004), Glasserman et al. (2000a,b, 2002). The algorithms from Sections 4.1.3 and 4.1.4 are direct consequences of the theory from Sections 2.1.2 and 2.1.3. However, in the context of market risk estimation, they are first applied in the paper of mine (Pupashenko (2014)) and here. The main advantage of these new methods is, that they can use any function of the changes in risk factors Q which approximates the original loss function L in order to guide the choice of the importance sampling change of measure. These also include the original loss function L, that gives the best precision of an approximation (exact approximation), but is much less time efficient according to Assumption 4.1. Only in the case of an extremely inaccurate approximation Q, the methods based on the original loss function L show better performance (for details, see Chapter 6).

4.1.1 Exponential twisting importance sampling: Gaussian case

In this section we present the variance reduction method for the estimation of the probability of high portfolio losses based on the exponential twisting in the case of Gaussian changes in risk factors, i.e. under Assumption 3.1. The corresponding method is developed in Glasserman (2004), Glasserman et al. (2000a,b) and is called Importance Sampling (IS) algorithm.

According to Glasserman (2004), Glasserman et al. (2000a,b), under Assumption 3.1, the correlated changes in risk factors $\Delta\mathbf{S}$ are modeled via independent normals $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$ in the following way (see Section 3.3.1):

$$\Delta\mathbf{S} = \mathbf{C}\mathbf{Z},$$

where the matrix $\mathbf{C}$ is chosen is such a way, that $\mathbf{C}\mathbf{C}^T = \mathbf{\Sigma}_S$ and the Delta-Gamma approximation can be written as (for details on how to choose such a matrix see Section 3.3.1):

$$L(\Delta\mathbf{S}) = L(\mathbf{C}\mathbf{Z}) \approx Q(\mathbf{Z}) = a + \sum_{i=1}^{d}(\mathbf{b}_i\mathbf{Z}_i + \lambda_i\mathbf{Z}_i^2),$$

under the notations from the Section 3.3.1.

Since we perform the exponential twisting for the approximation $Q(\mathbf{Z})$, in order to keep notations consistent, in this section we treat $\mathbf{Z}$ as original changes in risk factors, denoting its density by $f(\mathbf{x}; \mathbf{u})$ as before. This can be done simply by applying a composition

of portfolio loss function and its Delta-Gamma approximation together with the function $\Delta\mathbf{S}(\mathbf{Z}) = \mathbf{CZ}$. The original parameter $\mathbf{u}$ corresponds then to the identity matrix, since $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$.

It is important to note, that the IS algorithm in this section is restricted to the Delta-Gamma approximation expressed in the form above. Assumption 3.1 is crucial here, since an explicit form of the distribution of the Delta-Gamma approximation is found in this case (see Section 3.3.1). Hence, along this section, the role of the approximation Q is played only by the Delta-Gamma approximation.

Figures 3.1 and 3.2 illustrate a strong correlation of the exact and approximate losses in the corresponding examples. Because of the Assumption 4.1, it is worth performing an importance sampling to estimate $\mathbb{P}(Q(\mathbf{Z}) > x)$. The achieved change of measure is, therefore, also a good candidate as an importance sampling change of measure for the original estimate l in (4.1). Hence, an exponential twisting is first applied to estimate the probability $\mathbb{P}(Q(\mathbf{Z}) > x)$.

The importance sampling density is then achieved by the exponential change of measure (2.9) with $G(\mathbf{X})$ being $Q(\mathbf{Z})$:

$$f_\theta(\mathbf{x}; \mathbf{u}) = e^{\theta Q(\mathbf{x}) - \psi(\theta)} f(\mathbf{x}; \mathbf{u}),$$

where $\psi(\theta)$ denotes a cumulant generating function of $Q(\mathbf{Z})$ and, under Assumption 3.1, is given by (3.9).

The parameter θ is chosen in order to minimize the variance of the importance sampling estimator of $\mathbb{P}(Q(\mathbf{Z}) > x)$ under the above mentioned exponential change of measure. It is given as the root of the following equation (see Section 2.1.1):

$$\psi'(\theta) = \mathbb{E}_\theta[Q(\mathbf{Z})] = x.$$

We should also remind, that the exponential change of measure is defined only for those θ, for which $\psi(\theta) < \infty$. In our case this condition is satisfied if $\max_i \lambda_i \theta < \frac{1}{2}$.

Assuming an optimal choice of the parameter θ according to the above mentioned equation, the corresponding exponential change of measure is then applied to estimate the original probability l in (4.1):

$$l = \mathbb{P}(L(\mathbf{CZ}) > x) = \mathbb{E}_\theta \left[e^{-\theta Q(\mathbf{Z}) + \psi(\theta)} I_{(x, +\infty)}(L(\mathbf{CZ})) \right].$$

Again, due to unbiasedness, instead of investigating the variance, we can consider only a second moment of the form:

$$\mathbb{E}_\theta\left[e^{-2\theta Q(\mathbf{Z})+2\psi(\theta)}I_{(x,+\infty)}(L(\mathbf{CZ}))\right]=\mathbb{E}\left[e^{-\theta Q(\mathbf{Z})+\psi(\theta)}I_{(x,+\infty)}(L(\mathbf{CZ}))\right],$$

which is small if $\theta>0$ (under our choice) and Q is large whenever $L>x$, which is the case due to their correlation.

As we mentioned at the end of Section 2.1.1, in order to define the IS algorithm based on the exponential twisting and Delta-Gamma approximation, it is only left to define how to sample from the achieved exponential change of measure. For that purpose it is enough here to derive the distribution of $\mathbf{Z}$ under this exponential change of measure.

In Glasserman et al. (2000b) is shown that, under the exponential change of measure, the vector $\mathbf{Z}$ still has multivariate normal distribution, but with a new mean vector and covariance matrix, that depend on the parameter θ and the Delta-Gamma approximation. More precisely, under $f_\theta(\mathbf{x};\mathbf{u})$ holds $\mathbf{Z}\sim N_d(\boldsymbol{\mu}(\theta),\boldsymbol{\Sigma}(\theta))$, where $\boldsymbol{\Sigma}(\theta)$ is a diagonal matrix with diagonal entries $\sigma_i^2(\theta)$,

$$\boldsymbol{\mu}_i(\theta)=\frac{\theta b_i}{1-2\lambda_i\theta},\quad \sigma_i^2(\theta)=\frac{1}{1-2\lambda_i\theta},\quad i=1,\dots,d. \tag{4.2}$$

This gives a simple procedure of sampling from the exponential change of measure. Recall that, in order to define the exponential change of measure, we require $\max_i\theta\lambda_i<\frac{1}{2}$, which is important for the correctness of the above defined covariance matrix.

All said above summarizes in the following IS algorithm for the case of Gaussian distributed changes in risk factors, i.e. under Assumption 3.1:

Algorithm 4.1 (IS: Gaussian case)**.**

1. *For the changes in risk factors $\Delta\mathbf{S}\sim N_d(\mathbf{0},\boldsymbol{\Sigma}_S)$ calculate the decomposition of the covariance matrix $\boldsymbol{\Sigma}_S=\mathbf{CC}^T$, such that the Delta-Gamma approximation takes the form:*

$$Q(\mathbf{Z})=a+\sum_{i=1}^{d}(\mathbf{b}_i\mathbf{Z}_i+\lambda_i\mathbf{Z}_i^2),$$

for $\mathbf{Z}\sim N_d(\mathbf{0},\mathbf{I})$ (see Section 3.3.1). Fix input parameter N.

2. *Denote by $\theta>0$ the positive solution to the equation:*

$$\psi'(\theta)=x,$$

such that $\max_i\theta\lambda_i<\frac{1}{2}$, where $\psi(\theta)$ is the cumulant generating function of $Q(\mathbf{Z})$ and is given by (3.9).

3. Estimate the probability l using the importance sampling estimator (2.4) in the form

$$\hat{l} = \frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\mathbf{C}\mathbf{Z}_i))e^{-\theta Q(\mathbf{Z}_i)+\psi(\theta)}, \tag{4.3}$$

where $\mathbf{Z}_1, ..., \mathbf{Z}_N$ is a sample generated from the density $N_d(\boldsymbol{\mu}(\theta), \boldsymbol{\Sigma}(\theta))$ with mean vector and diagonal covariance matrix given by (4.2).

One should note, that the solution to the equation $\psi'(\theta) = x$ in step 2 of Algorithm 4.1 has to be found numerically, and in a high-dimensional setting it might become challenging. By the dimension we do not mean the dimension of the parameter θ which is always one. We rather mean the dimension of the changes in risk factors, which influence the expression of the function $\psi(\theta)$ and makes its evaluation more time consuming.

4.1.2 Exponential twisting importance sampling: Student t-case

This section is very similar to Section 4.1.1. Here we present the variance reduction method for the estimation of the probability of high portfolio losses based on the exponential twisting in the case of Student t-distributed changes in risk factors, i.e. under Assumption 3.2. The corresponding method is developed in Glasserman (2004), Glasserman et al. (2002) and is called Importance Sampling (IS) algorithm. We should note, that the name of this algorithm is the same as in Section 4.1.1, and we differ them only by the distribution of the changes in risk factors. In the numerical study in Chapter 6 it is rather clear, which version of the IS algorithm is used.

According to Glasserman (2004), Glasserman et al. (2002), under Assumption 3.2, the Student t-distributed changes in risk factors $\Delta\mathbf{S}$ are modeled via independent normals $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$ and the chi-squared distributed random variable $Y \sim \chi^2_\nu$ with ν degrees of freedom, that is independent of the vector $\mathbf{Z}$, in the following way (see Section 3.3.2):

$$\Delta\mathbf{S} = \frac{\mathbf{C}\mathbf{Z}}{\sqrt{Y/\nu}},$$

where the matrix $\mathbf{C}$ is chosen in such a way, that $\mathbf{C}\mathbf{C}^T = \boldsymbol{\Sigma}_S$ and the Delta-Gamma approximation can be written as (for details on how to choose such a matrix see Section 3.3.1):

$$L(\Delta\mathbf{S}) = L\left(\frac{\mathbf{C}\mathbf{Z}}{\sqrt{Y/\nu}}\right) \approx Q(\mathbf{Z}, Y) = a + \sum_{i=1}^{d}\left(\mathbf{b}_i\frac{\mathbf{Z}_i}{\sqrt{Y/\nu}} + \lambda_i\frac{\mathbf{Z}_i^2}{Y/\nu}\right),$$

under the notations from Sections 3.3.1 and 3.3.2.

As already mentioned in Section 3.3.2, in the Student t-case it is easier to work with the indirect Delta-Gamma approximation $Q_x(\mathbf{Z}, Y)$ for some fixed real value x instead of working with $Q(\mathbf{Z}, Y)$ directly. This is given by (see Section 3.3.2):

$$Q_x(\mathbf{Z}, Y) = (a - x)(Y/\nu) + \sum_{i=1}^{d} \left(\mathbf{b}_i \sqrt{Y/\nu} \mathbf{Z}_i + \lambda_i \mathbf{Z}_i^2 \right),$$

under the notations from Sections 3.3.1 and 3.3.2. The role of the fixed real value x here is played by the level x from the probability of interest (4.1), therefore, it is denoted accordingly.

Similarly to Section 3.3.1, the importance sampling density is achieved by the exponential change of measure (2.9) with $G(\mathbf{X})$ being $Q_x(\mathbf{Z}, Y)$:

$$f_\theta(\mathbf{x}, y; \mathbf{u}) = e^{\theta Q_x(\mathbf{x}, y) - \psi_x(\theta)} f(\mathbf{x}, y; \mathbf{u}),$$

where $\psi_x(\theta)$ denotes the cumulant generating function of $Q_x(\mathbf{Z}, Y)$ and, under Assumption 3.2, can easily be achieved from (3.10) and Definition 2.7. The notation of the original density $f(\mathbf{x}, y; \mathbf{u})$ can be understood as a familiar notation $f(\tilde{\mathbf{x}}; \mathbf{u})$ with $\tilde{\mathbf{x}}$ being the vector that consists of the vector $\mathbf{x}$ extended by the value y.

The corresponding exponential change of measure is then applied to estimate the original probability l in (4.1):

$$l = \mathbb{P}\left(L\left(\frac{\mathbf{CZ}}{\sqrt{Y/\nu}} \right) > x \right) = \mathbb{E}_\theta \left[e^{-\theta Q_x(\mathbf{Z}, Y) + \psi_x(\theta)} I_{(x, +\infty)} \left(L\left(\frac{\mathbf{CZ}}{\sqrt{Y/\nu}} \right) \right) \right].$$

Due to unbiasedness of the estimator, instead of investigating the variance, we can consider only the second moment of the form:

$$\begin{aligned} &\mathbb{E}_\theta \left[e^{-2\theta Q_x(\mathbf{Z}, Y) + 2\psi_x(\theta)} I_{(x, +\infty)} \left(L\left(\frac{\mathbf{CZ}}{\sqrt{Y/\nu}} \right) \right) \right] = \\ &= \mathbb{E} \left[e^{-\theta Q_x(\mathbf{Z}, Y) + \psi_x(\theta)} I_{(x, +\infty)} \left(L\left(\frac{\mathbf{CZ}}{\sqrt{Y/\nu}} \right) \right) \right], \end{aligned}$$

which is small if $-\theta Q_x(\mathbf{Z}, Y) + \psi_x(\theta)$ is small whenever $L\left(\frac{\mathbf{CZ}}{\sqrt{Y/\nu}} \right) > x$.

As already mentioned in Section 4.1.1, when $L\left(\frac{\mathbf{CZ}}{\sqrt{Y/\nu}} \right) > x$ we often have $Q(\mathbf{Z}, Y) > x$ and thus $Q_x(\mathbf{Z}, Y) > 0$. Hence, in order to keep $-\theta Q_x(\mathbf{Z}, Y) + \psi_x(\theta)$ small, under the choice of the positive parameter $\theta > 0$, the value of $\psi_x(\theta)$ should be small. Glasserman (2004), Glasserman et al. (2002) define, therefore, the optimal choice of the parameter

θ to be the positive root of the following equation:

$$\psi'_x(\theta) = \mathbb{E}_\theta[Q_x(\mathbf{Z}, Y)] = 0.$$

This correspond to the familiar choice of the exponential twisting parameter according to $\mathbb{E}_\theta[Q(\mathbf{Z}, Y)] = x$ (see Sections 2.1.1 and 4.1.1).

Similarly to Section 4.1.1, in order to define the IS algorithm based on the exponential twisting and Delta-Gamma approximation it is only left to define how to sample from the achieved exponential change of measure. For that purpose, it is enough here to derive the distribution of the vector $\mathbf{Z}$ and variable Y under this exponential change of measure. With the help of such simulated quantities the approximation and the losses are evaluated accordingly. One should also note, that under the original measure, the vector $\mathbf{Z}$ and the variable Y are independent. This, however, is not the case under the exponential change of measure.

In Glasserman et al. (2002) is shown, that under exponential change of measure, the vector $\mathbf{Z}$ conditional on Y has multivariate normal distribution $N_d(\boldsymbol{\mu}(\theta, Y), \boldsymbol{\Sigma}(\theta))$ with mean vector $\boldsymbol{\mu}(\theta, Y)$ and diagonal covariance matrix $\boldsymbol{\Sigma}(\theta)$ with diagonal entries $\sigma_i^2(\theta)$ given by:

$$\boldsymbol{\mu}_i(\theta, Y) = \frac{\theta b_i \sqrt{Y/\nu}}{1 - 2\lambda_i \theta}, \quad \sigma_i^2(\theta) = \frac{1}{1 - 2\lambda_i \theta}, \quad i = 1, \ldots, d. \tag{4.4}$$

Under the exponential change of measure, the random variable Y has the gamma distribution $\Gamma\left(\frac{\nu}{2}, \frac{2}{1-2a(\theta)}\right)$ with shape parameter $\nu/2$ and scale parameter $2/(1 - 2a(\theta))$, where $a(\theta)$ is defined in Section 3.3.2.

This all gives a simple procedure of sampling from the exponential change of measure. More precisely, the variable Y is generated first from the gamma distribution, and then is used to change the mean and the covariance matrix of the normally distributed vector $\mathbf{Z}$, that is further generated. Recall that, in order to define exponential change of measure, we require $\max_i \theta\lambda_i < \frac{1}{2}$ and $a(\theta) < 0.5$, which is important for the correctness of the above defined distributions.

All said above summarizes in the following IS algorithm for the case of Student t-distributed changes in risk factors, i.e. under Assumption 3.2:

Algorithm 4.2 (IS: Student t-case)**.**

1. *For the changes in risk factors $\Delta\mathbf{S} \sim t_d(\nu, \mathbf{0}, \boldsymbol{\Sigma}_S)$ calculate the decomposition of the covariance matrix $\boldsymbol{\Sigma}_S = \mathbf{C}\mathbf{C}^T$, such that the indirect Delta-Gamma approximation*

takes the form:

$$Q_x(\mathbf{Z}, Y) = (a - x)(Y/\nu) + \sum_{i=1}^{d} \left(\mathbf{b}_i \sqrt{Y/\nu} \mathbf{Z}_i + \lambda_i \mathbf{Z}_i^2 \right),$$

for $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$ *and* $Y \sim \chi^2_\nu$*, that is independent of the vector* $\mathbf{Z}$ *(see Section 3.3.2). Fix input parameter* N*.*

2. *Denote by* $\theta > 0$ *the positive solution to the equation:*

$$\psi'_x(\theta) = 0,$$

such that $\max_i \theta\lambda_i < \frac{1}{2}$ *and* $a(\theta) < 0.5$*, where* $\psi_x(\theta) = \log \phi_x(\theta)$ *is the cumulant generating function of* $Q_x(\mathbf{Z}, Y)$ *and is given via (3.10) with* $a(\theta)$ *defined in Section 3.3.2.*

3. *Estimate the probability* l *using the importance sampling estimator (2.4) in the form*

$$\hat{l} = \frac{1}{N} \sum_{i=1}^{N} I_{(x,\infty)} \left(L \left(\frac{\mathbf{C}\mathbf{Z}_i}{\sqrt{Y_i/\nu}} \right) \right) e^{-\theta Q_x(\mathbf{Z}_i, Y_i) + \psi_x(\theta)}, \tag{4.5}$$

where, given the sample $Y_1, ..., Y_N$ *generated from the density* $\Gamma\left(\frac{\nu}{2}, \frac{2}{1-2a(\theta)}\right)$*, the vectors* $\mathbf{Z}_i$ *are generated from the densities* $N_d(\boldsymbol{\mu}(\theta, Y_i), \boldsymbol{\Sigma}(\theta))$ *for* $i = 1, \ldots, N$ *with mean vector and diagonal covariance matrix given by (4.4).*

Similarly to the Algorithm 4.1, the solution to the equation $\psi'_x(\theta) = 0$ in step 2 of Algorithm 4.2 has to be also found numerically.

4.1.3 Cross - Entropy importance sampling

In this Section we present the variance reduction method for the estimation of the probability of high portfolio losses based on the CE algorithm from the Section 2.1.2. According to our knowledge, this algorithm has not been applied before in the application of the market risk estimation. Our choice of this method is explained by the ability to apply it for any distribution of the changes in risk factors, no matter if heavy- or light-tailed. Besides, in contrast to the IS algorithm from Sections 4.1.1 and 4.1.2, the derivation of this method does not involve any exact form of the distribution of the approximation of the loss function or its characteristic function. This gives us an extra flexibility in the choice of the approximation function, which theoretically can now also be chosen to be exact, i.e. simply a portfolio loss function. First, we applied the CE

algorithm in the context of the market risk estimation in the paper of mine (Pupashenko (2014)).

Similarly to previous algorithms, we first apply the CE algorithm (see Section 2.1.2) to some function $Q(\Delta\mathbf{S})$ of the changes in risk factors which approximates the original loss function $L(\Delta\mathbf{S})$. The optimal reference parameter, found together with the corresponding change of measure, is used to construct the importance sampling estimator of the probability l from (4.1). For practical reasons, Assumption 4.1 has also to be satisfied. In such a case we call the corresponding algorithm the Approximate Cross-Entropy (ACE) algorithm. It can easily be achieved from Algorithm 2.1 by substituting the level γ by the level x and the function G by the function Q in steps 1-4 (choice of the optimal reference parameter) and by the function L in step 5 (importance sampling estimator). The role of the random vector $\mathbf{X}$ now is played by the vector of the changes in risk factors $\Delta\mathbf{S}$.

All said above summarizes in the following ACE algorithm under Assumption 4.1:

Algorithm 4.3 (ACE)**.**

1. *Define* $\hat{\mathbf{v}}_0 = \mathbf{u}$ *and set* $j = 1$*. Fix input parameters* ρ*,* M *and* N*.*

2. *Generate a sample* $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$ *from the density* $f(\mathbf{x}; \hat{\mathbf{v}}_{j-1})$ *and compute the sample* $(1-\rho)-$ *quantile* $\hat{x}_j$ *of the sample* $Q(\Delta\mathbf{S}_1), ..., Q(\Delta\mathbf{S}_M)$*. If* $\hat{x}_j > x$*, set* $\hat{x}_j = x$*.*

3. *Using the same sample* $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$*, find an estimate from the adaptive optimization problem in the form of (2.12) or (2.13),*

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M} \sum_{i=1}^{M} I_{(\hat{x}_j,\infty)}(Q(\Delta\mathbf{S}_i)) W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}_{j-1}) \log f(\Delta\mathbf{S}_i; \mathbf{v}), \tag{4.6}$$

   and denote the solution as $\hat{\mathbf{v}}_j$*.*

4. *If* $\hat{x}_j < x$*, set* $j = j + 1$ *and reiterate from step 2, else set an optimal reference parameter estimate* $\hat{\mathbf{v}}^* = \hat{\mathbf{v}}_j$ *and proceed with step 5.*

5. *Estimate the probability* l *using the importance sampling estimator (2.4) in the form*

$$\hat{l} = \frac{1}{N} \sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i)) W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*), \tag{4.7}$$

   where $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ *is a sample generated from the density* $f(\mathbf{x}; \hat{\mathbf{v}}^*)$*.*

As we already mentioned before, the original portfolio loss function L, theoretically, can also be used as an approximation Q. In this case Assumption 4.1 is, however, violated,

making such an approach useless in practice. But, in order to investigate the influence of the quality of the approximation on the variance reduction rate, the Full Cross-Entropy (FCE) algorithm is introduced here. This method is a simple CE algorithm applied directly to the problem of the estimation of the probability of high portfolio losses (4.1). As the numerical study in the Chapter 6 shows, in the most situations, only a little improvement in performance can be achieved by using the original loss instead of its approximation.

The FCE algorithm is achieved by substituting the function Q in Algorithm 4.3 by the function L, more precisely:

Algorithm 4.4 (FCE)**.**

1. *Define* $\hat{\mathbf{v}}_0 = \mathbf{u}$ *and set* $j = 1$*. Fix input parameters* ρ*,* M *and* N*.*

2. *Generate a sample* $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$ *from the density* $f(\mathbf{x}; \hat{\mathbf{v}}_{j-1})$ *and compute the sample* $(1-\rho)-$ *quantile* $\hat{x}_j$ *of the sample* $L(\Delta\mathbf{S}_1), ..., L(\Delta\mathbf{S}_M)$*. If* $\hat{x}_j > x$*, set* $\hat{x}_j = x$*.*

3. *Using the same sample* $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$*, find an estimate from the adaptive optimization problem in the form of (2.12) or (2.13),*

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M} \sum_{i=1}^{M} I_{(\hat{x}_j,\infty)}(L(\Delta\mathbf{S}_i)) W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}_{j-1}) \log f(\Delta\mathbf{S}_i; \mathbf{v}), \tag{4.8}$$

   and denote the solution as $\hat{\mathbf{v}}_j$*.*

4. *If* $\hat{x}_j < x$*, set* $j = j + 1$ *and reiterate from step 2, else set an optimal reference parameter estimate* $\hat{\mathbf{v}}^* = \hat{\mathbf{v}}_j$ *and proceed with step 5.*

5. *Estimate the probability* l *using the importance sampling estimator (2.4) in the form*

$$\hat{l} = \frac{1}{N} \sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i)) W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*), \tag{4.9}$$

   where $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ *is a sample generated from the density* $f(\mathbf{x}; \hat{\mathbf{v}}^*)$*.*

We should note, that the optimal reference parameter estimation involves M multiplied by the number of the iterations j of the revaluation of the portfolio loss function. This is obviously extremely computationally heavy. In the numerical study in Chapter 6, the number of iterations is often 4. Together with the choice of $M = 10000$ it involves, therefore, 40000 revaluations of the portfolio. This is approximately the same time as necessary in order to achieve the MC estimator with the sample size of $N = 40000$.

4.1.4 Improved Cross - Entropy importance sampling

This section is related to applying ICE algorithm from Section 2.1.3 to the problem of estimating the probability of high portfolio losses. This is already done in Chan and Kroese (2012), but for the credit risk application, which differs from the market risk estimation of our interest. According to our knowledge, in the application of the market risk estimation, this algorithm is first applied here. As one can conclude from the name, this is an improved version of the CE algorithm. As already mentioned in Section 2.1.3, it is introduced in order to deal with the degeneracy of the likelihood ratio estimator problem. Since, in the context of market risk estimation we usually deal with not very small probabilities (near 0.01), we do not expect to face this problem. But, for completeness of our investigation, it is important to apply the ICE algorithm and compare it to the CE method.

Similarly to Section 4.1.3, the ICE algorithm (see Section 2.1.3) is first applied to some function $Q(\Delta\mathbf{S})$ of the changes in risk factors which approximates the original loss function $L(\Delta\mathbf{S})$. The optimal reference parameter found, together with the corresponding change of measure, is used to construct the importance sampling estimator of the probability l from (4.1). In such a case we call the corresponding algorithm the Approximate Improved Cross-Entropy (AICE) algorithm. It can easily be achieved from the Algorithm 2.3 by substituting the level γ by the level x and the function G by the function Q in steps 1-3 (choice of the optimal reference parameter) and by the function L in step 4 (importance sampling estimator). The role of the random vector $\mathbf{X}$ now is played by the vector of the changes in risk factors $\Delta\mathbf{S}$.

All said above summarizes in the following AICE algorithm under Assumption 4.1:

Algorithm 4.5 (AICE)**.**

1. *Fix input parameters M and N.*
2. *Generate a sample $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$ from the density*

$$g^*(\mathbf{x}) = \frac{I_{(x,+\infty)}(Q(\mathbf{x}))f(\mathbf{x};\mathbf{u})}{\mathbb{P}(Q(\Delta\mathbf{S}) > x)},$$

   with the help of the methods from Appendix A.
3. *Using this sample find an estimate from the optimization problem,*

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M}\sum_{i=1}^{M} \log f(\Delta\mathbf{S}_i;\mathbf{v}), \tag{4.10}$$

   and denote the solution as $\hat{\mathbf{v}}^*$.

4. *Estimate the probability l using the importance sampling estimator (2.4) in the form*

$$\hat{l} = \frac{1}{N} \sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta \mathbf{S}_i)) W(\Delta \mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*), \tag{4.11}$$

where $\Delta \mathbf{S}_1, ..., \Delta \mathbf{S}_N$ is a sample generated from the density $f(\mathbf{x}; \hat{\mathbf{v}}^*)$.

For the same reasons as in Section 4.1.3, the Full Improved Cross-Entropy (FICE) algorithm is introduced here. This method is a simple ICE algorithm applied directly to the problem of the estimation of the probability of high portfolio losses (4.1). Again, according to the numerical study in the Chapter 6, in most situations, only a little improvement in performance can be achieved by using the original losses instead of its approximation. But it can be done only with significant extra time costs, since the optimal reference parameter estimator involves some portfolio revaluations.

FICE algorithm is achieved by substituting the function Q in the Algorithm 2.3 by the function L, more precisely:

Algorithm 4.6 (FICE)**.**

1. *Fix input parameters M and N.*

2. *Generate a sample $\Delta \mathbf{S}_1, ..., \Delta \mathbf{S}_M$ from the density*

$$g^*(\mathbf{x}) = \frac{I_{(x,+\infty)}(L(\mathbf{x})) f(\mathbf{x}; \mathbf{u})}{l},$$

with the help of the methods from Appendix A.

3. *Using this sample find an estimate from the optimization problem,*

$$\max_{\mathbf{v} \in \mathcal{V}} \frac{1}{M} \sum_{i=1}^{M} \log f(\Delta \mathbf{S}_i; \mathbf{v}), \tag{4.12}$$

and denote the solution as $\hat{\mathbf{v}}^*$.

4. *Estimate the probability l using the importance sampling estimator (2.4) in the form*

$$\hat{l} = \frac{1}{N} \sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta \mathbf{S}_i)) W(\Delta \mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*), \tag{4.13}$$

where $\Delta \mathbf{S}_1, ..., \Delta \mathbf{S}_N$ is a sample generated from the density $f(\mathbf{x}; \hat{\mathbf{v}}^*)$.

All approaches, based on CE and ICE algorithms, estimate the same optimal reference parameter, but they use different estimators. According to the numerical study in

Chapter 6, they all give similar variance reduction rate in all examples considered there. The choice of the particular method is a matter of taste.

4.2 Importance Sampling with Stratification based algorithms

In this section we apply stratification to the algorithms from Section 4.1. As the numerical study in Chapter 6 shows, it gives significant extra variance reduction compared to the methods based only on the importance sampling. Since all importance sampling algorithms introduced in Section 4.1 show approximately the same performance, as mentioned there, the same stratification leads to the same improvement of all these methods.

In order to perform stratification, a stratified variable with known distribution should be introduced (see Section 2.2). Glasserman (2004), Glasserman et al. (2000a,b, 2002) use the Delta-Gamma approximation of the changes in risk factors as such a variable, whose distribution is found in two cases: of Gaussian and of Student t-distributed changes in risk factors (see Section 3.3). An algorithm based on the exponential twisting together with such a stratification (for both cases), called Importance Sampling with Stratification (ISS) algorithm, is given by Glasserman (2004), Glasserman et al. (2000a,b, 2002). This algorithm is presented in Section 4.2.1 for the Gaussian case and in Section 4.2.2 for the Student t-case.

Since, the CE and ICE based algorithms from Sections 4.1.3 and 4.1.4 are introduced here to build a flexible approach of measuring the probability of high portfolio losses under any model of changes in risk factors, stratification is not of particular interest for them. A special stratification, that does not involve a knowledge of the distribution of the stratified variable, is introduced instead in Section 4.3.1. In Section 4.2.3 we, however, apply stratification to the CE and ICE based algorithms. But, since applying the stratification leads potentially to the same performance as for the ISS algorithm, we do not perform the numerical study for these algorithms.

4.2.1 Exponential twisting with original Stratification: Gaussian case

In this section we present the ISS algorithm in the case of Gaussian changes in risk factors from Glasserman (2004), Glasserman et al. (2000a,b). It is based on the IS algorithm combined with stratification. The Delta-Gamma approximation $Q = Q(\mathbf{Z})$, introduced in Section 3.3.1, is used as a stratified variable.

In such a case, stratification consists of the following two steps (see Section 2.2):

- for some chosen number of strata K, the strata $A_1, \ldots, A_K$ and the allocation $N_1, \ldots, N_K$ have to be defined,
- the mechanism of sampling from the conditional distribution of $(Q(\mathbf{Z}), L(\mathbf{CZ}))$ given $Q(\mathbf{Z}) \in A_i$ for $i = 1, \ldots, K$ should be found.

Before considering these two steps in more details, we have to adjust the stratified estimator (2.23), which is simply a sample counterpart of the relation (2.22). Applying stratification to the IS algorithm is, therefore, achieved by substituting $\mathbb{E}$ and $H(\mathbf{X})$ in (2.22) by $\mathbb{E}_\theta$ and $e^{-\theta Q(\mathbf{Z})+\psi(\theta)} I_{(x,+\infty)}(L(\mathbf{CZ}))$ respectively, more precisely:

$$\begin{aligned} l &= \mathbb{E}_\theta \left[e^{-\theta Q(\mathbf{Z})+\psi(\theta)} I_{(x,+\infty)}(L(\mathbf{CZ})) \right] = \\ &= \sum_{i=1}^{K} p_i \mathbb{E}_\theta \left[e^{-\theta Q(\mathbf{Z})+\psi(\theta)} I_{(x,+\infty)}(L(\mathbf{CZ})) | Q(\mathbf{Z}) \in A_i \right], \end{aligned}$$

where $p_i = \mathbb{P}_\theta(Q(\mathbf{Z}) \in A_i)$ have to be known by Assumption 2.11.

The latter requires the knowledge of the distribution of the Delta-Gamma approximation $Q = Q(\mathbf{Z})$ under $\mathbb{P}_\theta$, i.e. when $\mathbf{Z}$ has the density $f_\theta(\mathbf{x}; \mathbf{u})$ (or equivalently $\mathbf{Z} \sim N_d(\mu(\theta), \Sigma(\theta))$, see Section 4.1.1). The parameter θ is an exponential twisting parameter achieved in step 2 of Algorithm 4.1.

Let us remind, that the distribution of Q in case of Gaussian changes in risk factors is given by the inversion integral (3.8) with the help of the cumulant generating function ψ of the form (3.9). In Glasserman (2004), Glasserman et al. (2000a,b) the cumulant generating function of Q, under exponential change of measure defined by the twisting parameter θ, is given by:

$$\psi_\theta(u) = \log \mathbb{E}_\theta[e^{uQ(\mathbf{Z})}] = \log \mathbb{E}[e^{uQ(\mathbf{Z})} e^{\theta Q(\mathbf{Z})-\psi(\theta)}] = \psi(\theta+u) - \psi(\theta). \tag{4.14}$$

The analogue to (3.8) looks then as:

$$\mathbb{P}_\theta(Q \le x) - \mathbb{P}_\theta(Q \le x-y) = \frac{1}{\pi} \int_0^\infty Re \left(\hat{\phi}_\theta(u) \left[\frac{e^{iuy}-1}{iu} \right] e^{-iux} \right) du, \tag{4.15}$$

with $\hat{\phi}_\theta(u) = e^{\psi(\theta+iu)-\psi(\theta)}$. To find $\mathbb{P}_\theta(Q \le x)$ one should choose y large enough so that $\mathbb{P}_\theta(Q \le x-y) \approx 0$, see Glasserman (2004), Glasserman et al. (2000a,b).

Now, we can return to the two steps in performing the stratification. The number of strata K is taken as an input parameter of the ISS algorithm. The strata $A_1, \ldots, A_K$

are chosen by Glasserman (2004), Glasserman et al. (2000a,b) to be a partition of the real line of equal probabilities for $Q = Q(\mathbf{Z})$ under $\mathbb{P}_\theta$, more precisely they are given by the points a_i, such that $A_i = (a_{i-1}, a_i)$ and the following holds true:

$$\mathbb{P}_\theta(Q(\mathbf{Z}) \in A_1) = \ldots = \mathbb{P}_\theta(Q(\mathbf{Z}) \in A_K) = \frac{1}{K}.$$

Such points a_i for $i = 0, \ldots, K$ are then defined as:

$$a_0 = -\infty, \quad \mathbb{P}_\theta(Q(\mathbf{Z}) \leq a_i) = \frac{i}{K} \text{ for } i = 1, \ldots, K-1, \quad a_K = +\infty. \tag{4.16}$$

An arbitrary choice of the allocation finalizes the first step. It can not be used as an input parameter, since e.g. proportional allocation (see Section 2.2) requires the knowledge of the strata, which are defined along the algorithm.

In the second step, it is easier to sample $\mathbf{Z}$ given $Q(\mathbf{Z}) \in A_i$ and evaluate $(Q(\mathbf{Z}), L(\mathbf{CZ}))$ rather than sample $(Q(\mathbf{Z}), L(\mathbf{CZ}))$ given $Q(\mathbf{Z}) \in A_i$ directly. The procedure of sampling $\mathbf{Z}$ given $Q(\mathbf{Z}) \in A_i$ is, therefore, necessary. In Section 2.2 we referred the reader to Appendix A on the methods for sampling from the conditional densities. One of those methods has a particular value in the ISS algorithm. The algorithm by Glasserman (2004), Glasserman et al. (2000a,b) uses the brute-force acceptance-rejection method along the lines (see Appendix A.1). Its main disadvantage of the low acceptance rate (and therefore wasting time on simulating unnecessary values) is not that crucial here, because we need conditional samples of $\mathbf{Z}$ given that $Q(\mathbf{Z})$ falls in all strata, which in the union give the whole real line. This means that, if the simulated value of $\mathbf{Z}$ given $Q(\mathbf{Z}) \in A_i$ is rejected, then it is accepted as $\mathbf{Z}$ given $Q(\mathbf{Z}) \in A_j$ for some other $j \neq i$. The algorithm starts to waste the samples only when the corresponding sample of $\mathbf{Z}$ given $Q(\mathbf{Z}) \in A_j$ is full according to the defined allocation. But the total number of wasted samples is not significant. This procedure in general case is described via the modification of Algorithm A.1, namely the Algorithm A.2.

All said above summarizes in the following ISS algorithm for the case of Gaussian distributed changes in risk factors, i.e. under Assumption 3.1:

Algorithm 4.7 (ISS: Gaussian case)**.**

1. *For the changes in risk factors $\Delta\mathbf{S} \sim N_d(\mathbf{0}, \mathbf{\Sigma}_S)$ calculate the decomposition of the covariance matrix $\mathbf{\Sigma}_S = \mathbf{C}\mathbf{C}^T$, such that the Delta-Gamma approximation takes the form:*

$$Q(\mathbf{Z}) = a + \sum_{i=1}^{d} (\mathbf{b}_i \mathbf{Z}_i + \lambda_i \mathbf{Z}_i^2),$$

for $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$ (see Section 3.3.1). Fix input parameters N and K.

2. *Denote by $\theta > 0$ the positive solution to the equation:*

$$\psi'(\theta) = x,$$

such that $\max_i \theta\lambda_i < \frac{1}{2}$, where $\psi(\theta)$ is the cumulant generating function of $Q(\mathbf{Z})$ and is given by (3.9).

3. *Define the strata $A_i = (a_{i-1}, a_i)$, $i = 1, \ldots, K$ via (4.16). Fix the allocation $N_1, \ldots, N_K$, such that $N = N_1 + \ldots + N_K$.*

4. *Estimate the probability l using the stratified estimator (2.23) in the form*

$$\hat{l} = \sum_{i=1}^{K} \frac{1}{K \cdot N_i} \sum_{j=1}^{N_i} I_{(x,\infty)}(L(\mathbf{C}\mathbf{Z}_{ij}))e^{-\theta Q(\mathbf{Z}_{ij})+\psi(\theta)}, \tag{4.17}$$

where $\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{iN_i}$ for $i = 1, \ldots, K$ are the samples generated from the conditional densities $\mathbf{Z} \sim N_d(\mu(\theta), \Sigma(\theta))$ given $Q(\mathbf{Z}) \in A_i$ (according to the brute-force acceptance-rejection method along the lines (see Algorithm A.2)). The mean vector and the diagonal covariance matrix are given by (4.2).

The brute-force acceptance-rejection method along the lines (see Algorithm A.2) simulates $\mathbf{Z}$ from the unconditional density $N_d(\mu(\theta), \Sigma(\theta))$ and evaluates $Q(\mathbf{Z})$. The vector $\mathbf{Z}$ is then assigned to the stratum A_i such that $Q(\mathbf{Z}) \in A_i$. It is done until all strata are full according to the allocation $N_1, \ldots, N_K$. If no more samples are necessary for a particular stratum, $\mathbf{Z}$ is simply discarded.

As we already mentioned before, the last step of the algorithm potentially generates much more values $\mathbf{Z}_{ij}$ than the given sample size N for the estimator $\hat{l}$. But, as mentioned in Glasserman (2004), executing these extra steps takes far less time than the portfolio revaluation step. The corresponding analysis is given in Glasserman et al. (2000b).

4.2.2 Exponential twisting with original Stratification: Student t-case

In this section we present the ISS algorithm in the case of Student t-distributed changes in risk factors from Glasserman (2004), Glasserman et al. (2002). It is based on the IS algorithm combined with stratification. The procedure of applying the stratification is very similar to Section 4.2.1. The only difference is that the indirect Delta-Gamma approximation $Q_x = Q_x(\mathbf{Z}, Y)$, introduced in Section 3.3.2, is used as a stratified variable. The distribution of the indirect Delta-Gamma approximation under the exponential change of measure is, therefore, required. This is derived in Glasserman (2004),

Glasserman et al. (2002) similarly to the Gaussian case, i.e. via cumulant generating functions (see (4.14)).

More precisely, the cumulant generating function of the indirect Delta-Gamma approximation $Q_x(\mathbf{Z}, Y)$ under the exponential change of measure is given by $u \longmapsto \psi_x(\theta + u) - \psi_x(\theta)$, where $\psi_x(\theta) = \log \phi_x(\theta)$ is the cumulant generating function of $Q_x(\mathbf{Z}, Y)$ under the original measure and is given via (3.10). The analogue to (4.15) can then easily be written for Q_x, giving the formula for calculating the distribution of the indirect Delta-Gamma approximation. This can further be used in order define the strata in an analogous way to (4.16). More precisely, the strata $A_i = (a_{i-1}, a_i)$, $i = 1, \ldots, K$ are given by the points a_i, such that:

$$a_0 = -\infty, \quad \mathbb{P}_\theta(Q_x(\mathbf{Z}, Y) \leq a_i) = \frac{i}{K} \text{ for } i = 1, \ldots, K-1, \quad a_K = +\infty, \tag{4.18}$$

where $\mathbb{P}_\theta$ denotes the probability measure under the exponential change of measure.

All said above summarizes in the following ISS algorithm for the case of Student t-distributed changes in risk factors, i.e. under the Assumption 3.2:

Algorithm 4.8 (ISS: Student t-case)**.**

1. *For the changes in risk factors $\Delta\mathbf{S} \sim t_d(\nu, \mathbf{0}, \mathbf{\Sigma}_S)$ calculate the decomposition of the covariance matrix $\mathbf{\Sigma}_S = \mathbf{C}\mathbf{C}^T$, such that the indirect Delta-Gamma approximation takes the form:*

$$Q_x(\mathbf{Z}, Y) = (a - x)(Y/\nu) + \sum_{i=1}^{d} \left(\mathbf{b}_i \sqrt{Y/\nu} \mathbf{Z}_i + \lambda_i \mathbf{Z}_i^2 \right),$$

   for $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$ and $Y \sim \chi^2_\nu$, that is independent of the vector $\mathbf{Z}$ (see Section 3.3.2). Fix input parameter N and K.

2. *Denote by $\theta > 0$ the positive solution to the equation:*

$$\psi'_x(\theta) = 0,$$

   such that $\max_i \theta\lambda_i < \frac{1}{2}$ and $a(\theta) < 0.5$, where $\psi_x(\theta) = \log \phi_x(\theta)$ is the cumulant generating function of $Q_x(\mathbf{Z}, Y)$ and is given via (3.10) with $a(\theta)$ defined in the Section 3.3.2.

3. *Define the strata $A_i = (a_{i-1}, a_i)$, $i = 1, \ldots, K$ via (4.18). Fix the allocation $N_1, \ldots, N_K$, such that $N = N_1 + \ldots + N_K$.*

4. Estimate the probability l using the stratified estimator (2.23) in the form

$$\hat{l} = \sum_{i=1}^{K} \frac{1}{K \cdot N_i} \sum_{j=1}^{N_i} I_{(x,\infty)} \left(L \left(\frac{\mathbf{C}\mathbf{Z}_{ij}}{\sqrt{Y_{ij}/\nu}} \right) \right) e^{-\theta Q_x(\mathbf{Z}_{ij}, Y_{ij}) + \psi_x(\theta)}, \tag{4.19}$$

where given the sample $Y_{i1}, ..., Y_{iN_i}$ for $i = 1, \ldots, K$ generated from the density $\Gamma\left(\frac{\nu}{2}, \frac{2}{1-2a(\theta)}\right)$, the vectors $\mathbf{Z}_{ij}$ are generated from the conditional densities $\mathbf{Z} \sim N_d(\boldsymbol{\mu}(\theta, Y_{ij}), \boldsymbol{\Sigma}(\theta))$ for $j = 1, \ldots, N_i$ given $Q_x(\mathbf{Z}, Y_{ij}) \in A_i$ (according to the brute-force acceptance-rejection method along the lines (see Algorithm A.2)). The mean vector and the diagonal covariance matrix are given by (4.4).

According to the brute-force acceptance-rejection method along the lines (see Algorithm A.2), the last step in Algorithm 4.8 may be explained easier as follows. We first generate Y from the density $\Gamma\left(\frac{\nu}{2}, \frac{2}{1-2a(\theta)}\right)$. Then, given this value, we simulate $\mathbf{Z}$ from the density $N_d(\boldsymbol{\mu}(\theta, Y), \boldsymbol{\Sigma}(\theta))$. For these value of Y and of the vector $\mathbf{Z}$ we find i, such that $Q_x(\mathbf{Z}, Y) \in A_i$. Then we denote $Y_{i1} = Y$ and $\mathbf{Z}_{i1} = \mathbf{Z}$. We continue sampling and denote the next value Y and vector $\mathbf{Z}$ such that $Q_x(\mathbf{Z}, Y) \in A_i$ by consequent numbers, e.g. $Y_{i2} = Y$ and $\mathbf{Z}_{i2} = \mathbf{Z}$. We perform such a simulation until we fill all the strata, i.e. we get the samples of values $Y_{i1}, ..., Y_{iN_i}$ and vectors $\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{iN_i}$ for $i = 1, \ldots, K$. It is important to understand, that each simulated vector $\mathbf{Z}_{ij}$ depends on the previously simulated value Y_{ij}, and they are accepted to the stratum only together. It can be also understood as applying the brute-force acceptance-rejection method to the vector $\mathbf{Z}$ extended by the value Y.

4.2.3 Original and improved Cross - Entropy with original Stratification

In this section we show how to apply stratification to the ACE, FCE, AICE and FICE algorithms introduced in Sections 4.1.3 and 4.1.4. Similarly to Sections 4.2.1 and 4.2.2, the approximation $Q(\Delta\mathbf{S})$ plays the role of the stratified variable. It can differ from the approximation function used in ACE, FCE, AICE and FICE algorithms in order to estimate the optimal reference parameter (as you might remember, in the FCE and FICE algorithms the approximation function is the original loss function, which, obviously, can not be used as a stratified variable). The knowledge of the distribution of the stratified variable restricts the choice of the approximation and the model of the changes in risk factors. In the cases of Gaussian and Student t-distributed changes in risk factors, the distribution of the Delta-Gamma approximation is found explicitly (see Sections 3.3.1 and 3.3.2). In these two cases the original stratification is applied to the IS algorithm

(see Sections 4.2.1 and 4.2.2). In a similar way, in both cases, it can be applied to the ACE, FCE, AICE and FICE algorithms here.

Due to the interest in developing a flexible approach of estimating the probability of high portfolio losses under any model of the changes in risk factors, applying the original stratification to the ACE, FCE, AICE and FICE algorithms is of little interest. Therefore, we only briefly describe here how to apply an extra stratification to these algorithms. The analogues to the ISS algorithms are introduced in Section 4.3 and are based on the special stratification introduced in Section 4.3.1.

ACE, FCE, AICE and FICE first estimate the optimal reference parameter $\hat{\mathbf{v}}^*$, and then, in the last step, use the corresponding change of measure to estimate the probability of interest with the help of the importance sampling estimator. In order to additionally apply stratification, the steps of the optimal reference parameter estimation stay unchanged. Therefore, we assume, that the estimate $\hat{\mathbf{v}}^*$ is found.

The stratification consists then of the following two steps (for more details see Sections 2.2 and 4.2.1):

- In order to define the strata $A_1, \ldots, A_K$, the distribution of the approximation $Q(\Delta\mathbf{S})$ under the density $f(\mathbf{x}; \hat{\mathbf{v}}^*)$ should be found
- the mechanism of sampling from the conditional distribution of $(Q(\Delta\mathbf{S}), L(\Delta\mathbf{S}))$ given $Q(\Delta\mathbf{S}) \in A_i$ for $i = 1, \ldots, K$ should be found.

The second step is very similar to the situation of ISS algorithm and is discussed in details in Sections 4.2.1 and 4.2.2. The first step should be done extra. We only note that the distribution of the approximation $Q(\Delta\mathbf{S})$ under the density $f(\mathbf{x}; \hat{\mathbf{v}}^*)$ can be found in a similar way as for the ISS algorithm.

After strata $A_i = (a_{i-1}, a_i)$, $i = 1, \ldots, K$ are defined in a similar way to Sections 4.2.1 and 4.2.2, the corresponding estimator based on CE or ICE importance sampling, together with stratification, is defined as:

$$\hat{l} = \sum_{i=1}^{K} \frac{1}{K \cdot N_i} \sum_{j=1}^{N_i} I_{(x,\infty)}(L(\Delta\mathbf{S}_{ij}))W(\Delta\mathbf{S}_{ij}; \mathbf{u}, \hat{\mathbf{v}}^*),$$

where $\Delta\mathbf{S}_{i1}, ..., \Delta\mathbf{S}_{iN_i}$ for $i = 1, \ldots, K$ are the samples generated from the conditional density $f(\mathbf{x}; \hat{\mathbf{v}}^*)$ given $Q(\Delta\mathbf{S}) \in A_i$ (according to the brute-force acceptance-rejection method along the lines (see Algorithm A.2)).

4.3 Importance Sampling with special Stratification based algorithms

As already mentioned in Section 4.2, the stratification applied there should be modified in order to bring flexibility to our new algorithms. Of course, it should be done in a way, as to achieve as higher variance reduction rate as possible. The corresponding modification to the stratification algorithm, here called special stratification, is done in Section 4.3.1 and is further applied to the ACE and AICE algorithms in Sections 4.3.2 and 4.3.3, respectively. We should note, that there is no interest in applying the special stratification to IS algorithm, since this algorithm is developed only in two cases of changes in risk factors (see Assumptions 3.1 and 3.2), where the original stratification is also developed. The special stratification can easily be applied to FCE and FICE algorithms in a similar way, but is also not of a particular interest here due to their time consuming steps for the estimation of the reference parameter.

4.3.1 Special Stratification

In this section we develop the special stratification in the case of estimating the probability of high portfolio losses (4.1). The corresponding idea is first applied in the paper of mine Pupashenko (2014). It is further improved in my joint paper with R. Korn, Korn and Pupashenko (2014). Here we give a more detailed discussion on the development of this procedure.

Under Assumption 4.1, the approximation $Q(\Delta\mathbf{S})$ is a good candidate to be a stratified variable (that was the choice in the previous sections). However, as it might be hard or even impossible to derive the distribution of the approximation $Q(\Delta\mathbf{S})$ for any distribution of the changes in risk factors $\Delta\mathbf{S}$, Assumption 2.11, which is crucial in developing the stratified estimator (see (2.23)), may be violated. The main problem lies here in the knowledge of the probabilities of Q falling in a particular stratum. In order to build a flexible algorithm, the special stratification is based, therefore, on the MC estimation of these probabilities.

In the general case of the stratification for the estimation of the mean (2.1), the corresponding stratified estimator (2.23) under the notations from Chapter 2 takes the form:

$$\hat{l} = \sum_{i=1}^{K} \hat{p}_i \frac{1}{N_i} \sum_{j=1}^{N_i} H(\mathbf{X}_{ij}), \tag{4.20}$$

where $\hat{p}_i$ are MC estimators of the probabilities $p_i = \mathbb{P}(\mathbf{Y} \in A_i)$.

First, we should mention, that the variance of the estimator (4.20) depends on the variances of the estimators $\hat{p}_i$ and is larger than the variance of the estimator (2.23). In order to keep the variance of (4.20) as close as possible to (2.23) (therefore achieve the same variance reduction rate), the sample size used to estimate all probabilities $\hat{p}_i$ should be set as high as possible. It is, however, restricted by the extra time costs, which have to be kept as small as possible. In general, according to Assumption 4.1, it is possible to use large sample sizes in estimating $\hat{p}_i$. One should, however, keep in mind, that the total time costs are also influenced by the number of probabilities to estimate K (the number of strata).

In order to make the problem more clear, let us consider a simple example. Assume, that the stratified estimator (4.20) is achieved by using a total sample size of losses $N = N_1 + \ldots + N_K$, where K is the number of strata. Assume also that, according to Assumption 4.1, the approximation Q can be calculated 100 times faster than the loss function L. Then, if we want, for example, to spend twice more time for the estimator (4.20) compared to (2.23), we approximately have the total sample size $100 \cdot N$ of revaluations of the approximation Q in estimating the probabilities $\hat{p}_i$. If we have decided to use $K = 50$ strata, this means, that each MC estimator of the probabilities $\hat{p}_i$ is based on the sample of the size $2 \cdot N$. This might even lead to the variance of each separate probability $\hat{p}_i$ being larger than the variance of the estimator (2.23), making the method useless in practice. This simple example motivates us to use as few strata as possible, which in turn reduces the variance reduction rate (compared to large number of strata), unless the strata and their allocation are chosen in a special way.

Our investigation in Korn and Pupashenko (2014), Pupashenko (2014) shows, that the best compromise between the speed and the performance of the special stratification in the case of estimating the probability of high portfolio losses is achieved by splitting the real line into three intervals in a special way. More precisely, for the loss function L, its approximation Q and the level x, we define two values:

$$y = y(x) = \inf_{\mathbf{x} \in \mathbb{R}_{\Delta \mathbf{S}}} \{Q(\mathbf{x}) | L(\mathbf{x}) > x\}, \quad z = z(x) = \sup_{\mathbf{x} \in \mathbb{R}_{\Delta \mathbf{S}}} \{Q(\mathbf{x}) | L(\mathbf{x}) < x\}, \tag{4.21}$$

where $\mathbb{R}_{\Delta \mathbf{S}}$ is the support of the distribution of the changes in risk factors $\Delta \mathbf{S}$. In general, both values y and z depend on the level x which is, however, fixed in the context of the probability of high portfolio losses estimation (4.1). Therefore, all over Chapter 4 they are also fixed for each loss function L and its approximation Q.

The following remark holds for the rest of the section:

Remark 4.1. *Since we deal with the case of continuous distributions (see Assumption 2.1), it is not important whether we use strict inequalities or not.*

The value y is chosen in such a way, that the following implication holds true (by the definition of inf):

$$L(\mathbf{x}) > x \Rightarrow Q(\mathbf{x}) > y, \text{ for all } \mathbf{x} \in \mathbb{R}_{\Delta\mathbf{S}},$$

which is equivalent to:

$$Q(\mathbf{x}) \leq y \Rightarrow L(\mathbf{x}) \leq x, \text{ for all } \mathbf{x} \in \mathbb{R}_{\Delta\mathbf{S}}. \tag{4.22}$$

In other words this means that, whenever the approximation falls below the level y, the loss falls below the level x. This means, that the bottom right quadrant between the vertical line at x and the horizontal line at y does not contain any points $(L(\Delta\mathbf{S}), Q(\Delta\mathbf{S}))$, see Figure 4.1.

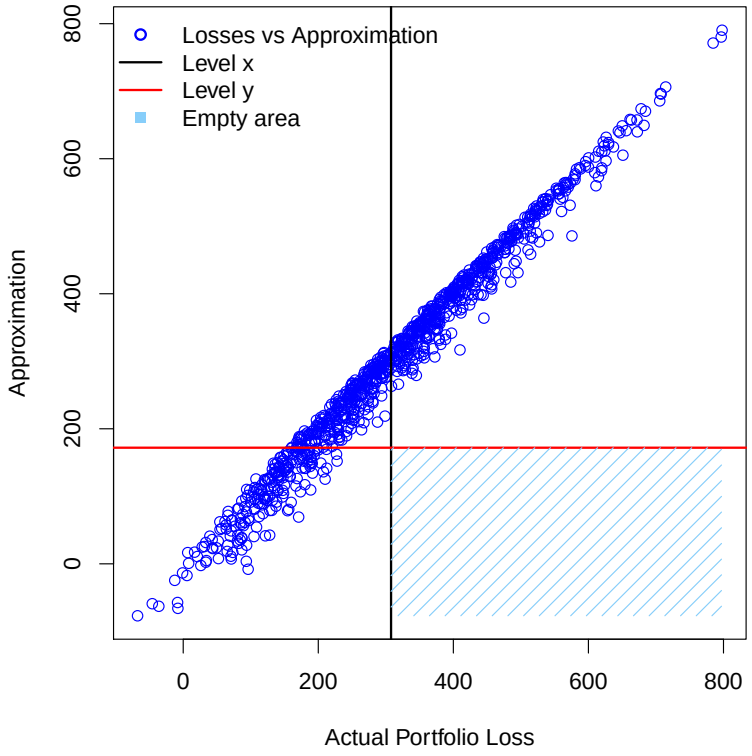


FIGURE 4.1: Level y for Delta-Gamma approximation against actual portfolio losses in 1000 randomly generated scenarios for a portfolio of short 10 ATM down-and-out calls, 10 uncorrelated assets, 0.1 years maturity, Gaussian changes in risk factors, probability nearly 1% (for details on the portfolio see Chapter 6).

Similarly, the value z is chosen in such a way, that the following implication holds true (by the definition of sup):

$$L(\mathbf{x}) < x \Rightarrow Q(\mathbf{x}) < z, \text{ for all } \mathbf{x} \in \mathbb{R}_{\Delta\mathbf{S}},$$

which is equivalent to:

$$Q(\mathbf{x}) \geq z \Rightarrow L(\mathbf{x}) \geq x, \text{ for all } \mathbf{x} \in \mathbb{R}_{\Delta\mathbf{S}}. \tag{4.23}$$

In other words this means that, whenever the approximation falls above the level z, the loss falls above the level x. This means, that the top left quadrant between the vertical line at x and the horizontal line at z does not contain any points $(L(\Delta\mathbf{S}), Q(\Delta\mathbf{S}))$, see Figure 4.2.

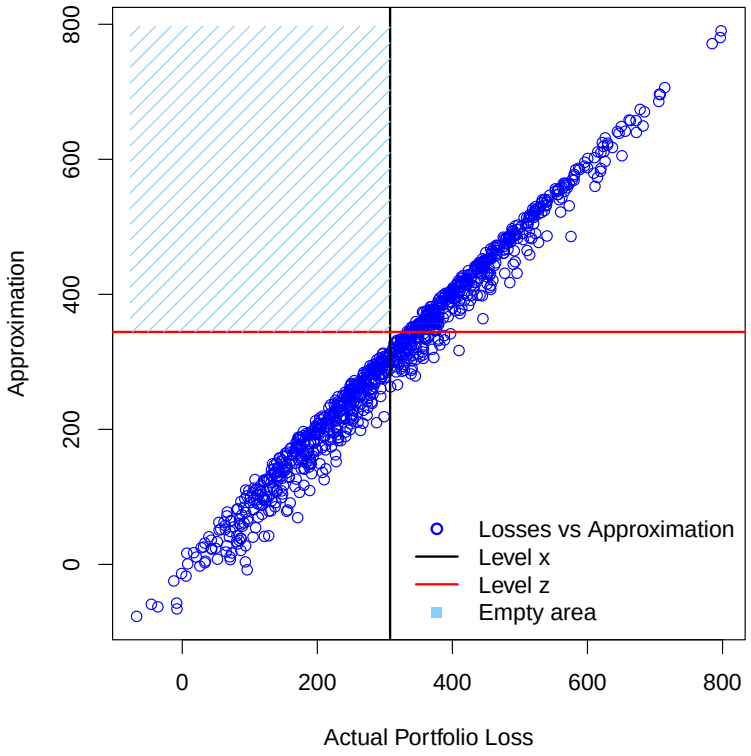


FIGURE 4.2: Level z for Delta-Gamma approximation against actual portfolio losses in 1000 randomly generated scenarios for a portfolio of short 10 ATM down-and-out calls, 10 uncorrelated assets, 0.1 years maturity, Gaussian changes in risk factors, probability nearly 1% (for details on the portfolio see Chapter 6).

The strata is then defined as a partition into three intervals:

$$A_1 = (-\infty, y], \quad A_2 = (y, z), \quad A_3 = [z, +\infty). \tag{4.24}$$

In order to get the maximum performance out of the stratification with the strata given by (4.24), we have to define an optimal allocation. Summarizing the above said, the

value of the indicator $I_{(x,\infty)}(L(\Delta\mathbf{S}))$ is known, if $Q(\Delta\mathbf{S})) \notin (y,z)$. More precisely:

$$I_{(x,\infty)}(L(\Delta\mathbf{S})) = \begin{cases} 0, & \text{if } Q(\Delta\mathbf{S})) \leq y, \\ 1, & \text{if } Q(\Delta\mathbf{S})) \geq z. \end{cases}$$

This leads to the zero variance of the estimators in the strata $A_1 = (-\infty, y]$ and $A_3 = [z, +\infty)$. Now it is logical not to use any sample for strata A_1 and A_3, since the corresponding estimators are actually known constants. Therefore, the optimal allocation for the strata given by (4.24) is defined in such a way, that the whole sample of size N is allocated to the stratum $A_2 = (y,z)$, i.e. $N_1 = 0, N_2 = N, N_3 = 0$. The corresponding interval of interest (y,z) is given in Figure 4.3.

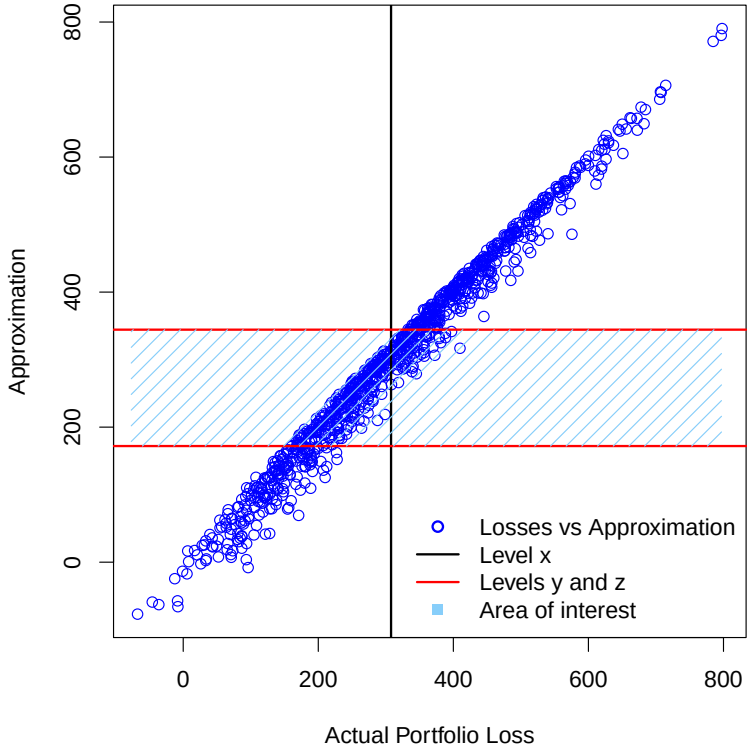


FIGURE 4.3: Levels y and z for Delta-Gamma approximation against actual portfolio losses in 1000 randomly generated scenarios for a portfolio of short 10 ATM down-and-out calls, 10 uncorrelated assets, 0.1 years maturity, Gaussian changes in risk factors, probability nearly 1% (for details on the portfolio see Chapter 6).

Remark 4.2. *Note, that the stratification does not have to be defined via exact values of y and z. A wider region between $y_1 \leq y$ and $z_1 \geq z$ can be used. This gives the advantage of not solving the optimizations (4.21) precisely. Of course, as narrower the interval (y,z), that better the performance of the special stratification. But the influence is not crucial in the neighborhoods of the original levels.*

The stratified estimator in the context of estimating the probability of high portfolio losses in the ideal situation of known distribution of the approximation Q takes the form:

$$\hat{l} = \mathbb{P}(Q(\Delta\mathbf{S}) \in (y,z))\frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i)) + \mathbb{P}(Q(\Delta\mathbf{S}) \geq z), \tag{4.25}$$

where $(\Delta\mathbf{S}_1, \ldots, \Delta\mathbf{S}_N)$ is a sample from the conditional density $f(\mathbf{x};\mathbf{u})$ of $\Delta\mathbf{S}$, given that $Q(\Delta\mathbf{S}) \in (y,z)$, which can be achieved by the methods given in Appendix A.

The corresponding special stratification, in a more general case, when the distribution of the approximation Q is not given explicitly, is achieved from (4.25) by substituting the probabilities $\mathbb{P}(Q(\Delta\mathbf{S}) \in (y,z))$ and $\mathbb{P}(Q(\Delta\mathbf{S}) \geq z)$ with their MC estimators:

$$\hat{l} = \hat{\mathbb{P}}(Q(\Delta\mathbf{S}) \in (y,z))\frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i)) + \hat{\mathbb{P}}(Q(\Delta\mathbf{S}) \geq z), \tag{4.26}$$

where $(\Delta\mathbf{S}_1, \ldots, \Delta\mathbf{S}_N)$ is a sample from the conditional density $f(\mathbf{x};\mathbf{u})$ of $\Delta\mathbf{S}$, given that $Q(\Delta\mathbf{S}) \in (y,z)$, which can be achieved by the methods given in Appendix A.

As one can see, the special stratification (4.26) involves an extra estimation of two probabilities $\mathbb{P}(Q(\Delta\mathbf{S}) \in (y,z))$ and $\mathbb{P}(Q(\Delta\mathbf{S}) \geq z)$, or equivalently $\mathbb{P}(Q(\Delta\mathbf{S}) > y)$ and $\mathbb{P}(Q(\Delta\mathbf{S}) \geq z)$, that should be done independently. As mentioned above, one should control the time necessary for this extra estimation by using sample sizes for the corresponding MC estimators as small as possible, however, keeping them as large as possible in order to achieve the variance of the estimator (4.26) as close as possible to the variance of the estimator (4.25).

Remark 4.3. *In order to improve the speed of the estimator (4.26) one can think of improving the speed of the estimators $\hat{\mathbb{P}}(Q(\Delta\mathbf{S}) > y)$ and $\hat{\mathbb{P}}(Q(\Delta\mathbf{S}) \geq z)$ by applying some variance reduction procedures to them. Although, assuming that the level y is large (it depends on the level x and the approximation Q and is large, whenever x is large and the approximation Q is good enough), these probabilities are similar to the probability of high portfolio losses, one should keep in mind, that they do not involve the loss function and, therefore, do not face similar problems of the heavy set up of the importance sampling estimator. We suggest to use flexible importance sampling tool, namely the CE algorithm. It can be applied directly in the form of the FCE algorithm (see Algorithm 4.4) by substituting the function L by Q in the whole algorithm. One should also note, that there is no gain in applying any stratification, since, generally, there is no good candidate of the stratified variable for the approximation $Q(\Delta\mathbf{S})$.*

We now make the procedure of controlling the variance of the estimator (4.26) more rigorous. We assume, that the sample size N (number of the total revaluations of

the loss function L) is fixed in both estimators (4.25) and (4.26). By $\hat{p}_1(m_1 \cdot N)$ and $\hat{p}_2(m_2 \cdot N)$ we denote the estimators $\hat{\mathbb{P}}(Q(\Delta\mathbf{S}) \in (y,z))$ and $\hat{\mathbb{P}}(Q(\Delta\mathbf{S}) \geq z)$, achieved by some MC or variance reduction algorithm with sample sizes $m_1 \cdot N$ and $m_2 \cdot N$, respectively. The multipliers m_1 and m_2 control the variance of the corresponding estimators (and therefore of the estimator (4.26)) and the extra time necessary to achieve estimator (4.26) in comparison to the estimator (4.25). By $\hat{l}_0$ we denote the common term $\frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i))$, which does not depend on m_1 or m_2. The estimator (4.26) then takes the form:

$$\hat{l} = \hat{p}_1(m_1 \cdot N)\hat{l}_0 + \hat{p}_2(m_2 \cdot N).$$

We use the variance reduction factor r (ratio of the variance of the MC method to the corresponding variance of the improved method, see Chapter 6) of the estimator (4.25) as a reference. We assume, that by applying the estimator (4.26) we want to achieve the variance reduction factor of approximately $\delta \cdot r$ for some $\delta \in (0,1)$, which should be close to 1. Under this assumption, the variance of the estimator (4.26), denoted here by $\mathbb{V}\mathbf{ar}[\hat{l}]$, should be equal to:

$$\mathbb{V}\mathbf{ar}[\hat{l}] = \frac{\mathbb{P}^2(Q(\Delta\mathbf{S}) \in (y,z))\mathbb{V}\mathbf{ar}[\hat{l}_0]}{\delta},$$

where the nominator is a variance of the estimator (4.25). This variance can be written as (due to the unbiasedness of the MC estimator $\hat{p}_1(N)$):

$$\mathbb{V}\mathbf{ar}[\hat{l}] = \frac{\mathbb{E}[\hat{p}_1^2(N)]\mathbb{V}\mathbf{ar}[\hat{l}_0]}{\delta}.$$

In order to achieve such variance, we find necessary multipliers m_1 and m_2. The variance $\mathbb{V}\mathbf{ar}[\hat{l}]$ can be written as (due to the mutual independence of $\hat{p}_1(m_1 \cdot N)$, $\hat{p}_1(m_2 \cdot N)$ and $\hat{l}_0$):

$$\begin{aligned}
\mathbb{V}\mathbf{ar}[\hat{l}] &= \mathbb{E}[\hat{p}_1^2(m_1 \cdot N)]\mathbb{V}\mathbf{ar}[\hat{l}_0] + \mathbb{V}\mathbf{ar}[\hat{p}_1(m_1 \cdot N)](\mathbb{E}[\hat{l}_0])^2 + \\
&+ \mathbb{V}\mathbf{ar}[\hat{p}_1(m_1 \cdot N)]\mathbb{V}\mathbf{ar}[\hat{l}_0] + \mathbb{V}\mathbf{ar}[\hat{p}_2(m_2 \cdot N)] \approx \\
&\approx \mathbb{E}[\hat{p}_1^2(N)]\mathbb{V}\mathbf{ar}[\hat{l}_0] + \mathbb{V}\mathbf{ar}[\hat{p}_1(m_1 \cdot N)](\mathbb{E}[\hat{l}_0])^2 + \mathbb{V}\mathbf{ar}[\hat{p}_2(m_2 \cdot N)],
\end{aligned}$$

where the third summand in the equality can be neglected, and the first summand is the variance of the estimator (4.25). In order to follow the above assumption on the variance reduction factor, the following condition should, therefore, be satisfied:

$$\mathbb{E}[\hat{p}_1^2(N)]\mathbb{V}\mathbf{ar}[\hat{l}_0]\frac{1-\delta}{\delta} = \mathbb{V}\mathbf{ar}[\hat{p}_1(m_1 \cdot N)](\mathbb{E}[\hat{l}_0])^2 + \mathbb{V}\mathbf{ar}[\hat{p}_2(m_2 \cdot N)].$$

Hence, we choose multipliers m_1 and m_2 such, that the following holds true:

$$\mathbb{E}[\hat{p}_1^2(N)]\mathbb{V}\mathbf{ar}[\hat{l}_0]\frac{1-\delta}{2\delta} = \mathbb{V}\mathbf{ar}[\hat{p}_1(m_1 \cdot N)](\mathbb{E}[\hat{l}_0])^2,$$
$$\mathbb{E}[\hat{p}_1^2(N)]\mathbb{V}\mathbf{ar}[\hat{l}_0]\frac{1-\delta}{2\delta} = \mathbb{V}\mathbf{ar}[\hat{p}_2(m_2 \cdot N)].$$

Note, that for MC or importance sampling method (e.g. CE), the variances of the probability estimators $\hat{p}_1(m_1 \cdot N)$ and $\hat{p}_1(m_2 \cdot N)$ satisfy:

$$\mathbb{V}\mathbf{ar}[\hat{p}_1(m_1 \cdot N)] = \frac{\mathbb{V}\mathbf{ar}[\hat{p}_1(N)]}{m_1}, \quad \mathbb{V}\mathbf{ar}[\hat{p}_2(m_2 \cdot N)] = \frac{\mathbb{V}\mathbf{ar}[\hat{p}_2(N)]}{m_2}.$$

Then, the choice of the multipliers m_1 and m_2 is done via the following estimators:

$$\hat{m}_1 = \frac{2\delta}{1-\delta}\frac{\mathbb{V}\mathbf{ar}[\hat{p}_1(N)]\hat{l}_0^2}{\hat{p}_1^2(N)\mathbb{V}\mathbf{ar}[\hat{l}_0]},$$
$$\hat{m}_2 = \frac{2\delta}{1-\delta}\frac{\mathbb{V}\mathbf{ar}[\hat{p}_2(N)]}{\hat{p}_1^2(N)\mathbb{V}\mathbf{ar}[\hat{l}_0]}. \tag{4.27}$$

This means, that in order to define the multipliers m_1 and m_2, such that the variance reduction factor of the estimator (4.26) is approximately δ times the variance reduction factor of the estimator (4.25) (i.e. according to (4.27)), we first perform the estimator (4.26) with sample sizes N for all involved estimators and then we evaluate the necessary information and plug it in (4.27). To finalize the estimator (4.26) we need to generate extra samples of $\Delta\mathbf{S}$ of the sizes $(m_1 - 1) \cdot N$ and $(m_2 - 1) \cdot N$, and reevaluate at these samples the approximation Q in order to reestimate the probabilities $\hat{\mathbb{P}}(Q(\Delta\mathbf{S}) \in (y, z))$ and $\hat{\mathbb{P}}(Q(\Delta\mathbf{S}) \geq z)$. This does not involve any extra revaluation of the portfolio loss function L.

Remark 4.4. *In the case that we do not want to perform the special stratification with both levels y and z, but just with one of them, the corresponding multipliers m_1 and m_2 should be adjusted accordingly. One of them should be set to 0 and the second should be halved. For example, if we want to divide the real line only into two parts by defining the level y, we need to estimate only the probability $\mathbb{P}(Q(\Delta\mathbf{S}) > y)$, that corresponds to setting $m_2 = 0$ and halving m_1.*

In order to make the special stratification easy for the user to implement, we suggest in Appendix B the method of the multi-dimensional optimization over the large space, which showed to be the most practical in the numerical study (see Chapter 6) for evaluating the levels y and z from (4.21). These evaluations involve the revaluation of the portfolio loss function, but it is not crucial, as the number of revaluations is often small (100-1000 in the numerical study in Chapter 6). Besides, as our numerical study in Chapter 6 shows, the total time for calculating the special strata is comparable to the

time necessary to calculate 40 strata according to the ISS algorithm (as they involve optimization with inversion integral evaluations).

In further sections we apply the special stratification in both modifications, when the distribution of the approximation Q is known according to (4.25) and when it is not known according to (4.26).

4.3.2 Cross - Entropy with special Stratification in Ideal setting

In this section we improve the performance of the ACE algorithm (see Algorithm 4.3) by applying the special stratification in the ideal setting of known distribution of the approximation Q in the form (4.25). In order to do this, we first have to adjust the final estimator. In Section 4.2 it was done in the following order: we first applied the importance sampling and then performed stratification to the modified expectation under different changes of measure. For the case of special stratification it is easier to perform the special stratification first, and then to apply the importance sampling for the part which is under estimation. More precisely, by the law of total probability, the probability of interest can be decomposed according to the special stratification in the following way:

$$\begin{aligned} \hat{l} &= \mathbb{P}(L(\Delta\mathbf{S}) > x) = \mathbb{P}((L(\Delta\mathbf{S}) > x) \cap (Q(\Delta\mathbf{S}) \leq y)) + \\ &+ \mathbb{P}((L(\Delta\mathbf{S}) > x) \cap (Q(\Delta\mathbf{S}) \in (y, z))) + \mathbb{P}((L(\Delta\mathbf{S}) > x) \cap (Q(\Delta\mathbf{S}) \geq z)). \end{aligned}$$

Due to the choice of the levels y and z it can be simplified in the following way:

$$\hat{l} = \mathbb{P}((L(\Delta\mathbf{S}) > x) \cap (Q(\Delta\mathbf{S}) \in (y, z))) + \mathbb{P}(Q(\Delta\mathbf{S}) \geq z).$$

In order to apply the importance sampling to the first summand (the second is known and stays unchanged), we have to represent it in terms of the expectation:

$$\hat{l} = \mathbb{E}[I_{(x,\infty)}(L(\Delta\mathbf{S}))I_{(y,z)}(Q(\Delta\mathbf{S}))] + \mathbb{P}(Q(\Delta\mathbf{S}) \geq z).$$

Now, the importance sampling can be applied to the expectation part on the right hand side in the standard way. One should, however, keep in mind, that the importance sampling should be performed not w.r.t. the classical change of measure, but w.r.t. the conditional change of measure given that $Q(\Delta\mathbf{S}) \in (y, z)$.

Now we come directly to the formulation of the algorithm, which we call the Approximate Cross-Entropy with Stratification in Ideal setting (ACESI) algorithm:

Algorithm 4.9 (ACESI)**.**

1. *Define* $\hat{\mathbf{v}}_0 = \mathbf{u}$ *and set* $j = 1$. *Fix input parameters* ρ, M *and* N.

2. *Generate a sample* $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$ *from the density* $f(\mathbf{x}; \hat{\mathbf{v}}_{j-1})$ *and compute the sample* $(1-\rho)-$ *quantile* $\hat{x}_j$ *of the sample* $Q(\Delta\mathbf{S}_1), ..., Q(\Delta\mathbf{S}_M)$. *If* $\hat{x}_j > x$, *set* $\hat{x}_j = x$.

3. *Using the same sample* $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$, *find an estimate from the adaptive optimization problem in the form of (2.12) or (2.13),*

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M} \sum_{i=1}^{M} I_{(\hat{x}_j,\infty)}(Q(\Delta\mathbf{S}_i)) W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}_{j-1}) \log f(\Delta\mathbf{S}_i; \mathbf{v}), \tag{4.28}$$

and denote the solution as $\hat{\mathbf{v}}_j$.

4. *If* $\hat{x}_j < x$, *set* $j = j + 1$ *and reiterate from step 2, else set the optimal reference parameter estimate* $\hat{\mathbf{v}}^* = \hat{\mathbf{v}}_j$ *and proceed with step 5.*

5. *Compute the levels* y *and* z *according to (4.21).*

6. *Estimate the probability* l *using the special stratified estimator (4.25) in the form*

$$\begin{aligned} \hat{l} &= \mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \in (y,z)) \frac{1}{N} \sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i)) W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*) + \\ &+ \mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z), \end{aligned} \tag{4.29}$$

where $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ *is a sample generated from the conditional density* $f(\mathbf{x}; \hat{\mathbf{v}}^*)$ *of* $\Delta\mathbf{S}$, *given that* $Q(\Delta\mathbf{S}) \in (y,z)$, *which can be achieved by methods given in Appendix A.*

Remark 4.5. *The ACESI algorithm (Algorithm 4.9) is restricted to the case when the distribution of the approximation* Q *is known. Following the study of this work, this obliges us to use the Delta-Gamma approximation in the case of either Gaussian or Student t-distributed changes in risk factors (see Section 3.3).*

4.3.3 Cross - Entropy with special Stratification

A more general algorithm based on the combination of the ACE algorithm and the special stratification in the form (4.26), named the Approximate Cross-Entropy with Stratification algorithm (ACES), can be achieved from ACESI algorithm (Algorithm 4.9) by estimating the probabilities $\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \in (y,z))$ and $\mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z)$ as described in Section 4.3.1.

We should note, that in contrast to Section 4.3.1, the probability $\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \in (y,z))$ is given under the change of measure. Therefore it is useless to apply the variance

reduction procedures in order to estimate this probability. Instead, we simply apply the MC algorithm. We denote the MC estimator of this probability with sample size $m_1 \cdot N$ by $\hat{p}_1(m_1 \cdot N)$, for some positive number m_1 (notation similar to Section 4.3.1).

From the other side, the probability $\mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z)$ is given under the original density and is often small (if z is large, which is the case for large x and good enough approximation Q). Hence, the variance reduction procedure can be very helpful. As already mentioned in Section 4.3.1, the CE algorithm (FCE algorithm with L substituted by Q, see Algorithm 4.4) is the method of our choice. The first reason is that it is flexible and easy. The second reason is that the estimate of the optimal reference parameter found in ACESI algorithm (the same parameter as in ACES algorithm, see Algorithms 4.9 and 4.10) can be used as an estimate of the optimal reference parameter for estimation of the probability $\mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z)$. We denote CE estimator of this probability with sample size $m_2 \cdot N$ by $\hat{p}_2(m_2 \cdot N)$, for some positive number m_2 (notation similar to Section 4.3.1).

We remind the reader that the performance of the ACES algorithm depends on the multipliers m_1 and m_2 of the corresponding probabilities estimators. These parameters are defined according to the procedure described in Section 4.3.1. This means, that they are set such that the variance reduction factor of the ACES algorithm is approximately δ times the variance reduction factor of the ACESI algorithm, for some $\delta \in (0, 1)$ defined by the user.

All said above summarizes in the following Aproximate Cross-Entropy with Stratification (ACES) algorithm:

Algorithm 4.10 (ACES)**.**

1. *Define* $\hat{\mathbf{v}}_0 = \mathbf{u}$ *and set* $j = 1$*. Fix input parameters* ρ*,* δ*,* M *and* N*.*

2. *Generate a sample* $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$ *from the density* $f(\mathbf{x}; \hat{\mathbf{v}}_{j-1})$ *and compute the sample* $(1-\rho)-$ *quantile* $\hat{x}_j$ *of the sample* $Q(\Delta\mathbf{S}_1), ..., Q(\Delta\mathbf{S}_M)$*. If* $\hat{x}_j > x$*, set* $\hat{x}_j = x$*.*

3. *Using the same sample* $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$*, find an estimate from the adaptive optimization problem in the form of (2.12) or (2.13),*

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M} \sum_{i=1}^{M} I_{(\hat{x}_j,\infty)}(Q(\Delta\mathbf{S}_i)) W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}_{j-1}) \log f(\Delta\mathbf{S}_i; \mathbf{v}), \tag{4.30}$$

   and denote the solution as $\hat{\mathbf{v}}_j$*.*

4. *If* $\hat{x}_j < x$*, set* $j = j + 1$ *and reiterate from step 2, else set the optimal reference parameter estimate* $\hat{\mathbf{v}}^* = \hat{\mathbf{v}}_j$ *and proceed with step 5.*

5. *Compute the levels y and z according to (4.21).*

6. *Generate a sample $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ from the conditional density $f(\mathbf{x}; \hat{\mathbf{v}}^*)$ of $\Delta\mathbf{S}$ given that $Q(\Delta\mathbf{S}) \in (y, z)$ (by methods given in Appendix A). Set $m_1 = 1$ and $m_2 = 1$.*

7. *Calculate the MC estimate $\hat{p}_1(m_1 \cdot N)$ of the probability $\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \in (y, z))$.*

8. *Calculate the CE estimate $\hat{p}_2(m_2 \cdot N)$ of the probability $\mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z)$.*

9. *Estimate the probability l using the special stratified estimator (4.26) in the form*

$$\hat{l} = \hat{p}_1(m_1 \cdot N)\frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*) + \hat{p}_2(m_2 \cdot N), \tag{4.31}$$

   where $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ is the sample from the step 6. Denote also by $\hat{l}_0$ the following term:

$$\hat{l}_0 = \frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*).$$

10. *Calculate m_1 and m_2 by (see (4.27)):*

$$m_1 = \frac{2\delta}{1-\delta}\frac{\mathbb{V}\mathbf{ar}[\hat{p}_1(N)]\hat{l}_0^2}{\hat{p}_1^2(N)\mathbb{V}\mathbf{ar}[\hat{l}_0]},$$
$$m_2 = \frac{2\delta}{1-\delta}\frac{\mathbb{V}\mathbf{ar}[\hat{p}_2(N)]}{\hat{p}_1^2(N)\mathbb{V}\mathbf{ar}[\hat{l}_0]}.$$

    and perform steps 7-9 again.

In the last step, the term $\hat{l}_0$ does not have to be revaluated, therefore, no extra loss revaluations are involved.

Remark 4.6. *The ACES algorithm (see Algorithm 4.10) can be easily modified in such a way, that the multipliers $m_1 \geq 1$ and $m_2 \geq 1$ are taken as input parameters. In such a case, there is no need to define the input parameter δ. Besides, step 10 is, then, redundant.*

Remark 4.7. *In contrast to the ACESI algorithm, any approximation function Q can be used in the ACES algorithm. This is because no information about the function Q, except its revaluation, is involved.*

Remark 4.8. *Although by Assumption 4.1 the approximation Q is significantly faster to revaluate than the portfolio loss function L, for large multipliers, the advantage can vanish. Hence, as we already mentioned in Section 4.3.1, one should find a compromise between the performance, defined by the multipliers m_1 and m_2, and the time necessary to perform step 10. By Remark 4.6, one can always set an upper limit for these multipliers.*

4.3.4 Improved Cross - Entropy with special Stratification in Ideal setting

In this section we apply the special stratification in the ideal setting of know distribution of the approximation Q in the form (4.25) to the AICE algorithm (see Algorithm 4.5). Since the AICE algorithm is different from ACE only in estimating the optimal reference parameter, the so called Approximate Improved Cross-Entropy with Stratification in Ideal setting (AICESI) algorithm, is easily achieved from Algorithm 4.9:

Algorithm 4.11 (AICESI)**.**

1. *Fix input parameters M and N.*
2. *Generate a sample $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$ from the density*

$$g^*(\mathbf{x}) = \frac{I_{(x,+\infty)}(Q(\mathbf{x}))f(\mathbf{x};\mathbf{u})}{\mathbb{P}(Q(\Delta\mathbf{S}) > x)},$$

   with help of methods from Appendix A.

3. *Using this sample find an estimate from the optimization problem,*

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M}\sum_{i=1}^{M} \log f(\Delta\mathbf{S}_i;\mathbf{v}), \tag{4.32}$$

   and denote the solution as $\hat{\mathbf{v}}^*$.

4. *Compute the levels y and z according to (4.21).*
5. *Estimate the probability l using the special stratified estimator (4.25) in the form*

$$\begin{aligned} \hat{l} &= \mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \in (y,z))\frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i;\mathbf{u},\hat{\mathbf{v}}^*) + \\ &+ \mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z), \end{aligned} \tag{4.33}$$

   where $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ is a sample generated from the conditional density $f(\mathbf{x};\hat{\mathbf{v}}^*)$ of $\Delta\mathbf{S}$ given that $Q(\Delta\mathbf{S}) \in (y,z)$, which can be achieved by methods given in Appendix A.

All remarks of Section 4.3.2 apply here.

4.3.5 Improved Cross - Entropy with special Stratification

Similarly to Section 4.3.4, the AICE algorithm (see Algorithm 4.5) combined with the special stratification in the form (4.26), named the Approximate Improved Cross-Entropy with Stratification (AICES) algorithm, can be easily achieved from the Algorithm 4.10:

Algorithm 4.12 (AICES)**.**

1. *Fix input parameters δ, M and N.*
2. *Generate a sample $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$ from the density*
$$g^*(\mathbf{x}) = \frac{I_{(x,+\infty)}(Q(\mathbf{x}))f(\mathbf{x};\mathbf{u})}{\mathbb{P}(Q(\Delta\mathbf{S}) > x)},$$
with help of methods from Appendix A.
3. *Using this sample find an estimate from the optimization problem,*
$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M}\sum_{i=1}^{M} \log f(\Delta\mathbf{S}_i;\mathbf{v}), \tag{4.34}$$
and denote the solution as $\hat{\mathbf{v}}^*$.
4. *Compute the levels y and z according to (4.21).*
5. *Generate a sample $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ from the conditional density $f(\mathbf{x};\hat{\mathbf{v}}^*)$ of $\Delta\mathbf{S}$ given that $Q(\Delta\mathbf{S}) \in (y,z)$ (by methods given in Appendix A). Set $m_1 = 1$ and $m_2 = 1$.*
6. *Calculate the MC estimate $\hat{p}_1(m_1 \cdot N)$ of the probability $\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \in (y,z))$.*
7. *Calculate the CE estimate $\hat{p}_2(m_2 \cdot N)$ of the probability $\mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z)$.*
8. *Estimate the probability l using the special stratified estimator (4.26) in the form*
$$\hat{l} = \hat{p}_1(m_1 \cdot N)\frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i;\mathbf{u},\hat{\mathbf{v}}^*) + \hat{p}_2(m_2 \cdot N), \tag{4.35}$$
where $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ is the sample from step 5. Denote also by $\hat{l}_0$ the following term:
$$\hat{l}_0 = \frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i;\mathbf{u},\hat{\mathbf{v}}^*).$$

9. Calculate m_1 and m_2 by (see (4.27)):

$$m_1 = \frac{2\delta}{1-\delta}\frac{\mathbb{V}\mathbf{ar}[\hat{p}_1(N)]\hat{l}_0^2}{\hat{p}_1^2(N)\mathbb{V}\mathbf{ar}[\hat{l}_0]},$$
$$m_2 = \frac{2\delta}{1-\delta}\frac{\mathbb{V}\mathbf{ar}[\hat{p}_2(N)]}{\hat{p}_1^2(N)\mathbb{V}\mathbf{ar}[\hat{l}_0]}.$$

and perform steps 6-8 again.

All remarks of Section 4.3.3 apply here with the corresponding references to the correct step numbers.

4.4 Algorithms based on approximation of the Optimal Importance Sampling density with Stratification

In the following sections we improve the performance of the methods given in Section 4.3 by improving the importance sampling part. Due to the purpose of a flexible algorithm, according to all said before, our choice of the importance sampling method is the CE algorithm. The ICE algorithm has the same purpose as the CE algorithm, namely to estimate the optimal reference parameter $\mathbf{v}^*$, which minimizes the Kullback-Leibler cross-entropy between the optimal importance sampling density $g^*(\mathbf{x})$ and the parametric family of densities $\mathcal{G} = \{f(\mathbf{x};\mathbf{v}), \mathbf{v} \in \mathcal{V}\}$ of the original density $f(\mathbf{x};\mathbf{u})$ (see Section 2.1.2). Hence, if we forget for the moment about the stratification, the quality of the estimate of the optimal reference parameter $\mathbf{v}^*$, together with the chosen family of densities $\mathcal{G} = \{f(\mathbf{x};\mathbf{v}), \mathbf{v} \in \mathcal{V}\}$ for optimization play the main role in the performance of the importance sampling part of the ACESI, ACES, AICESI and AICES algorithms from Section 4.3. Even if the optimal reference parameter is very well estimated, we can see from Figure 4.4, that the shape of the CE (or ICE) algorithms is not even close to the shape of the optimal importance sampling density $g^*(\mathbf{x})$.

The idea of the new importance sampling algorithm is very simple and is first presented in my joint paper with Ralf Korn (see Korn and Pupashenko (2014)). Here we give, however, a more detailed derivation of the method, together with different variants of it. We chose the family of densities for the change of measure closer to the shape of the optimal importance sampling density $g^*(\mathbf{x})$. The shape of the optimal importance sampling density $g^*(\mathbf{x})$, i.e. with the support bounded from the left by the level x, can be thankful to the fact that it is a conditional density of the changes in risk factors, given that the losses exceed the level x. As we already mentioned before, generally, we can sample from such a density, but it is time inefficient due to the portfolio loss

revaluations. Therefore, the first idea to overcome this problem is to use a much faster approximation Q instead. The family of densities parametrized by the parameter $\mathbf{v}$ then looks like:

$$\frac{I_{(x,+\infty)}(Q(\mathbf{x}))f(\mathbf{x};\mathbf{v})}{\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>x)}.$$

The problem of using this family as the change of measure in the importance sampling algorithm lies in the violation of (2.3). In more details, if the loss exceeds the level x, the approximation Q is not necessary larger than the level x, and the density on which we divide in the importance sampling estimator (2.4) is zero in this case. This problem can be solved by adjusting the above mentioned family of densities by introducing another level y, such that, whenever $L(\Delta\mathbf{S})>x$, the inequality $Q(\Delta\mathbf{S})>y$ holds true. This can be done with help of already introduced level y in the Section 4.3.1, defined by (4.21). The adjusted family of densities parametrized by the parameter $\mathbf{v}$ then looks like:

$$g_Q^*(\mathbf{x},\mathbf{v})=\frac{I_{(y,+\infty)}(Q(\mathbf{x}))f(\mathbf{x};\mathbf{v})}{\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)}. \tag{4.36}$$

One should also note that the density $g_Q^*(\mathbf{x},\mathbf{v})$ is a conditional density of $\Delta\mathbf{S}$ with density $f(\mathbf{x};\mathbf{v})$ given that $Q(\Delta\mathbf{S})>y$. One can sample from such densities by one of the methods from Appendix A. Moreover, these methods do not need the estimation of the probability $\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)$.

The next step is borrowed from the CE algorithm. We find the optimal reference parameter $\mathbf{v}^*$ such, that the density $g_Q^*(\mathbf{x},\mathbf{v}^*)$ has the minimum Kullback-Leibler cross-entropy to the optimal importance sampling density $g^*(\mathbf{x})$, given by (2.11) with $G=L$ and $\gamma=x$. This means, that the parameter $\mathbf{v}^*$ is given by the solution to the following optimization problem:

$$\min_{\mathbf{v}\in\mathcal{V}}\mathcal{D}(g^*,g_Q^*(\cdot;\mathbf{v})).$$

Such a density found for a particular portfolio is drawn in Figure 4.4. One can see, that it is much closer to the optimal importance sampling density than the corresponding change of measure under the CE algorithm.

Similarly to the CE algorithm, we need to simplify the above minimization problem:

$$\min_{\mathbf{v}\in\mathcal{V}}\mathcal{D}(g^*(\cdot),g_Q^*(\cdot;\mathbf{v}))=\min_{\mathbf{v}\in\mathcal{V}}\left(\int g^*(\mathbf{x})\log g^*(\mathbf{x})d\mathbf{x}-\int g^*(\mathbf{x})\log g_Q^*(\mathbf{x};\mathbf{v})d\mathbf{x}\right)$$

in the following way:

$$\begin{aligned}&\max_{\mathbf{v}\in\mathcal{V}}\int I_{(x,+\infty)}(L(\mathbf{x}))f(\mathbf{x};\mathbf{u})\log\frac{I_{(y,+\infty)}(Q(\mathbf{x}))f(\mathbf{x};\mathbf{v})}{\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)}d\mathbf{x}=\\ =\;&\max_{\mathbf{v}\in\mathcal{V}}\mathbb{E}_{\mathbf{u}}\left[I_{(x,+\infty)}(L(\Delta\mathbf{S}))\log\frac{f(\Delta\mathbf{S};\mathbf{v})}{\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)}\right],\end{aligned}$$

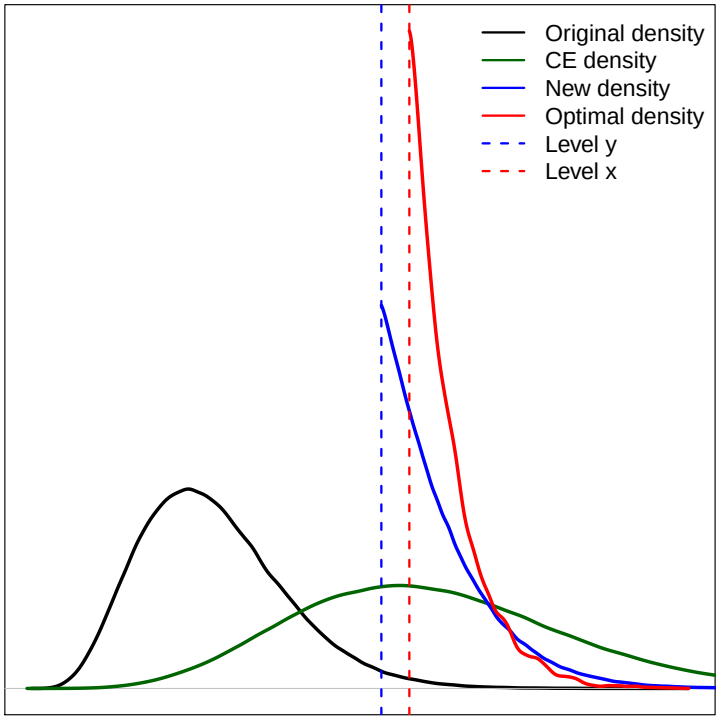


FIGURE 4.4: Estimated densities of losses under the original, CE, new importance sampling and optimal importance sampling densities in 500000 randomly generated scenarios for a portfolio of short 10 ATM calls, 5 ATM puts, 10 uncorrelated assets, 0.1 years maturity, Gaussian changes in risk factors, probability nearly 1% (for details on the portfolio see Chapter 6).

where we have used the fact, that $I_{(y,+\infty)}(Q(\mathbf{x})) = 1$ if $I_{(x,+\infty)}(L(\mathbf{x})) = 1$, otherwise the whole function under integral is zero.

Since we start sampling from the density $g_Q^*(\mathbf{x}, \mathbf{u})$, for which the event of interest is not of small probability, there is no need to perform an update step for the reference parameter given in the CE algorithm. It can be seen as an analogue to the idea of the ICE algorithm, where we sample first from the density similar to $g_Q^*(\mathbf{x}, \mathbf{u})$ and then estimate the optimal reference parameter directly. Therefore, the above maximization problem reads as follows:

$$\max_{\mathbf{v}\in\mathcal{V}} \mathbb{E}^*_{\mathbf{u},Q}\left[I_{(x,+\infty)}(L(\Delta\mathbf{S})) \frac{f(\Delta\mathbf{S};\mathbf{u})}{g_Q^*(\Delta\mathbf{S};\mathbf{u})} \log \frac{f(\Delta\mathbf{S};\mathbf{v})}{\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S}) > y)} \right],$$

where $\mathbb{E}^*_{\mathbf{u},Q}$ means the expectation w.r.t. the density $g_Q^*(\mathbf{x};\mathbf{u})$. Since $\mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) > y)$ involved in the density $g_Q^*(\mathbf{x};\mathbf{u})$ does not depend on the maximization parameter $\mathbf{v}$, it

can be simplified even further:

$$\max_{\mathbf{v}\in\mathcal{V}} \mathbb{E}^*_{\mathbf{u},Q}\left[I_{(x,+\infty)}(L(\Delta\mathbf{S}))\log\frac{f(\Delta\mathbf{S};\mathbf{v})}{\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)}\right].$$

The corresponding sample counterpart for estimating the solution of the above CE maximization problem reads as follows:

$$\max_{\mathbf{v}\in\mathcal{V}} \frac{1}{M}\sum_{i=1}^{M} I_{(x,+\infty)}(L(\Delta\mathbf{S}_i))\log\frac{f(\Delta\mathbf{S}_i;\mathbf{v})}{\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)}, \tag{4.37}$$

where $(\Delta\mathbf{S}_1,\ldots,\Delta\mathbf{S}_M)$ is a sample from the density $g^*_Q(\mathbf{x};\mathbf{u})$.

We should mention, that because $\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)$ depends on the parameter $\mathbf{v}$ w.r.t. which we optimize, we can not get rid of it in the above optimization. Hence, the main advantage of the CE algorithm compared to, e.g., the VM algorithm, namely the often exact solution to the above optimization problem (see Section 2.1.2), vanishes. Therefore, as an alternative, we give here also the corresponding derivation of the optimal reference parameter estimator in the case of the VM approach, which looks as follows (see Section 2.1.2):

$$\begin{aligned}
&\min_{\mathbf{v}\in\mathcal{V}} \mathbb{E}_{\mathbf{u}}\left[I_{(x,+\infty)}(L(\Delta\mathbf{S}))\frac{f(\Delta\mathbf{S};\mathbf{u})}{g^*_Q(\Delta\mathbf{S};\mathbf{v})}\right] = \\
= &\min_{\mathbf{v}\in\mathcal{V}} \mathbb{E}_{\mathbf{u}}\left[I_{(x,+\infty)}(L(\Delta\mathbf{S}))\frac{f(\Delta\mathbf{S};\mathbf{u})}{f(\Delta\mathbf{S};\mathbf{v})}\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)\right] = \\
= &\min_{\mathbf{v}\in\mathcal{V}} \mathbb{E}^*_{\mathbf{u},Q}\left[I_{(x,+\infty)}(L(\Delta\mathbf{S}))\frac{f(\Delta\mathbf{S};\mathbf{u})}{g^*_Q(\Delta\mathbf{S};\mathbf{u})}\frac{f(\Delta\mathbf{S};\mathbf{u})}{f(\Delta\mathbf{S};\mathbf{v})}\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)\right] = \\
= &\min_{\mathbf{v}\in\mathcal{V}} \mathbb{E}^*_{\mathbf{u},Q}\left[I_{(x,+\infty)}(L(\Delta\mathbf{S}))\frac{f(\Delta\mathbf{S};\mathbf{u})}{f(\Delta\mathbf{S};\mathbf{v})}\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)\right].
\end{aligned}$$

The corresponding sample counterpart for estimating the solution of the above VM minimization problem reads as follows:

$$\min_{\mathbf{v}\in\mathcal{V}} \frac{1}{M}\sum_{i=1}^{M} I_{(x,+\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i;\mathbf{u},\mathbf{v})\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y), \tag{4.38}$$

where $(\Delta\mathbf{S}_1,\ldots,\Delta\mathbf{S}_M)$ is a sample from the density $g^*_Q(\mathbf{x};\mathbf{u})$.

Remark 4.9. *Both optimizations (4.37) and (4.38), that are necessary to find an estimate of the optimal reference parameter in the context of the new algorithm, involve M revaluations of the portfolio loss function L. But this sample size M can be set to be small (in our numerical study in Chapter 6 it is around 100-1000), because under the density $g^*_Q(\mathbf{x};\mathbf{u})$ the probability $\mathbb{P}(L(\Delta\mathbf{S})>x)$ is not small, and most of the indicators*

in the corresponding sums in (4.37) or (4.38) are not zeros. One should also mention, that in contrast to the FICE algorithm, sampling from the density $g_Q^*(\mathbf{x};\mathbf{u})$ *does not involve any portfolio revaluations and can be done efficiently with the help of the methods presented in Appendix A.*

4.4.1 Approximate Optimal Importance Sampling algorithm in Ideal setting

In this section we give the precise formulation of the new method based on the new importance sampling algorithm presented at the beginning of Section 4.4. Here we deal with the ideal situation of known distribution of the approximation Q. Similarly to Section 4.3.2, such an ideal situation restricts the choice of the approximation Q to the Delta-Gamma approximation in the case of either Gaussian or Student t-distributed changes in risk factors (see Section 3.3). The corresponding algorithm is called Approximate Optimal Importance Sampling in Ideal setting (AOISI) algorithm.

This algorithm is based on the change of measure corresponding to the density $g_Q^*(\mathbf{x};\hat{\mathbf{v}}^*)$ given by (4.36), where $\hat{\mathbf{v}}^*$ is an estimate of the optimal reference parameter, that is found from one of the optimization problems, either the CE optimization (4.37) or the VM optimization (4.38).

The importance sampling estimator (2.4) then reads as follows:

$$\begin{aligned}\hat{l} &= \frac{1}{N}\sum_{i=1}^{N} I_{(x,+\infty)}(L(\Delta\mathbf{S}_i))\frac{f(\Delta\mathbf{S}_i;\mathbf{u})}{g_Q^*(\Delta\mathbf{S}_i;\hat{\mathbf{v}}^*)} = \\ &= \mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S})>y)\frac{1}{N}\sum_{i=1}^{N} I_{(x,+\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i;\mathbf{u},\hat{\mathbf{v}}^*),\end{aligned}$$

where $(\Delta\mathbf{S}_1,\ldots,\Delta\mathbf{S}_N)$ is a sample from the density $g_Q^*(\mathbf{x};\hat{\mathbf{v}}^*)$.

In the ideal situation the probability $\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S})>y)$ from the importance sampling estimator, as well as the probability $\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)$ from the optimal reference parameter estimators (4.37) and (4.38), are known. Therefore, we can now come directly to the formulation of the corresponding algorithm:

Algorithm 4.13 (AOISI)**.**

1. *Fix input parameters* M *and* N*.*
2. *Compute the level* y *according to (4.21).*
3. *Generate a sample* $\Delta\mathbf{S}_1,...,\Delta\mathbf{S}_M$ *from the density* $g_Q^*(\mathbf{x};\mathbf{u})$ *given by (4.36) with the help of the methods from Appendix A.*

4. *Using this sample find an estimate from one of the optimization problems (4.37) or (4.38) and denote the solution as* $\hat{\mathbf{v}}^*$.

5. *Estimate the probability* l *using the importance sampling estimator (2.4) in the form*

$$\hat{l} = \mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) > y)\frac{1}{N}\sum_{i=1}^{N} I_{(x,+\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i;\mathbf{u},\hat{\mathbf{v}}^*), \tag{4.39}$$

where $\Delta\mathbf{S}_1,...,\Delta\mathbf{S}_N$ *is a sample generated from the density* $g_Q^*(\mathbf{x};\hat{\mathbf{v}}^*)$ *by means of the methods given in Appendix A.*

4.4.2 Approximate Optimal Importance Sampling algorithm

In this section we extend the AOISI algorithm to the general case of an unknown distribution of the approximation Q. For that purpose, similarly to the idea of the special stratification (see Section 4.3.1), the probability $\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) > y)$ involved in the importance sampling estimator as well as the probability $\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S}) > y)$ from the optimal reference parameter estimators (4.37) and (4.38), have to be estimated by means of MC or some variance reduction method. Since all these probabilities are given not under the original density with parameter $\mathbf{u}$, the use of the variance reduction may be inefficient. Therefore we stop our choice on applying the simple MC estimator. The corresponding algorithm is called Approximate Optimal Importance Sampling (AOIS) algorithm.

The MC estimator of the probability $\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) > y)$ influences the variance of the final estimator directly (situation similar to Section 4.3.1), therefore, has to be estimated by MC simulation with sample size $m \cdot N$ for some multiplier m introduced in a similar way to the special stratification. The AOISI algorithm serves the role of the reference estimator here. This means, that the multiplier m is defined in such a way, that the variance reduction factor of the AOIS algorithm is approximately the variance reduction factor of the AOISI algorithm multiplied by some input parameter $\delta \in (0,1)$ (see Section 4.3.1). We denote the corresponding estimator $\hat{\mathbb{P}}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) > y)$ in the similar way to Section 4.3.1, i.e. by $\hat{p}(m \cdot N)$. According to Remark 4.4, the multiplier m can easily be found by:

$$\hat{m} = \frac{\delta}{1-\delta}\frac{\mathbb{V}\mathbf{ar}[\hat{p}(N)]\hat{l}_0^2}{\hat{p}^2(N)\mathbb{V}\mathbf{ar}[\hat{l}_0]}, \tag{4.40}$$

with the same notation for $\hat{l}_0$ and the only difference, that the underlying sample of $\Delta\mathbf{S}_i$ is generated from other density.

On the other hand, the probability $\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S}) > y)$ does not influence the variance reduction rate directly, rather the quality of the optimal reference parameter estimator

achieved by either (4.37) or (4.38). It can be estimated with some sample size K (more on it shortly). But one should notice, that this probability has to be estimated for all parameters $\mathbf{v}$ simultaneously. If we use a classical MC estimator, this means a separate sample of size K for each parameter $\mathbf{v}$, which is highly inefficient. Instead, we modify the estimator by applying a sort of importance sampling estimator with the change of measure being described by the density $g_Q^*(\mathbf{x};\mathbf{u})$. More precisely:

$$\begin{aligned}\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y) &= \mathbb{E}_{\mathbf{v}}[I_{(y,+\infty)}(Q(\Delta\mathbf{S}))] = \mathbb{E}^*_{\mathbf{u},Q}\left[I_{(y,+\infty)}(Q(\Delta\mathbf{S}))\frac{f(\Delta\mathbf{S}_i;\mathbf{v})}{g_Q^*(\Delta\mathbf{S}_i;\mathbf{u})}\right] = \\ &= \mathbb{E}^*_{\mathbf{u},Q}[I_{(y,+\infty)}(Q(\Delta\mathbf{S}))W(\Delta\mathbf{S};\mathbf{v},\mathbf{u})\mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S})>y)] = \\ &= \mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S})>y)\cdot\mathbb{E}^*_{\mathbf{u},Q}[W(\Delta\mathbf{S};\mathbf{v},\mathbf{u})],\end{aligned}$$

where the last equality holds true, since under the density $g_Q^*(\mathbf{x};\mathbf{u})$, the indicator under expectation $I_{(y,+\infty)}(Q(\Delta\mathbf{S}))$ is a constant and equals to 1.

Since the constant $\mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S})>y)$ does not influence the optimizations (4.37) and (4.38), the following estimator of the probability $\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)$ is proposed:

$$\frac{1}{K}\sum_{i=1}^{K}W(\Delta\mathbf{S}_i;\mathbf{v},\mathbf{u}),$$

where $(\Delta\mathbf{S}_1,\ldots,\Delta\mathbf{S}_K)$ is a sample from the density $g_Q^*(\mathbf{x};\mathbf{u})$, achieved by one of the methods from Appendix A. This estimator, obviously, uses only one sample of size K for estimating all the probabilities $\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S})>y)$ simultaneously.

The optimizations (4.37) and (4.38) take now the following forms - CE variant:

$$\max_{\mathbf{v}\in\mathcal{V}}\frac{1}{M}\sum_{i=1}^{M}I_{(x,+\infty)}(L(\Delta\mathbf{S}_i))\log\frac{f(\Delta\mathbf{S}_i;\mathbf{v})}{\frac{1}{K}\sum_{j=1}^{K}W(\Delta\mathbf{S}_{M+j};\mathbf{v},\mathbf{u})}, \tag{4.41}$$

and VM variant:

$$\min_{\mathbf{v}\in\mathcal{V}}\frac{1}{M}\sum_{i=1}^{M}I_{(x,+\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i;\mathbf{u},\mathbf{v})\frac{1}{K}\sum_{j=1}^{K}W(\Delta\mathbf{S}_{M+j};\mathbf{v},\mathbf{u}), \tag{4.42}$$

where $(\Delta\mathbf{S}_1,\ldots,\Delta\mathbf{S}_M,\Delta\mathbf{S}_{M+1},\ldots,\Delta\mathbf{S}_{M+K})$ is a sample from the density $g_Q^*(\mathbf{x};\mathbf{u})$.

Remark 4.10. *The sample size K does not have to be large, it can be considerably smaller than the sample size N for the estimator of the probability of high portfolio losses. It can be simply chosen to be equal to the sample size M.*

All said above summarizes in the following algorithm:

Algorithm 4.14 (AOIS)**.**

1. *Fix input parameters δ, M, K and N.*

2. *Compute the level y according to (4.21).*

3. *Generate a sample $(\Delta\mathbf{S}_1, \ldots, \Delta\mathbf{S}_{M+K})$ from the density $g_Q^*(\mathbf{x}; \mathbf{u})$ given by (4.36) with the help of the methods from Appendix A.*

4. *Using this sample find an estimate from one of the optimization problems (4.41) or (4.42) and denote the solution as $\hat{\mathbf{v}}^*$.*

5. *Generate a sample $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ from the density $g_Q^*(\mathbf{x}; \hat{\mathbf{v}}^*)$ by means of the methods given in Appendix A. Set $m = 1$.*

6. *Calculate the MC estimate $\hat{p}(m \cdot N)$ of the probability $\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) > y)$.*

7. *Estimate the probability l using the importance sampling estimator (2.4) in the form*

$$\hat{l} = \hat{p}(m \cdot N)\frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*), \tag{4.43}$$

   where $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ is the sample from step 5. Denote also by $\hat{l}_0$ the following term:

$$\hat{l}_0 = \frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*).$$

8. *Calculate m by (see (4.40)):*

$$m = \frac{\delta}{1-\delta}\frac{\mathbb{V}\mathbf{ar}[\hat{p}(N)]\hat{l}_0^2}{\hat{p}^2(N)\mathbb{V}\mathbf{ar}[\hat{l}_0]}.$$

   and perform steps 6-7 again.

In the next two sections we improve the AOISI and AOIS algorithms by applying the special stratification from Section 4.3.1. It is somehow tricky, since in these cases we deal with the support being not the whole real line, but with support bounded from the left by the level y. This can be either understood, as we stratify the interval $[y, \infty)$ instead of the real line by defining the level z, or as we already have performed the stratification with one level y and we want to divide the real line further by introducing the level z.

4.4.3 Approximate Optimal Importance Sampling with Stratification algorithm in Ideal setting

In this section we apply the special stratification from Section 4.3.1 to the AOISI algorithm, i.e. in the ideal setting, when the distribution of the approximation Q is known.

As we already mentioned in previous sections, there is no need to perform special stratification with both levels y and z, since the support of the density we are sampling from is already above the level y. Therefore, the special stratification in this case is done via the partition in two strata, (y, z) and (z, ∞), achieved by introducing the level z by (4.21).

The procedure of estimating the optimal reference parameter is the same. The way of adjusting the final estimator is very similar to Section 4.3.2. The only difference is that here we deal with the change of measure $g^*_Q(\mathbf{x}; \hat{\mathbf{v}}^*)$ instead of $f(\mathbf{x}; \hat{\mathbf{v}}^*)$. If we denote by $\mathbb{P}^*_{\hat{\mathbf{v}}^*,Q}$ the probability measure w.r.t. the density $g^*_Q(\mathbf{x}; \hat{\mathbf{v}}^*)$, then the corresponding final estimator looks as follows:

$$\begin{aligned} \hat{l} &= \mathbb{P}^*_{\hat{\mathbf{v}}^*,Q}(Q(\Delta\mathbf{S}) \in (y, z)) \frac{1}{N} \sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i)) W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*) \mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \geq y) + \\ &+ \mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z), \end{aligned}$$

where $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ is a sample generated from the conditional density $g^*_Q(\mathbf{x}; \hat{\mathbf{v}}^*)$ of $\Delta\mathbf{S}$, given that $Q(\Delta\mathbf{S}) \in (y, z)$, which can be achieved by methods given in Appendix A.

The latter one can be further simplified, since the following holds true (by definition of the conditional probabilities):

$$\mathbb{P}^*_{\hat{\mathbf{v}}^*,Q}(Q(\Delta\mathbf{S}) \in (y, z)) = \frac{\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \in (y, z))}{\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \geq y)}.$$

The final estimator then takes the form:

$$\hat{l} = \mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \in (y, z)) \frac{1}{N} \sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i)) W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*) + \mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z).$$

All said above summarizes in the following algorithm which we call Approximate Optimal Importance Sampling with Stratification in Ideal setting (AOISSI) algorithm:

Algorithm 4.15 (AOISSI)**.**

1. *Fix input parameters M and N.*
2. *Compute the levels y and z according to (4.21).*
3. *Generate a sample $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_M$ from the density $g^*_Q(\mathbf{x}; \mathbf{u})$ given by (4.36) with the help of the methods from Appendix A.*
4. *Using this sample find an estimate from one of the optimization problems (4.37) or (4.38) and denote the solution as $\hat{\mathbf{v}}^*$.*

5. *Estimate the probability l using the importance sampling estimator (2.4) in the form*

$$\begin{aligned} \hat{l} &= \mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \in (y,z)) \frac{1}{N} \sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i)) W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*) + \\ &+ \mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z), \end{aligned} \tag{4.44}$$

where $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ is a sample generated from the conditional density $g_Q^*(\mathbf{x}; \hat{\mathbf{v}}^*)$ of $\Delta\mathbf{S}$, given that $Q(\Delta\mathbf{S}) \in (y,z)$, which can be achieved by means of the methods given in Appendix A.

4.4.4 Approximate Optimal Importance Sampling with Stratification algorithm

In this section we extend the AOISSI algorithm to the general case of unknown distribution of the approximation Q in a similar way to Section 4.4.2. The procedure of estimating the probability $\mathbb{P}_{\mathbf{v}}(Q(\Delta\mathbf{S}) > y)$ from the reference parameter estimators (4.37) and (4.38) for all parameters $\mathbf{v}$ simultaneously is the same here and is described by the estimators (4.41) and (4.42). With respect to the estimators of the probabilities $\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \in (y,z))$ and $\mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z)$ we refer the reader to Section 4.3.1. The corresponding estimators involve the calculation of the multipliers m_1 and m_2 in order to achieve the variance reduction factor, approximately the variance reduction factor of AOISSI multiplied by some input parameter $\delta \in (0,1)$. These estimators, as well as the formula for the multipliers, can be taken directly from the AICES algorithm (see Algorithm 4.12), since the same probabilities are involved there.

The following algorithm, called the Approximate Optimal Importance Sampling with Stratification (AOISS) algorithm then reads as follows:

Algorithm 4.16 (AOISS)**.**

1. *Fix input parameters δ, M and N.*
2. *Compute the levels y and z according to (4.21).*
3. *Generate a sample $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_{M+K}$ from the density $g_Q^*(\mathbf{x}; \mathbf{u})$ given by (4.36) with the help of the methods from Appendix A.*
4. *Using this sample find an estimate from one of the optimization problems (4.41) or (4.42) and denote the solution as $\hat{\mathbf{v}}^*$.*

5. *Generate a sample* $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ *from the conditional density* $g_Q^*(\mathbf{x}; \hat{\mathbf{v}}^*)$ *of* $\Delta\mathbf{S}$ *given that* $Q(\Delta\mathbf{S}) \in (y, z)$ *(by methods given in Appendix A). Set* $m_1 = 1$ *and* $m_2 = 1$.

6. *Calculate the MC estimate* $\hat{p}_1(m_1 \cdot N)$ *of the probability* $\mathbb{P}_{\hat{\mathbf{v}}^*}(Q(\Delta\mathbf{S}) \in (y, z))$.

7. *Calculate the CE estimate* $\hat{p}_2(m_2 \cdot N)$ *of the probability* $\mathbb{P}_{\mathbf{u}}(Q(\Delta\mathbf{S}) \geq z)$.

8. *Estimate the probability* l *using the special stratified estimator (4.26) in the form*

$$\hat{l} = \hat{p}_1(m_1 \cdot N)\frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*) + \hat{p}_2(m_2 \cdot N), \tag{4.45}$$

   where $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ *is the sample from step 5. Denote also by* $\hat{l}_0$ *the following term:*

$$\hat{l}_0 = \frac{1}{N}\sum_{i=1}^{N} I_{(x,\infty)}(L(\Delta\mathbf{S}_i))W(\Delta\mathbf{S}_i; \mathbf{u}, \hat{\mathbf{v}}^*).$$

9. *Calculate* m_1 *and* m_2 *by (see (4.27)):*

$$m_1 = \frac{2\delta}{1-\delta}\frac{\mathbb{V}\mathbf{ar}[\hat{p}_1(N)]\hat{l}_0^2}{\hat{p}_1^2(N)\mathbb{V}\mathbf{ar}[\hat{l}_0]},$$
$$m_2 = \frac{2\delta}{1-\delta}\frac{\mathbb{V}\mathbf{ar}[\hat{p}_2(N)]}{\hat{p}_1^2(N)\mathbb{V}\mathbf{ar}[\hat{l}_0]}.$$

   and perform steps 6-8 again.

Remark 4.11. *One can notice, that the estimator (4.45) in the AOISS algorithm looks exactly like the estimators (4.31) and (4.35) in the ACES and AICES algorithms, respectively. The main difference is that the parameter* $\hat{\mathbf{v}}^*$ *is different, and that the sample* $\Delta\mathbf{S}_1, ..., \Delta\mathbf{S}_N$ *is achieved here from the conditional density* $g_Q^*(\mathbf{x}; \hat{\mathbf{v}}^*)$ *of* $\Delta\mathbf{S}$*, given that* $Q(\Delta\mathbf{S}) \in (y, z)$*, whereas in the algorithms ACES and AICES it is a sample from the conditional density* $f(\mathbf{x}; \hat{\mathbf{v}}^*)$ *of* $\Delta\mathbf{S}$*, given that* $Q(\Delta\mathbf{S}) \in (y, z)$.

4.5 Summary

In this section we review all methods introduced in the previous sections of Chapter 4. We compare the times necessary to prepare the final estimator by referring to the approximate number of revaluations of the approximation Q and the loss function L. The optimizations are usually also time consuming, therefore, we also include them in the analysis. We assume, that the probability of interest is estimated with the number N of revaluations of the portfolio loss function in all methods.

In Chapter 4 we introduced the following algorithms, which we combine according to their variance reduction rate (see the numerical study in Chapter 6):

No variance reduction: MC.

Importance sampling variance reduction: IS, ACE, FCE, AICE and FICE.

Importance sampling with stratification variance reduction: ISS, ACESI, ACES, AICESI and AICES.

Further improved variance reduction: AOISI, AOIS, AOISSI and AOISS.

The flexible algorithms (any approximation Q of the loss function L can be used, no distribution of Q is involved) with acceptable time necessary to set up the final estimator are ACE, AICE, ACES, AICES, AOIS and AOISS.

The following algorithms use the portfolio loss function in order to estimate the reference parameter: FCE, FICE, AOISI, AOIS, AOISSI and AOISS. Therefore, in the cases when there is no good enough approximation and other methods fail to give any variance reduction, these methods can be used. The AOIS and AOISS algorithms are the most useful in such cases, since they involve the lowest number of portfolio revaluations in the set up of the importance sampling density.

Now we go through all algorithms mentioned above and we record the most time consuming steps (heavy multidimensional optimizations, large samples of the approximation Q revaluation and, especially, the number of the portfolio losses revaluations):

MC: No extra time costs.

IS: One multidimensional optimization with inversion integral, which involves the revaluation of Q only.

ACE: M multiplied by the number of iterations (usually 4) of evaluations of Q. Optimization is often done via an explicit solution.

FCE: M multiplied by the number of iterations (usually 4) of evaluations of L. Optimization is often done via an explicit solution.

AICE: Sample of size M from the conditional density by MCMC method, involving revaluations of Q.

FICE: Sample of size M from the conditional density by MCMC method, involving revaluations of L.

ISS: $K+1$ - multidimensional optimizations with inversion integral, which involve the revaluation of Q only (K is the number of strata, in the numerical study of e.g. Glasserman (2004) chosen to be 40).

ACESI: M multiplied by the number of iterations (usually 4) of evaluations of Q. Optimization is often done via an explicit solution. Levels y and z involve a small number (100-1000) of Q and L revaluations.

ACES: M multiplied by the number of iterations (usually 4) of evaluations of Q. Optimization is often done via an explicit solution. Levels y and z involve a small number (100-1000) of Q and L revaluations. Extra sample of the total size $(m_1 + m_2) \cdot N$ of revaluations of Q.

AICESI: Sample of size M from the conditional density by MCMC method, involving revaluations of Q. Levels y and z involve a small number (100-1000) of Q and L revaluations.

AICES: Sample of size M from the conditional density by MCMC method, involving revaluations of Q. Levels y and z involve a small number (100-1000) of Q and L revaluations. Extra sample of the total size $(m_1 + m_2) \cdot N$ of revaluations of Q.

AOISI: Level y involves a small number (100-1000) of Q and L revaluations. Sample of small size (100-1000) from the conditional density by MCMC method, involving revaluations of Q and the same number of evaluations of L for optimization.

AOIS: Level y involves a small number (100-1000) of Q and L revaluations. Sample of small size (100-1000) from the conditional density by MCMC method, involving revaluations of Q and the same number of evaluations of L for optimization. Extra sample of size K from the conditional density by MCMC method, involving revaluations of Q. Extra sample of the total size $m \cdot N$ of revaluations of Q.

AOISSI: Levels y and z involve a small number (100-1000) of Q and L revaluations. Sample of small size (100-1000) from the conditional density by MCMC method, involving revaluations of Q and the same number of evaluations of L for optimization.

AOISS: Levels y and z involve a small number (100-1000) of Q and L revaluations. Sample of small size (100-1000) from the conditional density by MCMC method, involving revaluations of Q and the same number of evaluations of L for optimization. Extra sample of size K from the conditional density by MCMC method, involving revaluations of Q. Extra sample of the total size $(m_1 + m_2) \cdot N$ of revaluations of Q.

One should note that sampling from the conditional densities by the HITRO algorithm given in Appendix A is quite fast and usually involves less then 7 revaluations of the involved density function (see Karawatzki et al. (2005)). The sample size M is often taken smaller than the sample size N. Under Assumption 4.1, the revaluation of Q takes much less time than the revaluation of L. The total sum of the multipliers in the methods, where they are involved, are given in different examples in the numerical study in Chapter 6.

Now let us consider the corresponding time consumptions of all methods mentioned above for some particular choice of the parameters (similar to the numerical study in Chapter 6). We assume that $N = 100000$ and we set $M = 10000$, $K = 1000$ (for AOIS and AOISS algorithms), the number of Q and L revaluations is 100 for each value y and z. We also assume that for the AOISI, AOIS, AOISSI and AOISS algorithms we use a sample of size 100 from the conditional density (each involves the revaluation of L and Q). Besides, we assume the worst case of 7 revaluations of the corresponding density function in the HITRO algorithm and the number of 4 iterations in each CE based algorithm. Moreover, we assume, according to Assumption 4.1, that the approximation Q is 100 times faster than the portfolio loss function L. Then, the preparation steps of the importance sampling estimator in the corresponding algorithms have the following number of revaluations in terms of the loss function (e.g. 100 revaluations of Q is the same as 1 revaluation of L):

MC: 0 revaluations of L, 0% of the time spent for the final estimator (100000 revaluations of L).

IS: Not comparable, approximately 0.5% according to the numerical study.

ACE: 400 revaluations of L, 0.4% of the time spent for the final estimator.

FCE: 40000 revaluations of L, 40% of the time spent for the final estimator.

AICE: 700 revaluations of L, 0.7% of the time spent for the final estimator.

FICE: 70000 revaluations of L, 70% of the time spent for the final estimator.

ISS: Approximately 1% according to the numerical study (40 strata).

ACESI: 602 revaluations of L, 0.6% of the time spent for the final estimator.

ACES: $602 + 1000(m_1 + m_2)$ revaluations of L, $(m_1 + m_2 + 0.6)$% of the time spent for the final estimator.

AICESI: 902 revaluations of L, 0.9% of the time spent for the final estimator.

AICES: $902 + 1000(m_1 + m_2)$ revaluations of L, $(m_1 + m_2 + 0.9)\%$ of the time spent for the final estimator.

AOISI: 808 revaluations of L, 0.8% of the time spent for the final estimator.

AOIS: $878 + 1000m$ revaluations of L, $(m + 0.9)\%$ of the time spent for the final estimator.

AOISSI: 909 revaluations of L, 0.9% of the time spent for the final estimator.

AOISS: $979 + 1000(m_1 + m_2)$ revaluations of L, $(m_1 + m_2 + 1)\%$ of the time spent for the final estimator.

Besides the variance reduction factors, which we present in the numerical study in Chapter 6, it is always important to compare the extra time spent in order to set up the corresponding algorithm, as it is done above.

Chapter 5

Variance Reduction of Value-at-Risk and Conditional Value-at-Risk

Chapter 4 was devoted to the derivation of flexible and efficient tools in order to estimate the probability of high portfolio losses. But such a quantity is not a usual market risk measure. As we already mentioned in Chapter 1, VaR and CVaR have become very popular market risk measures in all financial institutions. Therefore, it is important to construct variance reduction procedures that estimate these quantities directly. This can be done with the help of the algorithms for the estimation of the probability of high portfolio losses, and is the goal of this chapter.

In Section 5.1 we start from the rigorous mathematical definitions of VaR and CVaR. Section 5.2 is devoted to the adjustment of the variance reduction procedures from Chapter 4 to the problem of VaR and CVaR estimation. In Section 5.3 we perform an asymptotic analysis of the corresponding estimators and describe the procedure of constructing confidence intervals. Finally, Section 5.4 summarizes the procedure for the transition from the estimation of the probability of high portfolio losses to the problem of estimating VaR and CVaR.

5.1 Value-at-Risk and Conditional Value-at-Risk definitions

We already mentioned in Chapter 1, that VaR measures the worst expected loss over a given time horizon at a given confidence level. The information about the average level of loss, given that VaR is exceeded, can be achieved using CVaR. In this section we

give more rigorous mathematical definitions for these risk measures for continuous distributions (see Assumption 2.1) as given in, e.g., Pflug (2000), Rockafellar and Uryasev (2000, 2002), Trindade et al. (2007).

Under the same notations from Chapters 3 and 4, we denote here by $\mathbb{F}(x) = \mathbb{P}(L \leq x)$ the cumulative distribution function of the portfolio losses L. Then, we can formulate the following definition (see e.g. Trindade et al. (2007)):

Definition 5.1. The portfolio VaR with confidence level $\alpha \in (0,1)$, denoted by v_α, is defined as the α-quantile of the cumulative distribution function $\mathbb{F}(x)$, i.e.:

$$v_\alpha = \mathbb{F}^{-1}(x) = \inf\{x : \mathbb{F}(x) \geq \alpha\} = \inf\{x : \mathbb{P}(L > x) \leq 1 - \alpha\}. \tag{5.1}$$

In practice, the confidence level α typically lies between 95% and 99.9%.

As we already mentioned in Section 1.1.2, there are different variants of CVaR which, however, coincide for continuous loss distributions, which is the case here (see Assumption 2.1). Further, we give the definitions of the most useful among them. Trindade et al. (2007) defines CVaR as follows:

Definition 5.2. The portfolio CVaR with confidence level $\alpha \in (0,1)$, denoted by c_α, is defined as:

$$c_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 v_\beta d\beta. \tag{5.2}$$

Pflug (2000) shows, that CVaR can also be defined as:

Definition 5.3. The portfolio CVaR with confidence level $\alpha \in (0,1)$, denoted by c_α, is defined as:

$$c_\alpha = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{1-\alpha} \mathbb{E}[L - t]_+ \right\}, \tag{5.3}$$

where $[\cdot]_+$ denotes the nonnegative part.

Trindade et al. (2007) also show the following definition:

Definition 5.4. The portfolio CVaR with confidence level $\alpha \in (0,1)$, denoted by c_α, is defined as:

$$c_\alpha = v_\alpha + \frac{1}{1-\alpha} \mathbb{E}[L - v_\alpha]_+. \tag{5.4}$$

Finally, Rockafellar and Uryasev (2000) define CVaR as ES or TCE of the form:

Definition 5.5. The portfolio CVaR with confidence level $\alpha \in (0,1)$, denoted by c_α, is defined as:

$$c_\alpha = \mathbb{E}[L | L \geq v_\alpha]. \tag{5.5}$$

All these definitions have their advantages. The definition of CVaR in the form (5.3) can be used to find VaR and CVaR simultaneously. The parameter t, at which the right hand side of (5.3) is minimized, is exactly VaR. This has its particular value when it is hard to decide which market risk measure to use (VaR or CVaR). The definition of CVaR in the form (5.4) is useful, e.g., when we want to investigate the asymptotics of CVaR, having already derived the asymptotics of VaR. The definition of CVaR in the form (5.5) is used by Glasserman et al. (2002) in order to apply the ISS algorithm for CVaR estimation.

5.2 Value-at-Risk and Conditional Value-at-Risk estimation

There is a various literature on MC and variance reduction procedures in the context of estimating VaR and CVaR. It is, however, not consecutive and not complete. Serfling (1980) has first developed MC estimators and their asymptotics for VaR and CVaR. Later, Glasserman (2004), Glasserman et al. (2000b), Glynn (1996) investigated the MC, IS and ISS estimators and their asymptotics for VaR only. Glasserman et al. (2002) looked at the IS estimator together with the asymptotics for CVaR. Hong and Liu (2011), Sun and Hong (2009, 2010) made a more complete investigation of MC and IS estimators and their asymptotics for both VaR and CVaR. However, only the authors in Chu (2010), Chu and Nakayama (2012) made a complete investigation of the MC, IS and ISS algorithms with their asymptotics, but only for VaR. In Section 5.2 we give a consecutive development of all variance reduction estimators from Chapter 4 for both market risk measures, namely VaR and CVaR. We further analyze their asymptotics in Section 5.3, and construct there the confidence intervals.

5.2.1 Monte Carlo simulation

In this chapter we introduce the MC estimator of VaR and CVaR. We start from VaR estimation. We first remind the reader the form of the MC estimator:

$$\hat{\mathbb{P}}(L(\Delta\mathbf{S}) \leq x) = \hat{\mathbb{F}}_N(x) = \frac{1}{N}\sum_{i=1}^{N} I_{(-\infty,x)}(L(\Delta\mathbf{S}_i)),$$

where $\Delta\mathbf{S}_1, \ldots, \Delta\mathbf{S}_N$ is a sample of independent identically distributed random vectors with joint density $f(\mathbf{x};\mathbf{u})$. Using one sample $\Delta\mathbf{S}_1, \ldots, \Delta\mathbf{S}_N$, the whole distribution function $\mathbb{F}$ is estimated (for all values x simultaneously) by the empirical distribution function $\hat{\mathbb{F}}_N$.

Glynn (1996) suggests another, more efficient estimator of the empirical distribution function, which is even more convenient in the context of our investigation, since it involves the probability of high portfolio losses:

$$\hat{\mathbb{P}}(L(\Delta\mathbf{S}) \leq x) = \hat{\mathbb{F}}_N(x) = 1 - \frac{1}{N}\sum_{i=1}^{N} I_{(x,+\infty)}(L(\Delta\mathbf{S}_i)), \tag{5.6}$$

where $\Delta\mathbf{S}_1, \ldots, \Delta\mathbf{S}_N$ is a sample of independent identically distributed random vectors with joint density $f(\mathbf{x};\mathbf{u})$.

Since VaR is a quantile of the cumulative distribution function of the portfolio loss, many authors suggest the order statistics estimator for VaR (see e.g. Chu (2010), Chu and Nakayama (2012), Hong and Liu (2011), Serfling (1980), Sun and Hong (2009)), which does not depend on the form of the estimator of the empirical distribution function:

$$\hat{v}_\alpha = L_{([N\cdot\alpha])},$$

where $L_{(1)} \leq \ldots \leq L_{(N)}$ denote the ordered statistics that are obtained from the sample $L(\Delta\mathbf{S}_1), \ldots, L(\Delta\mathbf{S}_N)$. This estimator is, however, not convenient for the purpose of extending the estimator to the variance reduction procedures.

An alternative estimator is achieved via an inverse of the empirical distribution function in the form (5.6) (see e.g. Glasserman (2004), Glasserman et al. (2000b), Glynn (1996), Sun and Hong (2010)):

$$\hat{v}_\alpha = \inf\{x : \hat{\mathbb{F}}_N(x) \geq \alpha\} = \inf\{x : \frac{1}{N}\sum_{i=1}^{N} I_{(x,+\infty)}(L(\Delta\mathbf{S}_i)) \leq 1-\alpha\}. \tag{5.7}$$

Glasserman (2004) mentions that, applying the piecewise linear interpolation to the empirical distribution function before taking the inverse, produces generally more accurate quantile estimates.

In regard to the estimator for CVaR, Glasserman et al. (2002) suggests to use the definition of CVaR according to (5.5), which we find, however, inconvenient for our purpose of further investigation of the different variance reduction procedures together with their asymptotics. Instead, we follow the research of Hong and Liu (2011), Sun and Hong (2009, 2010) and define the following estimator of CVaR, which is based on the definition of CVaR according to (5.4):

$$\hat{c}_\alpha = \hat{v}_\alpha + \frac{1}{N\cdot(1-\alpha)}\sum_{i=1}^{N}[L(\Delta\mathbf{S}_i) - \hat{v}_\alpha]_+, \tag{5.8}$$

where $\Delta\mathbf{S}_1, \ldots, \Delta\mathbf{S}_N$ is a sample of independent identically distributed random vectors with joint density $f(\mathbf{x}; \mathbf{u})$.

We further use the MC estimator $\hat{v}_\alpha$ of VaR in the form (5.7) and the MC estimator $\hat{c}_\alpha$ of CVaR in the form (5.8). In order to extend these MC estimators to the variance reduction procedures from Chapter 4, we have to look at the final estimators. In the notations of this chapter, we omit the dependency on the vector of the changes in risk factors, simply denoting by L_i the generated losses, by Q_i the generated approximations (also stratified variables), by LR_i the likelihood ratio terms in corresponding importance sampling or importance sampling with original or special stratification based algorithms, by p_i the probabilities for the strata under the corresponding measure and by $\hat{p}_i$ their estimators from the corresponding algorithms. One should only take care about the probability measure under which the corresponding changes in risk factors are generated and under which the corresponding probabilities are calculated or estimated. We group the final estimators in the simplified notations, according to their types, in the following way:

Type 1: IS, ACE, FCE, AICE, FICE

$$\hat{\mathbb{P}}(L > x) = \frac{1}{N} \sum_{i=1}^{N} I_{(x,+\infty)}(L_i) LR_i. \tag{5.9}$$

Type 2: ISS, ACESI without z, AICESI without z, AOISI

$$\hat{\mathbb{P}}(L > x) = \sum_{i=1}^{K} p_i \frac{1}{N_i} \sum_{j=1}^{N_i} I_{(x,+\infty)}(L_{ij}) LR_{ij}. \tag{5.10}$$

Type 3: ACESI with z, AICESI with z, AOISSI

$$\hat{\mathbb{P}}(L > x) = p_1 \frac{1}{N} \sum_{i=1}^{N} I_{(x,+\infty)}(L_i) LR_i + p_2. \tag{5.11}$$

Type 4: ACES without z, AICES without z, AOIS

$$\hat{\mathbb{P}}(L > x) = \hat{p}_1 \frac{1}{N} \sum_{i=1}^{N} I_{(x,+\infty)}(L_i) LR_i. \tag{5.12}$$

Type 5: ACES with z, AICES with z, AOISS

$$\hat{\mathbb{P}}(L > x) = \hat{p}_1 \frac{1}{N} \sum_{i=1}^{N} I_{(x,+\infty)}(L_i) LR_i + \hat{p}_2. \tag{5.13}$$

We should note that the estimator (5.9) of type 1 is covered by the type 2. More precisely, the estimator (5.9) can be achieved from the estimator (5.10) by setting $K = 1$ and, hence, $p_1 = 1$. Therefore, we skip the consideration of the type 1 below. We have also separated the type 3 from the type 2 and the type 5 from the type 4, because the estimator for CVaR can not be achieved by the special stratification with the level z, but only with the level x (see Section 5.2.2).

The type 2 estimator, which covers the importance sampling, the importance sampling with stratification and the importance sampling with special stratification in ideal setting without upper level z is considered in Section 5.2.2. The type 3 estimator, which covers the importance sampling with special stratification in ideal setting is considered in Section 5.2.3. The type 4 estimator, which covers the importance sampling with special stratification without upper level z is considered in Section 5.2.4. Finally, the type 5 estimator, which covers the importance sampling with special stratification is considered in Section 5.2.5.

5.2.2 Importance Sampling with original Stratification

In this section we derive the estimators for VaR and CVaR for the variance reduction procedures according to the type 2 estimator from Section 5.2.1. In order to do that, we first have to modify the corresponding MC estimators (5.7) and (5.8):

$$\hat{v}_\alpha = \inf\{x : \sum_{i=1}^{K} p_i \frac{1}{N_i} \sum_{j=1}^{N_i} I_{(x,+\infty)}(L_{ij}) LR_{ij} \leq 1 - \alpha\}. \tag{5.14}$$

$$\hat{c}_\alpha = \hat{v}_\alpha + \frac{1}{1-\alpha} \sum_{i=1}^{K} p_i \frac{1}{N_i} \sum_{j=1}^{N_i} [L_{ij} - \hat{v}_\alpha]_+ LR_{ij}. \tag{5.15}$$

The estimator (5.14) for VaR can be found in, e.g., Chu (2010), Chu and Nakayama (2012), Glasserman et al. (2000b), whereas the estimator (5.15) for CVaR, in the case of the importance sampling estimator only (without the stratification), can be found in Hong and Liu (2011), Sun and Hong (2009, 2010).

Glasserman et al. (2000b) mention, that in order to perform the optimization in (5.14), one has to perform the variance reduction algorithm for different values x around the true, but unknown, value v_α of VaR. Since the set up of the corresponding change of measure and stratification depends on the level x, it has to be performed many times (for each value x separately). This is, however, inconvenient, as the number of such levels x may be large. Instead, the change of measure and the stratified parameters of the particular variance reduction procedure are constructed only for one value x in

the neighborhood of the true value v_α (this value can be either a rough estimate of VaR obtained, say, by the MC method, or some good guess, if any information about the true VaR is given). One should, therefore, take care in choosing this level x, such that the corresponding variance reduction procedure works for the estimation of the probability of high portfolio losses $\mathbb{P}(L > \tilde{x})$ for a range of values $\tilde{x}$.

Glasserman et al. (2000b) show the robustness of the IS algorithm when the exponential change of measure is constructed for the level x, but the probability of high portfolio losses $\mathbb{P}(L > \tilde{x})$ for a range of values $\tilde{x}$ is estimated. Besides, they mention that the variance reduction of the IS and ISS algorithms for the estimation of the probability of high portfolio losses $\mathbb{P}(L > x)$ for the level x in the neighborhood of the true value v_α of VaR carries over to VaR estimation.

In the case of CVaR estimation in the form (5.15), the same result holds. Either the same change of measure as for the estimation of VaR can be used to estimate CVaR, or the new change of measure for the level $x = \hat{v}_\alpha$ can be performed.

With regard to the IS, ACE, FCE, AICE, FICE and ISS algorithms, there is no restriction on the choice of the level x for which the change of measure is constructed, such that the probability of high portfolio losses $\mathbb{P}(L > \tilde{x})$ can be estimated for a range of values $\tilde{x}$. With regard to the algorithms that involve the special stratification either in the ideal or in the general setting, we have to take more care. The problem is, that the special stratification levels y and z from (4.21) depend on the level x and are chosen in a special way, such that (see Section 4.3.1):

$$Q(\mathbf{x}) \leq y \Rightarrow L(\mathbf{x}) \leq x, \text{ for all } \mathbf{x} \in \mathbb{R}_{\Delta\mathbf{S}},$$

and

$$Q(\mathbf{x}) \geq z \Rightarrow L(\mathbf{x}) \geq x, \text{ for all } \mathbf{x} \in \mathbb{R}_{\Delta\mathbf{S}}.$$

Hence, if we want to estimate the probability of high portfolio losses $\mathbb{P}(L > \tilde{x})$ for a range of values $\tilde{x}_1 \leq \tilde{x} \leq \tilde{x}_2$, the following has to hold for the levels of the special stratification $y = y(x)$ and $z = z(x)$, chosen in advance for some fixed value x in the same range $(\tilde{x}_1, \tilde{x}_2)$:

$$Q(\mathbf{x}) \leq y \Rightarrow L(\mathbf{x}) \leq \tilde{x}, \text{ for all } \mathbf{x} \in \mathbb{R}_{\Delta\mathbf{S}} \text{ and all } \tilde{x} \in (\tilde{x}_1, \tilde{x}_2),$$

and

$$Q(\mathbf{x}) \geq z \Rightarrow L(\mathbf{x}) \geq \tilde{x}, \text{ for all } \mathbf{x} \in \mathbb{R}_{\Delta\mathbf{S}} \text{ and all } \tilde{x} \in (\tilde{x}_1, \tilde{x}_2).$$

This does not hold, in particular, for $\tilde{x} < x$ in the first implication and for $\tilde{x} > x$ in the second implication.

Thanks to Remark 4.2, the special stratification can be defined via a wider interval than the interval $(y(x), z(x))$, such that the above implications hold true. In order to build such a wider interval, we first show the monotonicity of the levels $y(\tilde{x})$ and $z(\tilde{x})$ as functions of $\tilde{x}$. More precisely, for $\tilde{x}_1 \leq \tilde{x}$ the following holds true:

$$y(\tilde{x}_1) = \inf_{\mathbf{x}\in\mathbb{R}_{\Delta\mathbf{S}}} \{Q(\mathbf{x}) | L(\mathbf{x}) > \tilde{x}_1\} \leq \inf_{\mathbf{x}\in\mathbb{R}_{\Delta\mathbf{S}}} \{Q(\mathbf{x}) | L(\mathbf{x}) > \tilde{x}\} = y(\tilde{x}),$$

and for $\tilde{x} \leq \tilde{x}_2$ the following holds true:

$$z(\tilde{x}) = \sup_{\mathbf{x}\in\mathbb{R}_{\Delta\mathbf{S}}} \{Q(\mathbf{x}) | L(\mathbf{x}) < \tilde{x}\} \leq \sup_{\mathbf{x}\in\mathbb{R}_{\Delta\mathbf{S}}} \{Q(\mathbf{x}) | L(\mathbf{x}) < \tilde{x}_2\} = z(\tilde{x}_2).$$

Due to this monotonicity, for all $\tilde{x} \in (\tilde{x}_1, \tilde{x}_2)$, we have that $(y(\tilde{x}), z(\tilde{x})) \subset (y(\tilde{x}_1), z(\tilde{x}_2))$, and, hence, the following implications hold true:

$$Q(\mathbf{x}) \leq y(\tilde{x}_1) \Rightarrow L(\mathbf{x}) \leq \tilde{x}_1 < \tilde{x}, \text{ for all } \mathbf{x} \in \mathbb{R}_{\Delta\mathbf{S}} \text{ and all } \tilde{x} \in (\tilde{x}_1, \tilde{x}_2),$$

and

$$Q(\mathbf{x}) \geq z(\tilde{x}_2) \Rightarrow L(\mathbf{x}) \geq \tilde{x}_2 > \tilde{x}, \text{ for all } \mathbf{x} \in \mathbb{R}_{\Delta\mathbf{S}} \text{ and all } \tilde{x} \in (\tilde{x}_1, \tilde{x}_2).$$

All said above gives us the following procedure of estimating VaR with the help of the methods based on the special stratification:

1. Estimate roughly the neighborhood of the true VaR: $v_\alpha \in (\tilde{x}_1, \tilde{x}_2)$.
2. Find the importance sampling change of measure based on the level x being the middle of this interval, i.e. $x = (\tilde{x}_1 + \tilde{x}_2)/2$.
3. Perform the variance reduction procedure based on the importance sampling together with the special stratification defined by the levels $y(\tilde{x}_1)$ and $z(\tilde{x}_2)$. The corresponding strata partition is demonstrated in the Figure 5.1.

The estimation of CVaR is slightly different. First of all, assuming that VaR is already estimated, it does not involve the range of the levels at which the estimator is performed (no optimization is involved). Therefore, it is enough to perform the importance sampling together with any stratification for the level given by the estimate $\hat{v}_\alpha$ of VaR.

Another problem, however, appears while using the special stratification in the estimation of CVaR. We remind the reader, that the main advantage of the special stratification is that the function $I_{(x,\infty)}(L(\Delta\mathbf{S}))$ is a constant and known, if $Q(\Delta\mathbf{S}) \notin (y, z)$. In the case of CVaR, the corresponding indicator function is substituted by the function $[L(\Delta\mathbf{S}) - \hat{v}_\alpha]_+$, whose value is fixed and known only if $Q(\Delta\mathbf{S}) \leq y$ but not if $Q(\Delta\mathbf{S}) \geq z$.

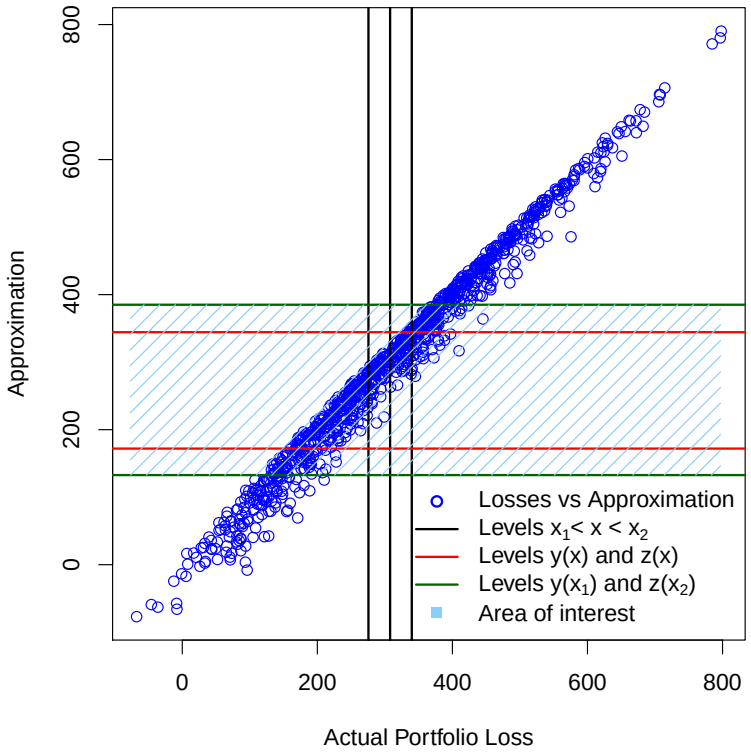


FIGURE 5.1: Wider range from $y(x_1)$ to $z(x_2)$ for some $x_1 < x < x_2$ for Delta-Gamma approximation against actual portfolio losses in 1000 randomly generated scenarios for a portfolio of short 10 ATM down-and-out calls, 10 uncorrelated assets, 0.1 years maturity, Gaussian changes in risk factors, probability nearly 1% (for details on the portfolio see Chapter 6).

Therefore, only the special stratification without the level z should be used in order to estimate CVaR. This is the reason, why we have separated ACESI, AICESI, ACES and AICES algorithms for the cases with the special stratification based on both levels y and z, and the special stratification based only on the level y. The type 2 estimator (5.10), in particular, can be used in order to estimate CVaR.

5.2.3 Importance Sampling with special Stratification in Ideal setting

In this section we are interested in the estimators for VaR and CVaR for the variance reduction procedures according to the type 3 estimator from Section 5.2.1. As mentioned in Section 5.2.2, CVaR can not be estimated for such variance reduction algorithms, because of the level z in the special stratification. With regard to VaR, we first have to modify the corresponding MC estimator (5.7):

$$\hat{v}_\alpha = \inf\{x : p_1 \frac{1}{N} \sum_{i=1}^{N} I_{(x,+\infty)}(L_i) LR_i + p_2 \leq 1 - \alpha\}. \tag{5.16}$$

The estimator (5.16) for VaR can be found in Pupashenko (2014) in a slightly modified form. The whole theory about the choice of the importance sampling change of measure and the levels of the special stratification is the same as in Section 5.2.2.

5.2.4 Importance Sampling with special Stratification without upper level

In this section we derive the estimators for VaR and CVaR for the variance reduction procedures according to the type 4 estimator from Section 5.2.1. This is achieved by the simple modification of the corresponding estimators (5.14) and (5.15) from Section 5.2.2:

$$\hat{v}_\alpha = \inf\{x : \hat{p}_1 \frac{1}{N} \sum_{i=1}^{N} I_{(x,+\infty)}(L_i) LR_i \leq 1 - \alpha\}. \tag{5.17}$$

$$\hat{c}_\alpha = \hat{v}_\alpha + \frac{1}{1-\alpha} \hat{p}_1 \frac{1}{N} \sum_{i=1}^{N} [L_i - \hat{v}_\alpha]_+ LR_i. \tag{5.18}$$

The whole theory about the choice of the importance sampling change of measure and the levels of the special stratification is the same as in Section 5.2.2.

5.2.5 Importance Sampling with special Stratification

In this section we derive the estimator for VaR for the variance reduction procedures according to the type 5 estimator from Section 5.2.1. As mentioned in Section 5.2.2, CVaR can not be estimated for such variance reduction algorithms because of the level z in the special stratification. In order to estimate VaR, we have to modify the corresponding estimator (5.16) from Section 5.2.3:

$$\hat{v}_\alpha = \inf\{x : \hat{p}_1 \frac{1}{N} \sum_{i=1}^{N} I_{(x,+\infty)}(L_i) LR_i + \hat{p}_2 \leq 1 - \alpha\}. \tag{5.19}$$

The whole theory about the choice of the importance sampling change of measure and the levels of the special stratification is the same as in Section 5.2.2.

5.3 Estimator analysis

After we have defined the variance reduction estimators for VaR and CVaR in Section 5.2, we should analyze their asymptotics. The most striking among them are the consistency (convergence of the estimator to its true value) and the asymptotic normality

(the limiting distribution of the difference of the estimator and its true value is normally distributed), for definitions see Billingsley (1995), Serfling (1980). In particular, the asymptotic normality can be used in order to construct the confidence intervals, which we do in Section 5.3.3.

The literature gives an intensive study of the asymptotics of MC and different variance reduction procedures for VaR, as well as for CVaR. Among them we would like to mention Chu (2010), Chu and Nakayama (2012), Glasserman (2004), Glasserman et al. (2000b, 2002), Hong and Liu (2011), Sun and Hong (2009, 2010). None of these literature, however, gives a complete study of the problem for both VaR and CVaR. Besides, they, of course, do not include the new variance reduction algorithms derived in Chapter 4. Therefore, we give here the consecutive study of the corresponding asymptotics of the VaR and CVaR estimators of all types of the variance reduction procedures given in Section 5.2.1.

5.3.1 Asymptotics of Monte Carlo simulation

In this section we give the asymptotic results for the MC estimators of VaR and CVaR. The consistency and the asymptotic normality of the MC estimators (5.7) and (5.8) have been studied in, e.g., Glasserman (2004), Serfling (1980), Trindade et al. (2007). But we present here a more general result based on the Bahadur (Bahadur (1966)) and the Bahadur-Ghosh (Ghosh (1971)) representations of the quantile, studied in Chu (2010), Chu and Nakayama (2012), Hong and Liu (2011), Sun and Hong (2009, 2010).

We first start from the asymptotics for VaR given by (5.7), which is derived in Chu (2010), Chu and Nakayama (2012). We further assume the following:

Assumption 5.6. The cumulative distribution function $\mathbb{F}$ is differentiable at v_α and $f(v_\alpha) > 0$, where f denotes in this section the density function of the portfolio losses L (do not confuse the notation with the density of the changes in risk factors, where we usually make an explicit dependency on the parameter).

The following assumption is crucial in the proof of the Bahadur representation (see Bahadur (1966)):

Assumption 5.7. The second derivative of the cumulative distribution function $\mathbb{F}$ is bounded in the neighborhood of v_α.

VaR is a quantile estimator, for which, under Assumptions 5.6 and 5.7, the following Bahadur representation holds true (see Bahadur (1966)):

$$\hat{v}_\alpha = v_\alpha - \frac{(1-\alpha) - \frac{1}{N}\sum_{i=1}^{N} I_{(v_\alpha,+\infty)}(L_i)}{f(v_\alpha)} + A'_N,$$

where almost surely (a.s.):

$$A'_N = O(N^{-3/4}(\log N)^{1/2}(\log\log N)^{1/4}) \text{ as } N \to \infty,$$

where $A'_N = O(g(N))$ a.s. means, that there exists a set of probability 1, such that for each ω from this set there exits a constant $c(\omega)$ such that $|A'_N| \leq c(\omega)g(N)$ for N sufficiently large (see Billingsley (1995), Serfling (1980)). Kiefer (1967) improves the rate of the convergence of A'_N even further by proving a.s. the following:

$$A'_N = O(N^{-3/4}(\log\log N)^{3/4}) \text{ as } N \to \infty.$$

This asymptotic representation of the MC estimator for VaR can be found in Chu (2010), Chu and Nakayama (2012), Hong and Liu (2011), Sun and Hong (2009, 2010), but only Chu (2010), Chu and Nakayama (2012) develop a weaker form of the Bahadur representation, namely the Bahadur-Ghosh representation (see Ghosh (1971)). In more details, for some sequence $\alpha_N = \alpha - O(N^{-1/2})$, under only Assumption 5.6, the Bahadur-Ghosh representation for a quantile looks like:

$$\hat{v}_{\alpha_N} = v_\alpha - \frac{(1-\alpha_N) - \frac{1}{N}\sum_{i=1}^{N} I_{(v_\alpha,+\infty)}(L_i)}{f(v_\alpha)} + A_N, \tag{5.20}$$

where A_N satisfies:

$$\sqrt{N}A_N \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty,$$

where $\xrightarrow{\mathbb{P}}$ denotes the convergence in probability (see Billingsley (1995), Serfling (1980)).

The Bahadur and the Bahadur-Ghosh representations of the MC estimator (5.7) of VaR give a speed at which this estimator converges to the true value. Besides, as proven in Chu (2010), Chu and Nakayama (2012), they have two important consequences, the consistency (by the Law of Large Numbers, see e.g. Billingsley (1995), Serfling (1980)):

$$\hat{v}_\alpha \xrightarrow{\text{a.s.}} v_\alpha \text{ as } N \to \infty,$$

$$\hat{v}_{\alpha_N} \xrightarrow{\mathbb{P}} v_\alpha \text{ as } N \to \infty, \tag{5.21}$$

where $\xrightarrow{\text{a.s.}}$ means the a.s. convergence or the convergence with probability 1 (see Billingsley

(1995), Serfling (1980)), and the asymptotic normality (by Slutsky's lemma and the Central Limit Theorem, see e.g. Billingsley (1995), Serfling (1980)):

$$\sqrt{N}(\hat{v}_\alpha - v_\alpha) \xrightarrow{\mathbb{D}} N_1\left(0, \frac{\mathbb{V}\mathbf{ar}[I_{(v_\alpha,+\infty)}(L)]}{f^2(v_\alpha)}\right) \text{ as } N \to \infty,$$

$$\sqrt{N}(\hat{v}_{\alpha_N} - v_\alpha) \xrightarrow{\mathbb{D}} N_1\left(0, \frac{\mathbb{V}\mathbf{ar}[I_{(v_\alpha,+\infty)}(L)]}{f^2(v_\alpha)}\right) \text{ as } N \to \infty, \tag{5.22}$$

where $\xrightarrow{\mathbb{D}}$ denotes the convergence in distribution (see Billingsley (1995), Serfling (1980)).

There are various ways to prove the strong consistency and the asymptotic normality of the MC estimator (5.7) of VaR, presented in, e.g., Glasserman (2004), Serfling (1980), Trindade et al. (2007), but we find the most advanced one the one presented here, given by (5.20), (5.21) and (5.22). As mentioned in Chu (2010), Chu and Nakayama (2012), the main advantage of the Bahadur-Ghosh representation (5.20) is that it further simplifies the derivation of the confidence intervals (see Section 5.3.3).

The main difference between the Bahadur and the Bahadur-Ghosh representations is that the result from Ghosh (1971) requires weaker assumptions, but the result from Bahadur (1966) gives an additional information about the exact rate of convergence of the last term A'_N.

We further continue with the same analysis of the MC estimator (5.8) of CVaR, developed in Hong and Liu (2011), Sun and Hong (2009, 2010) with the help of only the Bahadur representation. In this case, the authors do not face problems as for VaR, therefore they do not need to apply the weaker Bahadur-Ghosh representation. Under the same Assumptions 5.6 and 5.7, the following Bahadur representation for the MC estimator (5.8) of CVaR holds true:

$$\hat{c}_\alpha = c_\alpha + \left(\left[v_\alpha + \frac{1}{1-\alpha}\frac{1}{N}\sum_{i=1}^{N}[L_i - v_\alpha]_+\right] - c_\alpha\right) + B_N, \tag{5.23}$$

where B_N a.s. satisfies:

$$B_N = O(N^{-1}\log N) \text{ as } N \to \infty.$$

The corresponding strong consistency and the asymptotic normality hold true:

$$\hat{c}_\alpha \xrightarrow{\text{a.s.}} c_\alpha \text{ as } N \to \infty, \tag{5.24}$$

$$\sqrt{N}(\hat{c}_\alpha - c_\alpha) \xrightarrow{\mathbb{D}} N_1\left(0, \frac{\mathbb{V}\mathbf{ar}[[L - v_\alpha]_+]}{(1-\alpha)^2}\right) \text{ as } N \to \infty. \tag{5.25}$$

5.3.2 Asymptotics of Variance Reduction procedures

In this section we extend the asymptotic theory to the variance reduction procedures from Chapter 4 for VaR, according to Chu (2010), Chu and Nakayama (2012), and for CVaR, according to Hong and Liu (2011), Sun and Hong (2009, 2010).

Again, we start from the VaR estimation, for which all four types 2-5 of the variance reduction procedures (see Section 5.2.1) can be applied. The type 2 estimator covers the importance sampling and the importance sampling with stratification variance reduction algorithms, for which the Bahadur-Ghosh representation is proven in Chu (2010), Chu and Nakayama (2012) under Assumptions 5.6 and 5.8:

Assumption 5.8. For each stratum i assume, that there exist $\epsilon > 0$ and $\delta > 0$ such that the expectation of $I_{(v_\alpha-\delta,\infty)}(L_{ij})LR_{ij}^{2+\epsilon}$ under the importance sampling change of measure is finite.

The Bahadur-Ghosh representation for the importance sampling with original stratification estimator (5.14) of VaR looks as follows:

$$\hat{v}_{\alpha_N} = v_\alpha - \frac{(1-\alpha_N) - \sum_{i=1}^{K} p_i \frac{1}{N_i} \sum_{j=1}^{N_i} I_{(v_\alpha,+\infty)}(L_{ij})LR_{ij}}{f(v_\alpha)} + A_N, \tag{5.26}$$

where A_N satisfies:

$$\sqrt{N}A_N \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty.$$

The corresponding consistency of the estimator and the asymptotic normality read as follows:

$$\hat{v}_{\alpha_N} \xrightarrow{\mathbb{P}} v_\alpha \text{ as } N \to \infty, \tag{5.27}$$

$$\sqrt{N}(\hat{v}_{\alpha_N} - v_\alpha) \xrightarrow{\mathbb{D}} N_1\left(0, \frac{\sum_{i=1}^{K} p_i^2 \frac{N}{N_i} \mathbb{V}\mathbf{ar}[I_{(v_\alpha,+\infty)}(L_i)LR_i]}{f^2(v_\alpha)}\right) \text{ as } N \to \infty. \tag{5.28}$$

Note that, since the partition in the stratification part is fixed, the ratios $\frac{N}{N_i}$ can be treated as constants, which sum up to 1.

These results can be extended to the type 3 variance reduction procedure in the straightforward way. In particular, the Bahadur-Ghosh representation for the importance sampling with special stratification in ideal setting estimator (5.16) of VaR takes the form:

$$\hat{v}_{\alpha_N} = v_\alpha - \frac{(1-\alpha_N) - p_1 \frac{1}{N} \sum_{j=1}^{N} I_{(v_\alpha,+\infty)}(L_i)LR_i - p_2}{f(v_\alpha)} + A_N, \tag{5.29}$$

where A_N satisfies:

$$\sqrt{N}A_N \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty.$$

The corresponding consistency of the estimator and the asymptotic normality read as follows:

$$\hat{v}_{\alpha_N} \xrightarrow{\mathbb{P}} v_\alpha \text{ as } N \to \infty, \tag{5.30}$$

$$\sqrt{N}(\hat{v}_{\alpha_N} - v_\alpha) \xrightarrow{\mathbb{D}} N_1\left(0, \frac{p_1^2 \mathbb{V}\mathbf{ar}[I_{(v_\alpha,+\infty)}(L)LR]}{f^2(v_\alpha)}\right) \text{ as } N \to \infty. \tag{5.31}$$

Now we want to extend the above stated results for the type 2 estimator to the type 4 estimator, and the result for the type 3 estimator to the type 5 estimator. In both cases, the difference in the estimators lies in the use of the estimators $\hat{p}_1$ and $\hat{p}_2$ instead of the fixed values p_1 and p_2. Moreover, according to the algorithms from Chapter 4, the MC algorithm is used for the estimation of the probability $\hat{p}_1 = \mathbb{P}_{\hat{\mathbf{v}}^*}(Q \in (y, z))$ under the corresponding change of measure:

$$\frac{1}{m_1 \cdot N} \sum_{i=1}^{m_1 \cdot N} I_{(y,z)}(Q_i),$$

and the CE algorithm is used for estimation of the probability $\hat{p}_2 = \mathbb{P}_{\mathbf{u}}(Q \geq z)$ under the original measure:

$$\frac{1}{m_2 \cdot N} \sum_{i=1}^{m_2 \cdot N} I_{(z,\infty)}(Q_i)LR_i,$$

where y and z are the levels of the special stratification.

In order to do the proofs similar to Sun and Hong (2010), one has to state an assumption on the estimators $\hat{p}_1$ and $\hat{p}_2$ similar to Assumption 5.8:

Assumption 5.9. Assume, that there exist $\epsilon > 0$ and $\delta > 0$ such that the expectation of $I_{(y-\delta,z+\delta)}(Q_i)$ and the expectation of $I_{(z-\delta,\infty)}(Q_i)LR_i^{2+\epsilon}$, both under the importance sampling change of measure, are finite.

The corresponding asymptotic result for the type 2 estimator can then be easily extended to the type 4 estimator, and the type 3 to the type 5 estimator for the worst case of $m_1 = 1$ and $m_2 = 1$ (otherwise the sample sizes in estimators $\hat{p}_1$ and $\hat{p}_2$ are even larger than N).

More precisely, the Bahadur-Ghosh representation for the importance sampling with special stratification without upper level estimator (5.17) of VaR, under Assumptions 5.6, 5.8 and 5.9, takes the form:

$$\hat{v}_{\alpha_N} = v_\alpha - \frac{(1-\alpha_N) - \hat{p}_1 \frac{1}{N} \sum_{i=1}^{N} I_{(v_\alpha,+\infty)}(L_i)LR_i}{f(v_\alpha)} + A_N, \tag{5.32}$$

where A_N satisfies:

$$\sqrt{N}A_N \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty.$$

The corresponding consistency of the estimator and the asymptotic normality read as follows:

$$\hat{v}_{\alpha_N} \xrightarrow{\mathbb{P}} v_\alpha \text{ as } N \to \infty, \tag{5.33}$$

$$\sqrt{N}(\hat{v}_{\alpha_N} - v_\alpha) \xrightarrow{\mathbb{D}} N_1\left(0, \frac{\mathbb{V}\mathbf{ar}[\hat{p}_1 I_{(v_\alpha,+\infty)}(L)LR]}{f^2(v_\alpha)}\right) \text{ as } N \to \infty. \tag{5.34}$$

In order to show the Bahadur-Ghosh representation (5.32) from (5.26), it suffices to prove that:

$$\sqrt{N}\frac{(p_1 - \hat{p}_1)\frac{1}{N}\sum_{i=1}^{N} I_{(v_\alpha,+\infty)}(L_i)LR_i}{f(v_\alpha)} \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty,$$

which can be easily done, under Assumptions 5.6, 5.8 and 5.9, in a similar way to the proof of Lemma 2 in Sun and Hong (2010).

Similarly, the Bahadur-Ghosh representation for the importance sampling with special stratification estimator (5.19) of VaR, under Assumptions 5.6, 5.8 and 5.9, can be easily achieved from (5.29) by showing the following:

$$\sqrt{N}\frac{(p_1 - \hat{p}_1)\frac{1}{N}\sum_{i=1}^{N} I_{(v_\alpha,+\infty)}(L_i)LR_i + (p_2 - \hat{p}_2)}{f(v_\alpha)} \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty,$$

again in a straightforward way similarly to the proof of Lemma 2 in Sun and Hong (2010).

The Bahadur-Ghosh representation then takes the form:

$$\hat{v}_{\alpha_N} = v_\alpha - \frac{(1 - \alpha_N) - \hat{p}_1\frac{1}{N}\sum_{j=1}^{N} I_{(v_\alpha,+\infty)}(L_i)LR_i - \hat{p}_2}{f(v_\alpha)} + A_N, \tag{5.35}$$

where A_N satisfies:

$$\sqrt{N}A_N \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty.$$

The corresponding consistency of the estimator and the asymptotic normality read as follows:

$$\hat{v}_{\alpha_N} \xrightarrow{\mathbb{P}} v_\alpha \text{ as } N \to \infty, \tag{5.36}$$

$$\sqrt{N}(\hat{v}_{\alpha_N} - v_\alpha) \xrightarrow{\mathbb{D}} N_1\left(0, \frac{\mathbb{V}\mathbf{ar}[\hat{p}_1 I_{(v_\alpha,+\infty)}(L)LR + \hat{p}_2]}{f^2(v_\alpha)}\right) \text{ as } N \to \infty. \tag{5.37}$$

In the next step we move to the asymptotic theory of the estimation of CVaR. We remind the reader, that only the variance reduction methods of the types 1, 2 and 4 can be applied in the estimation of CVaR. We start here from the type 1 estimator of the variance reduction procedures, that covers the importance sampling algorithm,

for which the Bahadur representation is proven in Hong and Liu (2011), Sun and Hong (2009, 2010) under Assumptions 5.6, 5.7, 5.8 and 5.10:

Assumption 5.10. The likelihood ratio is bounded in the neighborhood of v_α.

Note, that in the case of the importance sampling algorithm without stratification, the number of strata in Assumptions 5.8 and 5.10 is simply taken to be 1.

The importance sampling estimator (5.15) of CVaR (importance sampling with stratification estimator with one stratum) allows the following Bahadur representation (see Hong and Liu (2011), Sun and Hong (2009, 2010)):

$$\hat{c}_\alpha = c_\alpha + \left(\left[v_\alpha + \frac{1}{1-\alpha} \frac{1}{N} \sum_{i=1}^{N} [L_i - v_\alpha]_+ LR_i \right] - c_\alpha \right) + B_N, \tag{5.38}$$

where B_N satisfies:

$$\sqrt{N} B_N \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty.$$

The corresponding consistency of the estimator and asymptotic normality read as follows:

$$\hat{c}_\alpha \xrightarrow{\mathbb{P}} c_\alpha \text{ as } N \to \infty, \tag{5.39}$$

$$\sqrt{N}(\hat{c}_\alpha - c_\alpha) \xrightarrow{\mathbb{D}} N_1 \left(0, \frac{\mathbb{V}\mathbf{ar}[[L - v_\alpha]_+ LR]}{(1-\alpha)^2} \right) \text{ as } N \to \infty. \tag{5.40}$$

Analogously to the asymptotic theory for VaR, the Bahadur representation (5.38), the consistency (5.39) and the asymptotic normality (5.40) of the importance sampling estimator (5.15) of CVaR can be easily extended to the importance sampling with stratification estimator (5.15).

More precisely, the Bahadur representation for the importance sampling with stratification estimator (5.15) of CVaR takes the form:

$$\hat{c}_\alpha = c_\alpha + \left(\left[v_\alpha + \frac{1}{1-\alpha} \sum_{i=1}^{K} p_i \frac{1}{N_i} \sum_{j=1}^{N_i} [L_{ij} - v_\alpha]_+ LR_{ij} \right] - c_\alpha \right) + B_N, \tag{5.41}$$

where B_N satisfies:

$$\sqrt{N} B_N \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty.$$

The corresponding consistency of the estimator and asymptotic normality read as follows:

$$\hat{c}_\alpha \xrightarrow{\mathbb{P}} c_\alpha \text{ as } N \to \infty, \tag{5.42}$$

$$\sqrt{N}(\hat{c}_\alpha - c_\alpha) \xrightarrow{\mathbb{D}} N_1 \left(0, \frac{\sum_{i=1}^{K} p_i^2 \frac{N}{N_i} \mathbb{V}\mathbf{ar}[[L_i - v_\alpha]_+ LR_i]}{(1-\alpha)^2} \right) \text{ as } N \to \infty. \tag{5.43}$$

Again, since the partition in the stratification part is fixed, the ratios $\frac{N}{N_i}$ can be treated as constants, which sum up to 1.

Finally, the Bahadur representation for the importance sampling with special stratification without upper level estimator (5.18) of CVaR can be achieved from (5.41) under Assumptions 5.6, 5.7, 5.8 and 5.10 in a similar way to VaR, i.e. by showing that:

$$\sqrt{N}\frac{1}{1-\alpha}(p_1-\hat{p}_1)\frac{1}{N}\sum_{i=1}^{N}[L_i-v_\alpha]_+LR_i \xrightarrow{\mathbb{P}} 0 \text{ as } N\to\infty,$$

in a similar way to the proof of Lemma 2 in Sun and Hong (2010).

The Bahadur representation then takes the form:

$$\hat{c}_\alpha = c_\alpha + \left(\left[v_\alpha + \frac{1}{1-\alpha}\hat{p}_1\frac{1}{N}\sum_{i=1}^{N}[L_i-v_\alpha]_+LR_i\right]-c_\alpha\right)+B_N, \tag{5.44}$$

where B_N satisfies:

$$\sqrt{N}B_N \xrightarrow{\mathbb{P}} 0 \text{ as } N\to\infty.$$

The corresponding consistency of the estimator and asymptotic normality read as follows:

$$\hat{c}_\alpha \xrightarrow{\mathbb{P}} c_\alpha \text{ as } N\to\infty, \tag{5.45}$$

$$\sqrt{N}(\hat{c}_\alpha - c_\alpha) \xrightarrow{\mathbb{D}} N_1\left(0, \frac{\mathbb{V}\mathbf{ar}[\hat{p}_1[L-v_\alpha]_+LR]}{(1-\alpha)^2}\right) \text{ as } N\to\infty. \tag{5.46}$$

In order to summarize, the Bahadur-Ghosh and Bahadur representations, the consistency and the asymptotic normality of the estimators from Chapter 4, combined by the types from Section 5.2.1, are given by:

Type 1 and type 2: (5.26), (5.27) and (5.28) for the estimator (5.14) of VaR and (5.41), (5.42) and (5.43) for the estimator (5.15) of CVaR

Type 3 (5.29), (5.30) and (5.31) for the estimator (5.16) of VaR

Type 4 (5.32), (5.33) and (5.34) for the estimator (5.17) of VaR and (5.44), (5.45) and (5.46) for the estimator (5.18) of CVaR

Type 5 (5.35), (5.36) and (5.37) for the estimator (5.19) of VaR

5.3.3 Confidence intervals

In this section we construct the confidence intervals for all estimators of VaR and CVaR. The asymptotic normality that results from Sections 5.3.1 and 5.3.2 help us to do this

in the standard way. More precisely, assume that some estimator $\hat{\theta}$ of the unknown parameter θ, achieved with the sample size N, is asymptotically normal with the variance σ^2:

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{\mathbb{D}} N_1(0, \sigma^2).$$

Then, the β confidence interval (usually β is close to 1) for the large sample size N is given by the following endpoints:

$$\hat{\theta} \pm z_{(1-\beta)/2} \frac{\sigma}{\sqrt{N}}, \tag{5.47}$$

where $1 - \Phi(z_{(1-\beta)/2}) = (1-\beta)/2$ with $\Phi(\cdot)$ being the cumulative distribution function of the standard normal random variable $N_1(0,1)$.

Usually, in practice, we do not know the variance σ^2. But, if we substitute σ in the confidence interval (5.47) by some consistent estimator $\hat{\sigma}$, the corresponding confidence interval remains asymptotically valid:

$$\hat{\theta} \pm z_{(1-\beta)/2} \frac{\hat{\sigma}}{\sqrt{N}}. \tag{5.48}$$

Hence, in order to construct the confidence intervals for the VaR and CVaR estimators from Section 5.2, one has to construct consistent estimators for the variances in the corresponding asymptotic normality results of Sections 5.3.1 and 5.3.2.

We further summarize the involved variances of the asymptotic normality results for MC and the variance reduction algorithms of VaR and CVaR estimators, which we again group by the type of the estimator (see Section 5.2.1):

MC:

$$\text{VaR variance from (5.22): } \frac{\mathbb{V}\mathbf{ar}[I_{(v_\alpha,+\infty)}(L)]}{f^2(v_\alpha)},$$

$$\text{and CVaR variance from (5.25): } \frac{\mathbb{V}\mathbf{ar}[[L - v_\alpha]_+]}{(1-\alpha)^2}.$$

Type 1 and type 2:

$$\text{VaR variance from (5.28): } \frac{\sum_{i=1}^K p_i^2 \frac{N}{N_i} \mathbb{V}\mathbf{ar}[I_{(v_\alpha,+\infty)}(L_i) LR_i]}{f^2(v_\alpha)},$$

$$\text{and CVaR variance from (5.43): } \frac{\sum_{i=1}^K p_i^2 \frac{N}{N_i} \mathbb{V}\mathbf{ar}[[L_i - v_\alpha]_+ LR_i]}{(1-\alpha)^2}.$$

Type 3:

$$\text{VaR variance from (5.31): } \frac{p_1^2 \mathbb{V}\mathbf{ar}[I_{(v_\alpha,+\infty)}(L) LR]}{f^2(v_\alpha)}.$$

Type 4:

$$\text{VaR variance from (5.34): } \frac{\mathbb{V}\mathbf{ar}[\hat{p}_1 I_{(v_\alpha,+\infty)}(L)LR]}{f^2(v_\alpha)},$$

$$\text{and CVaR variance from (5.46): } \frac{\mathbb{V}\mathbf{ar}[\hat{p}_1[L-v_\alpha]_+ LR]}{(1-\alpha)^2}.$$

Type 5:

$$\text{VaR variance from (5.37): } \frac{\mathbb{V}\mathbf{ar}[\hat{p}_1 I_{(v_\alpha,+\infty)}(L)LR + \hat{p}_2]}{f^2(v_\alpha)}.$$

The role of $\hat{\theta}$ in the confidence intervals for VaR is played by $\hat{v}_{\alpha_N}$ with $\alpha_N = \alpha - O(N^{-1/2})$, and in the confidence intervals for CVaR by $\hat{c}_\alpha$.

The variances for CVaR, as well as the numerators of the variances for VaR, can be easily estimated by the corresponding sample counterparts in parallel with the estimator itself. The problem appears with the denominators of the variances for VaR, which contain the value of the density function f at VaR. As mentioned in, e.g., Glasserman (2004), the estimation of this function is complicated, and another approach for constructing the confidence intervals is proposed. In more details, one can divide the whole sample into smaller subsamples (batches), compute the VaR estimate from each batch, and form a confidence interval based on the sample variance of the estimates across these batches. This method has, however, a crucial drawback. In order to achieve an accurate confidence interval, a large sample size is often required.

The authors in Chu (2010), Chu and Nakayama (2012), construct the consistent estimator for the corresponding density function. In particular, the Bahadur-Ghosh representation of VaR with the perturbed parameter $\alpha_N = \alpha - O(N^{-1/2})$ is a key step in developing this estimator. This is the reason, why it is important to use the Bahadur-Ghosh representation for VaR, but for CVaR, the Bahadur representation suffices.

Further, we state the results from Chu (2010), Chu and Nakayama (2012), with regard to the estimation procedure of $f(v_\alpha)$. First note, that:

$$\frac{1}{f(v_\alpha)} = \frac{d}{dp}\mathbb{F}^{-1}(p)\Big|_{p=v_\alpha}.$$

Hence, the authors in Chu (2010), Chu and Nakayama (2012), suggest the finite-difference estimators (see Glasserman (2004)) for $\frac{1}{f(v_\alpha)}$ and prove their consistency using the Bahadur-Ghosh representation:

$$\frac{\sqrt{N}}{c}(\hat{v}_\alpha - \hat{v}_{\alpha_N}),$$

$$\frac{\sqrt{N}}{2c}(\hat{v}_{\alpha'_N} - \hat{v}_{\alpha_N}),$$

$$\sum_{j=1}^{r} w_j \frac{\sqrt{N}}{c_j}(\hat{v}_\alpha - \hat{v}_{\alpha_N(j)}),$$

$$\sum_{j=1}^{r} w_j \frac{\sqrt{N}}{2c_j}(\hat{v}_{\alpha'_N(j)} - \hat{v}_{\alpha_N(j)}),$$

where $\alpha_N = \alpha - \frac{c}{\sqrt{N}}$ and $\alpha'_N = \alpha + \frac{c}{\sqrt{N}}$ for some constant $c > 0$, and $\alpha_N(j) = \alpha - \frac{c_j}{\sqrt{N}}$ and $\alpha'_N(j) = \alpha + \frac{c_j}{\sqrt{N}}$ for some constants $c_j > 0$ and some weights w_j such that $\sum_{j=1}^{r} w_j = 1$ for some integer r.

Each of these estimators can be used in order to construct the asymptotically valid confidence intervals.

5.4 Summary

In this section we want to summarize the process of transition from the estimation of the probability of high portfolio losses to the estimation of VaR and CVaR.

First of all, it is important to remind the reader, that all estimators of the probability of high portfolio losses, i.e. the probability that losses exceed some known and fixed level x, are based on the change of measure and the choice of the strata that depend on this level x. But, in order to estimate VaR and then CVaR, it is required to estimate the probability of high portfolio losses in some neighborhood of the true VaR v_α. In order to do it efficiently, the change of measure and the strata are constructed only once for some fixed level x in the middle of the corresponding neighborhood. As mentioned in Glasserman (2004), Glasserman et al. (2000a,b, 2002), the variance reduction under this fixed change of measure of the estimation of the probability $\mathbb{P}(L > x)$ is carried over to the estimation of the probability $\mathbb{P}(L > y)$ for all y in the neighborhood of x. Therefore, we can expect, that the variance reduction rates of the estimation of the probability $\mathbb{P}(L > x)$ are approximately the same as the variance reduction rates of the estimation of the probability $\mathbb{P}(L > y)$ for all y in the neighborhood of x.

Now, looking at the variances of the corresponding estimators of VaR and CVaR (see Section 5.3.3) for all variance reduction algorithms, one can conclude, that they are proportional to the variances of the corresponding estimators of the probability of high portfolio losses $\mathbb{P}(L > v_\alpha)$. This implies, that the variance reduction rate of estimating the probability of high portfolio losses $\mathbb{P}(L > v_\alpha)$ is carried over to the estimation of the market risk measures VaR and CVaR. Besides, according to the above said, this variance reduction rate is approximately the same as the one achieved in the estimation of the probability $\mathbb{P}(L > x)$, where the level x is the one for which the change of measure

and the strata are constructed. A similar result is discussed in e.g. Glasserman (2004), Glasserman et al. (2000a,b, 2002).

Chapter 6

Numerical examples

In this chapter we perform an in-depth numerical study in order to examine our new variance reduction algorithms on different realistic portfolios, first in the light-tailed case in Section 6.1 and then in the heavy-tailed case in Section 6.2. In order to compare our new algorithms with the IS and ISS algorithms given by Glasserman (2004), Glasserman et al. (2000a,b, 2002), we use the same set of portfolios as tested in, e.g., Glasserman et al. (2000b). Besides, these portfolios are constructed in such a way, that they cover a wide variety of different characteristics. This is done in order to examine the effect of the portfolio on the performance of the particular variance reduction algorithms.

We first should note that the numerical study is done directly to the problem of estimating the probability of high portfolio losses. The first reason for this is that this is exactly the quantity tested by Glasserman (2004), Glasserman et al. (2000a,b, 2002). Besides, none of these papers cover the VaR and CVaR estimation for both IS and ISS algorithms simultaneously. The second reason is that the variance reduction rate of estimating the probability of high portfolio losses is carried over to the estimation of VaR and CVaR, as discussed in Section 5.4. Besides, the process of transition from the estimation of the probability of high portfolio losses to the estimation of VaR and CVaR is well described in Chapter 5 and is easy to implement, given that the variance reduction algorithms from Chapter 4 are already implemented.

We should also note, that all portfolios tested in the following are quite simple compared to the portfolios on real financial markets. But, as already mentioned in Glasserman et al. (2002), the simple portfolios give a good indication of the variance reduction rate, even if more complex portfolio revaluation procedures were used. This is because the greatest computational gain from the variance reduction algorithms is obtained in cases,

where the revaluation is extremely time consuming (see Assumption 4.1), e.g. for finite-difference methods, trees or even simulations.

For all new algorithms which are based on some approximation, except the classical Delta-Gamma approximation, we often use the modified Delta-Gamma, Charm-Speed and the modified Charm-Speed approximation (see Chapter 3). By improved approximation we always mean one of these four approximations, which gives the best performance. The user is, however, not restricted to these four approximations and can use any other approximation, if he or she believes that it fits the portfolio loss function better.

6.1 Light - tailed case

In this section we examine the light-tailed case, i.e. the case when the underlying changes in risk factors are assumed to have a multivariate normal distribution, see Assumption 3.1.

6.1.1 Setup of the example

We assume 250 trading days in a year and a continuously compounded annual risk free interest rate of 5%. Both the real and the risk-neutral rates of return are assumed to be equal. For each case, losses over 10 days are investigated, i.e. $\Delta t = 0.04$. All assets have initial values of 100 and are modeled as geometric Brownian motions with an annual volatility of $\sigma = 0.3$. In order to make it consistent with Assumption 3.1, the standard deviation of the changes in risk factors $\Delta \mathbf{S}_j$ should be given by $\mathbf{S}_j \sigma \sqrt{\Delta t}$, which gives the covariance matrix $\mathbf{\Sigma}_S$.

The following portfolios are considered:

(a.1) Short 10 at-the-money (ATM) calls and 5 ATM puts on each of the 10 uncorrelated assets, all options having a 0.1 years maturity.

(a.2) Same as (a.1) but with the maturity being half a year.

(a.3) Long 10 ATM calls and 5 ATM puts on each of 10 uncorrelated assets, all options having a 0.1 years maturity.

(a.4) Same as (a.3) but with the maturity being half a year.

(a.5) Same as (a.1) but with the number of puts increased so that $\mathbf{b} = \mathbf{0}$ in the Delta-Gamma approximation (see Section 3.3.1).

(a.6) Short 10 ATM calls and 10 ATM puts on each of 100 uncorrelated assets, all options having a 0.1 years maturity.

(a.7) Same as (a.6) but with the correlation of distinct assets set to 0.2.

(b.1) Short 10 ATM calls on each of the 10 uncorrelated assets, all options having a 0.1 years maturity.

(b.2) Short 10 ATM down-and-out calls on each of the 10 uncorrelated assets, the barrier is set to 95.

(b.3) Short 10 ATM down-and-out calls and 5 ATM puts on each of the 10 uncorrelated assets.

(b.4) Same as (b.3) but with the number of puts increased so that $\mathbf{b} = \mathbf{0}$ in the Delta-Gamma approximation (see Section 3.3.1).

(b.5) Short 10 ATM down-and-out calls and 5 ATM cash-or-nothing puts on each of the 10 uncorrelated assets, the cash value is set to 100.

(b.6) Same as (b.5) but with the number of cash-or-nothing puts increased so that $\mathbf{b} = \mathbf{0}$ in the Delta-Gamma approximation (see Section 3.3.1).

The at-the-money (ATM) option means, that an option's strike price is identical to the price of the underlying security, i.e. the strike price in all portfolios is equal to 100.

Portfolios (a.1) - (a.7), (b.1) contain continuous payoffs, where the approximations work quite well, while portfolios (b.2) - (b.6) contain discontinuous payoffs with approximations performing quite poor. Note also, that all these portfolios are based on options, therefore, the modified approximations from Chapter 3 can be applied.

All option prices are calculated by the Black-Scholes formula or its extensions. We refer an interested reader to Appendix C for a description of all involved options and their pricing formulas. There is a certain inconsistency in using these formulas for the normal distribution directly, but this is the choice of Glasserman (2004), Glasserman et al. (2000a,b), and as mentioned there, it is very common in practice.

The level x is usually chosen such that the probability of high portfolio losses $\mathbb{P}(L > x)$ is 5%, 1% or 0.3%.

The total sample size is set to $N = 120000$ in order to achieve precise estimates of the variance reduction rates and compare them to the results from other papers. The variance reduction rates are given in Section 6.1.3 via the so called variance reduction factors, which are the ratios of the variance of the MC estimate (no variance reduction)

to the variance of the estimators obtained by the corresponding variance reduction algorithm. Thus, the larger variance reduction factor means larger variance reduction rate, and, therefore, better performance of the corresponding variance reduction algorithm.

The sample size for the parameter update in all CE based algorithms is chosen to be $M = 10000$ (it can be, however, often reduced to 1000 without loss of performance). Since the probability $\mathbb{P}(L > x)$ of interest varies around 1% and in practice is usually not smaller than 0.1%, the rarity parameter ρ in all CE based algorithms is chosen to be 0.1.

The 40 equally partitioned and equally allocated strata are chosen for the ISS method as it is done in Glasserman (2004), Glasserman et al. (2000a,b).

The special strata is achieved by the simulated annealing algorithm (see Appendix B) with the maximum number of iterations set to 100 per stratum. This means, that the evaluation of each level y and z of the special stratification involves 100 revaluation of the portfolio loss function, as well as 100 revaluations of the much faster approximation. The parameter δ for the calculation of the corresponding multipliers m or m_1 and m_2 is chosen to be 0.9 (90% of performance of the ideal algorithms). The multipliers themselves are calculated according to the algorithm and their sum (either m or $m_1 + m_2$) is presented in the results.

Finally, the parameter K for the AOIS and AOISS algorithms is chosen to be 1000.

We should note, that the choice of these parameters is similar to the example in Section 4.5, where the comparison of the time necessary for corresponding set-up of the variance reduction algorithm is given.

6.1.2 Choice of the parametric family for CE based algorithms

All new algorithms from Chapter 4 are based on the importance sampling change of measure, which is the closest one, out of some parametric family of densities, to the optimal importance sampling density. The performance and ease of these algorithms depends mainly on the choice of the corresponding parametric family, which from one point of view should be wide, but from another point of view should be easy to handle in the corresponding optimizations of the optimal reference parameter estimator. Therefore, in order to apply our new variance reduction algorithms efficiently, one has to take care in the choice of the parametric family of densities. In the general case it is very hard to give guidance on the choice of this family, but from the particular examples one can achieve a first impression on this problem. Therefore, here, in the scope of the

light-tailed example from Section 6.1.1 (as well for the heavy-tailed example from Section 6.2.1, later in Section 6.2.2) we emphasize on the main principles of the successful choice of the parametric family.

The first, and obvious, principle is that the parametric family should contain the original density of the changes in risk factors, which under Assumption 3.1 is a d-dimensional multivariate normal distribution with zero mean vector, i.e. $\Delta\mathbf{S} \sim N_d(\mathbf{0}, \mathbf{\Sigma}_S)$. The first idea about the parametric family is simply a d-dimensional multivariate normal distribution $N_d(\boldsymbol{\mu}, \mathbf{\Sigma})$ parametrized by the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\mathbf{\Sigma}$. In this case, the reference parameter to be estimated consists of the d-dimensional vector parameter $\boldsymbol{\mu}$ and the $d \times d$ covariance matrix $\mathbf{\Sigma}$. Recall the system of equations (2.13) which appear in the reference parameter update step. It involves the gradient of the joint density function w.r.t. the parameter we are searching for. Such a system becomes decoupled and is, therefore, much easier to solve, if we have the mutually independent normals (i.e. the diagonal covariance matrix). Besides, it would involve much less parameters to estimate (only the diagonal elements of the covariance matrix). This can be done, if we modify the original density of $\Delta\mathbf{S}$ to the density with the diagonal covariance matrix. And it can be easily done by expressing the correlated normals $\Delta\mathbf{S} \sim N_d(\mathbf{0}, \mathbf{\Sigma}_S)$ via independent normals $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$ by setting,

$$\Delta\mathbf{S} = \mathbf{C}\mathbf{Z}, \quad \text{with} \quad \mathbf{C}\mathbf{C}^T = \mathbf{\Sigma}_S,$$

with the special choice of the matrix $\mathbf{C}$ discussed in Section 3.3.1. The independent normals $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$ then can be treated as the changes in risk factors by adjusting the portfolio loss and the approximation functions accordingly (by taking the superposition of the corresponding function and the function $\Delta\mathbf{S}(\mathbf{Z}) = \mathbf{C}\mathbf{Z}$). This gives us an opportunity to define the parametric family of the changes of measure by $N_d(\boldsymbol{\mu}, \mathrm{diag}(\boldsymbol{\sigma}^2))$ parametrized by the d-dimensional mean vector $\boldsymbol{\mu}$ and the d-dimensional variance vector $\boldsymbol{\sigma}^2$, where $\mathrm{diag}(\boldsymbol{\sigma}^2)$ means the matrix with diagonal elements taken from the vector $\boldsymbol{\sigma}^2$, all other elements being zeros.

We further provide the exact solution (main advantage of the corresponding reference parameter update step in the CE based algorithms) to the system of equations (2.13) under the above defined choice of the parametric family. For simplicity of notations we forget about the matrix $\mathbf{C}$ and assume, that the changes in risk factors $\Delta\mathbf{S}$ have the standard normal distribution $N_d(\mathbf{0}, \mathbf{I})$. Under the above mentioned choice of the parametric family, the vector parameter $\mathbf{v}$ is a concatenation of the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$. Moreover, due to the choice of the diagonal covariance matrix, the joint density $f(\mathbf{X}_i; \mathbf{v})$ is a product of marginal densities $N_1(\boldsymbol{\mu}_j, \boldsymbol{\sigma}_j^2)$ for $j = 1, \ldots, d$ (do not get confused between indexes i and j, the index i is responsible for sample indexing, and the index j

is responsible for indexing the d-dimensional vectors). The system (2.13) then becomes decoupled and reads as follows:

$$\begin{cases} \frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{w})\frac{\partial \log f(\mathbf{X}_{ij};\boldsymbol{\mu}_j,\boldsymbol{\sigma}_j^2)}{\partial \boldsymbol{\mu}_j} = 0, \\ \frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{w})\frac{\partial \log f(\mathbf{X}_{ij};\boldsymbol{\mu}_j,\boldsymbol{\sigma}_j^2)}{\partial \boldsymbol{\sigma}_j^2} = 0, \end{cases}$$

for $j = 1, \ldots, d$, where $\mathbf{X}_{ij}$ denotes the j-th coordinate of the vector $\mathbf{X}_i$ and $f(\cdot;\boldsymbol{\mu}_j,\boldsymbol{\sigma}_j^2)$ is the density of the normal distribution $N_1(\boldsymbol{\mu}_j,\boldsymbol{\sigma}_j^2)$.

The next step is to derive the partial derivatives $\frac{\partial \log f(x;\mu,\sigma^2)}{\partial \mu}$ and $\frac{\partial \log f(x;\mu,\sigma^2)}{\partial \sigma^2}$, which is a standard material of any statistical book (see, e.g., Anderson (2003)):

$$\frac{\partial \log f(x;\mu,\sigma^2)}{\partial \mu} = \frac{x-\mu}{\sigma^2},$$

$$\frac{\partial \log f(x;\mu,\sigma^2)}{\partial \sigma^2} = \frac{(x-\mu)^2-\sigma^2}{2\sigma^4}.$$

The above system of equations can be simplified by plugging in the partial derivatives:

$$\begin{cases} \frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{w})(\mathbf{X}_{ij}-\boldsymbol{\mu}_j) = 0, \\ \frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{w})((\mathbf{X}_{ij}-\boldsymbol{\mu}_j)^2-\boldsymbol{\sigma}_j^2) = 0, \end{cases}$$

for $j = 1, \ldots, d$.

Finally, its solution after rearranging reads as follows:

$$\begin{cases} \boldsymbol{\mu}_j = \frac{\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{w})\mathbf{X}_{ij}}{\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{w})}, \\ \boldsymbol{\sigma}_j^2 = \frac{\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{w})(\mathbf{X}_{ij}-\boldsymbol{\mu}_j)^2}{\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i))W(\mathbf{X}_i;\mathbf{u},\mathbf{w})}, \end{cases}$$

for $j = 1, \ldots, d$.

6.1.3 Results

In this section we present the results for the light-tailed example from Section 6.1.1 and discuss them in the following. We first remind the reader the set of the variance reduction algorithms which we examine here:

Importance sampling variance reduction: IS, ACE, FCE, AICE and FICE.

Importance sampling with stratification variance reduction: ISS, ACESI, ACES, AICESI and AICES.

Further improved variance reduction: AOISI, AOIS, AOISSI and AOISS.

Since, the estimator of CVaR can be done with the help of the new algorithms based on the special stratification with only one level y (see Section 5.2), we often differ ACESI, ACES, AICESI and AICES algorithms with and without the level z. Note, that the variants of the AOISSI and AOISS algorithms without z are given explicitly by AOISI and AOIS algorithms respectively.

	x	$\widehat{l}$	IS	ACE	AICE	ISS	ACESI	AICESI	AOISSI
(a.1)	130	5.0%	7.0	7.2	7.2	28.9	19.5	19.9	25.0
	196	1.1%	22.5	23.5	23.8	71.2	61.8	63.2	89.0
	260	0.3%	71.1	73.7	75.5	182.7	144.6	148.7	259.4
(a.2)	120	5.3%	7.6	7.8	8.0	85.8	71.2	74.3	69.3
	185	1.0%	29.8	31.2	31.3	289.1	201.9	207.3	205.0
	208	0.5%	52.5	54.7	55.6	484.3	328.7	339.3	348.8
(a.3)	136	1.0%	39.7	42.0	42.3	60.9	48.0	48.7	80.2
(a.4)	153	1.0%	41.3	47.0	47.4	241.2	294.1	298.0	235.5
(a.5)	141	4.7%	6.3	6.3	6.2	13.7	10.6	10.6	17.8
	207	1.1%	16.6	16.4	16.6	28.8	25.9	26.3	44.8
	236	0.5%	28.5	29.0	28.0	45.7	45.7	44.7	79.2
(a.6)	545	1.0%	27.3	23.0	26.0	47.4	29.8	34.1	47.0
(a.7)	1827	1.0%	11.3	4.0	4.6	24.1	19.3	22.8	66.1
(b.1)	265	1.0%	30.7	32.9	33.3	162.5	214.3	221.1	212.6
(b.2)	308	1.1%	25.1	17.3	17.9	45.5	45.7	50.2	73.0
(b.3)	248	1.1%	14.3	7.0	7.3	17.3	11.3	11.8	33.2
(b.4)	308	1.0%	7.3	6.6	6.5	8.4	7.4	7.2	34.0
(b.5)	771	1.1%	20.3	19.1	19.4	29.7	20.9	21.1	32.2
(b.6)	165	1.0%	0.9	0.8	0.6	1.0	0.9	0.6	10.8

TABLE 6.1: Variance reduction factors in the light-tailed case for the Delta-Gamma approximation

We now start from the Delta-Gamma approximation Q used in all variance reduction algorithms, since the IS and ISS algorithms are restricted to this choice of the approximation function. Besides, it is the most common approximation used in practice, therefore, it is important to examine our new algorithms first on this approximation. Table 6.1 gives the variance reduction factors for the IS, ACE, AICE, ISS, ACESI, AICESI and AOISSI algorithms. For now we ignored the FCE and FICE algorithms due to their heavy set-up of the importance sampling change of measure (we compare these algorithms to ACE and AICE later in Table 6.2). We also ignored the ACES, AICES and AOISS algorithms, since their variance reduction rates are simply the variance reduction rates of ACESI, AICESI and AOISSI algorithms respectively, multiplied by some

input parameter $\delta \in (0,1)$ (given by the user and is 0.9 in our numerical study) due to their construction. It is more important to investigate the sum of the multipliers m or $m_1 + m_2$ involved in all these algorithms, which we do later in Tables 6.3 and 6.4. Finally, we ignored the AOISI and AOIS algorithms, since they are based on the reduced special stratification (without level z). The influence of the level z on the corresponding algorithms is tested later in Tables 6.3 and 6.4.

From Table 6.1 we can immediately conclude, that the performance of the IS, ACE and AICE algorithms is approximately the same. This is not a surprise, since they all are based on the similar importance sampling change of measure. The parametric families of the candidates to the importance sampling density in the cases of ACE and AICE (means and variances) are wider than in the case of the IS (twisting parameter θ) algorithms, therefore, the first two algorithms usually outperform the latter one, however, insignificantly. In the case when the Delta-Gamma approximation is not precise enough (examples (b.2) - (b.6)) the IS algorithm outperforms the ACE and AICE algorithms, seams to be, because the optimal reference parameter estimator of the ACE and AICE algorithms is more sensitive to the quality of the approximation. This is, however, not crucial for our new algorithms, because of their flexibility, i.e. the ability to apply some improved approximation. For example, in the case of the portfolio (b.3) (the largest difference in the performance), ACE and AICE have variance reduction factor 20.5 for the modified Delta-Gamma approximation and even outperform the IS algorithm.

We further compare ACE and AICE with the FCE and FICE algorithms in Table 6.2. We use the advantage of the ACE and AICE algorithms to use any approximation of the portfolio loss function and give the results for the improved approximation.

One can conclude, that there is no significant difference between the ACE, AICE, FCE and FICE algorithms, unless there exists no good enough approximation. The most interesting portfolio is (b.6), where most of the algorithms even fail to give any variance reduction (see also Table 6.1). The successful variance reduction is achieved only by the FCE, FICE (not very useful in practice) and the AOISSI (hence also AOISI, AOIS and AOISS) algorithms. Let us have a look at the quality of the Delta-Gamma and the modified Delta-Gamma approximations, see Figure 6.1.

It is not really satisfactory, in the sense that the values of the portfolio losses can highly deviate from the values of the corresponding approximation. The situation gets even worse in the case of the Charm-Speed and the modified Charm-Speed approximations, see e.g. Figure 6.2.

As one can conclude from Table 6.1, the algorithms based on the importance sampling only (without any kind of stratification) perform in the most situations significantly

	x	$\widehat{l}$	ACE	FCE	AICE	FICE
(a.1)	130	5.0%	7.2	7.1	7.2	7.2
	196	1.1%	23.5	22.8	23.8	23.3
	260	0.3%	74.4	73.5	75.5	74.5
(a.2)	120	5.3%	8.0	7.8	8.0	7.9
	185	1.0%	31.4	30.6	31.6	31.1
	208	0.5%	55.6	54.3	56.0	54.7
(a.3)	136	1.0%	52.7	59.6	53.4	60.3
(a.4)	153	1.0%	47.0	46.0	47.4	46.2
(a.5)	141	4.7%	6.6	6.5	6.6	6.6
	207	1.1%	17.9	18.0	18.0	18.2
	236	0.5%	31.6	31.6	31.8	32.0
(a.6)	545	1.0%	25.7	25.5	27.4	28.0
(a.7)	1827	1.0%	10.5	10.7	12.5	12.9
(b.1)	265	1.0%	32.9	32.2	33.3	32.4
(b.2)	308	1.1%	27.8	27.6	28.3	27.9
(b.3)	248	1.1%	20.8	22.1	21.4	22.4
(b.4)	308	1.0%	15.3	19.5	15.5	19.6
(b.5)	771	1.1%	22.9	18.4	23.2	20.9
(b.6)	165	1.0%	1.5	8.7	1.3	7.1

TABLE 6.2: Variance reduction factors: comparison of ACE and AICE with improved approximation against the FCE and FICE algorithms in the light-tailed case

worse than the same algorithms but with extra stratification. Because of this, we do not test the IS, ACE, AICE, FCE, FICE methods further, emphasizing on the fact that it is worth performing an extra stratification to any importance sampling algorithm.

The next step is to investigate the influence of the level z of the special stratification on the performance of our new algorithms and to give the sum of the multipliers involved in the ACES, AICES, AOIS and AOISS algorithms (variance reduction rate of these algorithms is simply the variance reduction rate of the ACESI, AICESI, AOISI and AOISSI algorithms respectively, multiplied by δ, which is chosen in our study to be 0.9). In Table 6.3 we investigate the algorithms based on the full special stratification and in Table 6.4 those which are based on the special stratification without the level z (the corresponding algorithms can be easily modified by setting $z = \infty$). Both tables involve only the Delta-Gamma approximation.

From Tables 6.3 and 6.4 one can observe that, the higher the variance reduction for a particular portfolio, the larger the sum of the multipliers is. For example, consider the portfolio (a.1) with level $x = 260$ in Table 6.3. We can observe, that the AOISSI algorithm significantly outperforms the ACESI and AICESI algorithms, but with the cost

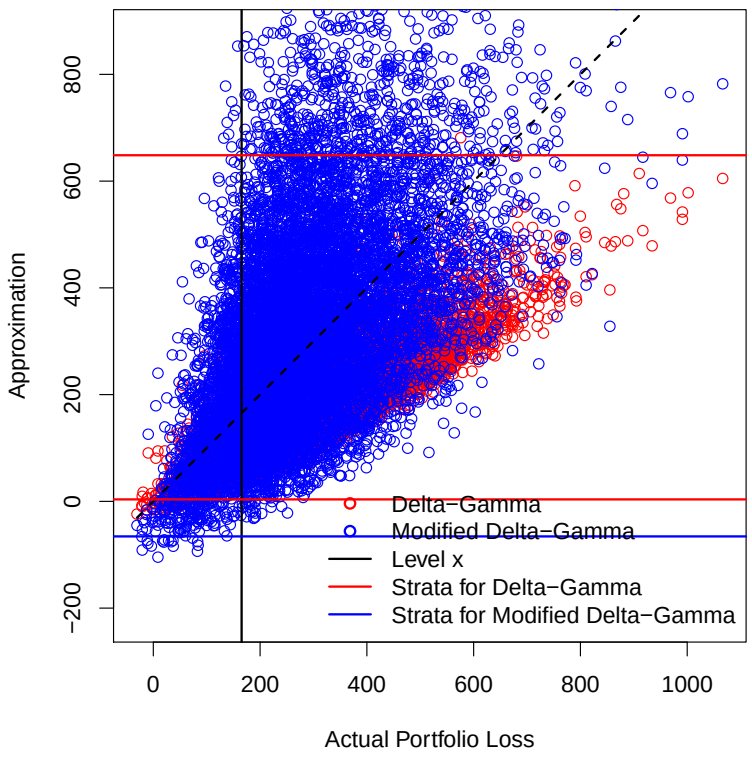


FIGURE 6.1: Comparison of Delta-Gamma and Modified Delta-Gamma approximations against actual portfolio losses in 10000 randomly generated scenarios for the portfolio (b.6) in the light-tailed case (for details on the portfolio see Section 6.1.1).

of a higher value of the multipliers. This is expected, since the multipliers control the variance of the corresponding estimators of the probabilities of strata (roughly speaking, they destroy the variance reduction rate of the corresponding variance reduction algorithm in comparison to the same variance reduction algorithm in the ideal setting), which should become smaller if the variance of the corresponding variance reduction algorithm in ideal setting becomes smaller.

If we compare Table 6.3 with Table 6.4, we can see, that the latter one has smaller multipliers, again because of the smaller variance reduction rate (variance reduction factors of the algorithms based on the special strata with two levels are larger than the variance reduction factors of the same algorithms based on the special strata without level z).

In Chapter 4 we have mentioned, that the multipliers influence the time necessary to set-up the corresponding variance reduction algorithm, and, therefore, should be kept as small as possible. On the other hand, as concluded from the numerical study (see Tables 6.3 and 6.4), a better performance of the variance reduction procedures involves larger multipliers. We also have mentioned, that the variance reduction algorithms that involve these multipliers, can be easily modified, so that the multipliers play the role of

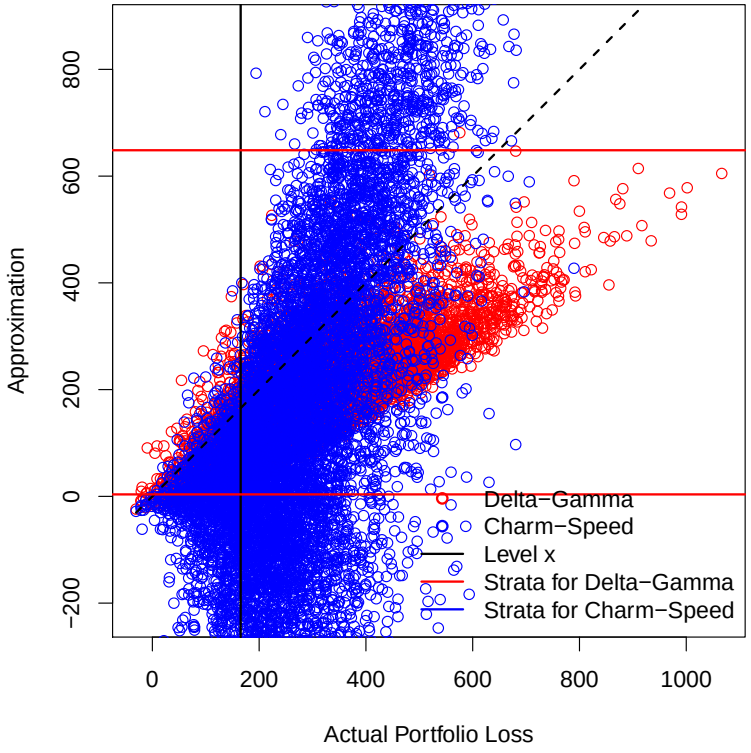


FIGURE 6.2: Comparison of Delta-Gamma and Charm-Speed approximations against actual portfolio losses in 10000 randomly generated scenarios for the portfolio (b.6) in the light-tailed case (for details on the portfolio see Section 6.1.1).

the input parameters. This means, that we can easily achieve a compromise between the speed of the corresponding variance reduction algorithms and their performance.

The final observation about the multipliers is that their value, generally, decreases as the level x increases (see portfolios (a.1), (a.2) and (a.5)). The rate of decrease is higher in the case of the full special stratification.

Let us now conclude about the influence of the second level z of the special stratification on the performance of the corresponding algorithms. A first observation from Tables 6.3 and 6.4 is that this influence vanishes as level x increases (see portfolios (a.1), (a.2) and (a.5)). This is because, as we already mentioned before, the multipliers value for the algorithms with special stratification without level z decreases slower than the one of the same algorithms, based on the full special stratification (and as we already mentioned, the multiplier value reflects the performance of the corresponding variance reduction algorithm).

The final and very important observation from Tables 6.3 and 6.4 is that the second level z of the special stratification has smaller effect on the AOISSI (and, therefore,

	x	$\widehat{l}$	ACESI	ACES m	AICESI	AICES m	AOISSI	AOISS m
(a.1)	130	5.0%	19.5	17	19.9	17	25.0	32
	196	1.1%	61.8	13	63.2	14	89.0	34
	260	0.3%	144.6	6	148.7	6	259.4	24
(a.2)	120	5.3%	71.2	123	74.3	130	69.3	139
	185	1.0%	201.9	67	207.3	69	205.0	90
	208	0.5%	328.7	57	339.3	60	348.8	83
(a.3)	136	1.0%	48.1	2	48.7	2	80.2	4
(a.4)	153	1.0%	294.1	77	298.0	78	235.5	72
(a.5)	141	4.7%	10.6	6	10.6	6	17.8	19
	207	1.1%	25.9	4	26.3	4	44.8	13
	236	0.5%	45.7	4	44.7	4	79.2	11
(a.6)	545	1.0%	29.8	3	34.1	3	47.0	5
(a.7)	1827	1.0%	19.3	11	22.8	14	66.1	93
(b.1)	265	1.0%	214.4	67	221.1	69	212.6	87
(b.2)	308	1.1%	45.7	11	50.2	13	73.0	17
(b.3)	248	1.1%	11.2	2	11.8	3	33.2	5
(b.4)	308	1.0%	7.4	2	7.2	2	34.0	6
(b.5)	771	1.1%	20.9	2	21.1	2	32.2	3
(b.6)	165	1.0%	0.9	2	0.6	2	10.8	2

TABLE 6.3: Variance reduction factors in the light-tailed case for the Delta-Gamma approximation and full special stratification together with the sums of the multipliers $m = m_1 + m_2$ for the corresponding algorithms not in the ideal setting

on AOISS) algorithm than on the ACESI and AICESI (and, therefore, on ACES and AICES) algorithms, but again with the cost of a larger value of the multiplier.

Next step in our numerical study is to test different approximations from Chapter 3. The variance reduction factors for our new algorithms in the case of the Delta-Gamma approximation are given in Table 6.1. We further present the variance reduction factors for our new algorithms in the case of the modified Delta-Gamma approximation in Table 6.5, in the case of the Charm-Speed approximation in Table 6.6 and in the case of the modified Charm-Speed approximation in Table 6.7. In these tables we present the results for the ACESI algorithm only (we do not investigate AICESI anymore), since, as all previous tables showed, it performs approximately as the AICESI algorithm. Among the ACESI and AICESI algorithms we have chosen ACESI, since its performance is usually slightly worse than the performance of the AICESI algorithm (we choose the worst case). We also include in these tables the variant of ACESI with the special stratification without level z, and the AOISI algorithm (variant of AOISSI with the special stratification without level z), because they are important for the further estimation of CVaR.

	x	$\widehat{l}$	ACESI	ACES m	AICESI	AICES m	AOISSI	AOISS m
(a.1)	130	5.0%	12.2	6	12.4	6	23.4	27
	196	1.1%	41.9	6	42.9	6	86.1	31
	260	0.3%	133.8	5	137.5	5	259.0	24
(a.2)	120	5.3%	14.5	8	14.9	9	49.3	82
	185	1.0%	57.1	7	58.2	7	167.9	63
	208	0.5%	100.4	7	103.2	7	292.3	60
(a.3)	136	1.0%	47.6	2	48.1	2	79.8	4
(a.4)	153	1.0%	78.8	7	79.7	7	165.9	37
(a.5)	141	4.7%	9.8	5	9.8	5	17.8	19
	207	1.1%	25.4	4	25.7	4	44.8	13
	236	0.5%	45.0	3	44.0	3	79.2	11
(a.6)	545	1.0%	28.5	2	32.5	3	46.6	5
(a.7)	1827	1.0%	8.9	3	10.2	3	60.8	91
(b.1)	265	1.0%	60.6	7	61.9	7	175.0	62
(b.2)	308	1.1%	19.6	2	20.4	2	42.4	5
(b.3)	248	1.1%	7.4	2	7.8	2	28.7	3
(b.4)	308	1.0%	7.3	2	7.1	2	34.0	6
(b.5)	771	1.1%	20.4	2	20.7	2	32.1	3
(b.6)	165	1.0%	0.8	2	0.6	2	10.8	2

TABLE 6.4: Variance reduction factors in the light-tailed case for the Delta-Gamma approximation and special stratification without z together with the multipliers m for the corresponding algorithms not in the ideal setting

It is hard to see any pattern in all four approximations from Tables 6.1, 6.5, 6.6 and 6.7. Different algorithms for the same portfolio might perform best for different approximations. For example, in the case of portfolio (a.5) with level $x = 236$, the best performance of the ACESI algorithm is achieved by the modified Charm-Speed approximation (variance reduction factor of 77.5), but the best performance of the AOISSI algorithm is achieved by the Charm-Speed approximation (variance reduction factor of 110.8). We can only notice, that the modified versions of both, the Delta-Gamma and the Charm-Speed approximations, work better for portfolios with discontinuous pay-offs, namely (b.2)-(b.6).

We further use the main advantage of the new algorithms, namely their flexibility in working with any approximation function, and present Table 6.8, which compares the ISS algorithm with our new algorithms based on the improved approximation (the one which gives the best variance reduction rate for a particular portfolio and algorithm among four recently mentioned approximations). The approximation used for the ISS algorithm is, of course, the Delta-Gamma approximation.

	x	$\widehat{l}$	ACESI (no z)	ACESI	AOISI	AOISSI
(a.1)	130	5.0%	15.5	18.1	27.4	27.8
	196	1.1%	47.8	55.4	81.0	81.8
	260	0.3%	145.0	150.6	236.2	236.5
(a.2)	120	5.3%	15.4	70.1	61.3	78.6
	185	1.0%	59.9	209.8	199.6	235.7
	208	0.5%	106.4	314.5	348.8	393.6
(a.3)	136	1.0%	54.5	182.8	71.4	145.1
(a.4)	153	1.0%	74.1	459.7	181.4	340.3
(a.5)	141	4.7%	14.1	14.6	22.0	22.1
	207	1.1%	36.7	37.2	50.7	50.7
	236	0.5%	66.7	67.4	96.6	96.6
(a.6)	545	1.0%	59.1	60.6	65.8	67.0
(a.7)	1827	1.0%	37.5	107.1	137.1	145.8
(b.1)	265	1.0%	64.7	117.2	138.5	148.5
(b.2)	308	1.1%	36.8	81.5	84.5	97.6
(b.3)	248	1.1%	31.3	38.5	58.5	59.6
(b.4)	308	1.0%	22.7	23.0	49.5	49.5
(b.5)	771	1.1%	29.0	29.2	31.0	31.1
(b.6)	165	1.0%	1.5	1.5	9.2	9.2

TABLE 6.5: Variance reduction factors in the light-tailed case for the modified Delta-Gamma approximation

The first and obvious conclusion after having a look at Tables 6.1 and 6.8 is that the variance reduction rate of the new algorithms can be significantly improved, if some better approximation is used instead of the Delta-Gamma approximation. Another direct observation is that the ISS algorithm has been outperformed by the AOISSI algorithm for all portfolios and by the AOISI algorithm in all portfolios except (a.4), where the difference is actually not crucial.

One can expect, that the AOISSI algorithm should perform, in general, better than the ACESI algorithm, due to the improved parametric family of densities among which an importance sampling density is chosen. But this is apparently not the case, see e.g. portfolios (a.2)-(a.4). In order to investigate this exceptional behavior of the AOISSI algorithm, we present a few plots of the corresponding estimated importance sampling densities before performing any stratification and the corresponding quality of the used approximation.

We first take the portfolio (a.4) for which, according to Table 6.8, ACESI has significantly outperformed the AOISSI algorithm (variance reduction factor of 459.7 against 340.3). Both results are achieved by using the modified Delta-Gamma approximation.

	x	$\widehat{l}$	ACESI (no z)	ACESI	AOISI	AOISSI
(a.1)	130	5.0%	13.2	16.0	38.5	38.7
	196	1.1%	41.5	47.0	112.2	112.5
	260	0.3%	131.3	152.5	342.8	343.5
(a.2)	120	5.3%	15.1	299.8	118.1	304.3
	185	1.0%	56.3	695.5	317.1	644.3
	208	0.5%	100.0	1121.5	547.9	1014.4
(a.3)	136	1.0%	56.5	70.8	94.2	100.7
(a.4)	153	1.0%	60.6	309.3	95.9	224.3
(a.5)	141	4.7%	11.9	13.9	27.3	27.4
	207	1.1%	32.2	33.8	66.8	66.8
	236	0.5%	57.3	59.0	116.2	116.2
(a.6)	545	1.0%	36.6	38.9	67.9	68.3
(a.7)	1827	1.0%	30.1	61.9	208.2	209.3
(b.1)	265	1.0%	56.4	169.3	198.9	225.4
(b.2)	308	1.1%	24.0	28.4	33.1	36.1
(b.3)	248	1.1%	13.6	14.3	24.4	25.9
(b.4)	308	1.0%	7.4	7.4	49.0	49.0
(b.5)	771	1.1%	5.5	6.7	24.0	25.8
(b.6)	165	1.0%	1.3	1.3	9.2	9.2

TABLE 6.6: Variance reduction factors in the light-tailed case for the Charm-Speed approximation

As we can see from Figure 6.3, the quality of this approximation in the case of the portfolio (a.4) is very good and the interval (y, z) of the special stratification is narrow. This, from our point of view, is the reason why the ACESI algorithm has performed better than AOISSI. It can only be confirmed by looking at Figure 6.4, where the importance sampling densities of the ACESI and AOISSI algorithms are drawn. Obviously, the importance sampling density of the AOISSI algorithm (blue line) approximates the optimal importance sampling density (red line) significantly better than the importance sampling density of the ACESI algorithm (green line). It also can be seen by comparing the variance reduction factors of the ACESI algorithm based on the special stratification without the level z and the variance reduction factor of the AOISI algorithm, namely 78.8 against 181.4. Removing the stratification in the ACESI algorithm will make its performance even worse.

We further take the portfolio (a.7), where the AOISSI algorithm has now outperformed the ACESI algorithm (variance reduction factor of 209.3 against 107.1, see Table 6.8), and perform the same analysis. We first plot the Charm-Speed approximation in Figure 6.5, which is involved in both results. The quality of this approximation is not bad either.

	x	$\widehat{l}$	ACESI (no z)	ACESI	AOISI	AOISSI
(a.1)	130	5.0%	14.8	15.0	32.7	32.7
	196	1.1%	42.7	43.3	90.6	90.6
	260	0.3%	124.7	125.7	262.7	262.7
(a.2)	120	5.3%	15.7	231.8	116.8	230.6
	185	1.0%	60.0	589.9	331.5	543.2
	208	0.5%	104.8	961.7	548.0	870.8
(a.3)	136	1.0%	38.8	125.5	67.8	125.9
(a.4)	153	1.0%	55.0	279.3	82.9	210.0
(a.5)	141	4.7%	16.1	17.5	26.1	26.2
	207	1.1%	44.0	44.8	64.6	64.7
	236	0.5%	76.8	77.5	110.8	110.8
(a.6)	545	1.0%	54.8	58.2	61.7	65.3
(a.7)	1827	1.0%	37.3	71.0	179.5	181.0
(b.1)	265	1.0%	65.6	103.0	169.7	175.7
(b.2)	308	1.1%	31.8	32.7	34.1	34.7
(b.3)	248	1.1%	23.9	24.3	27.4	27.6
(b.4)	308	1.0%	15.1	15.1	46.0	46.0
(b.5)	771	1.1%	7.1	7.1	24.1	24.1
(b.6)	165	1.0%	1.2	1.2	9.2	9.2

TABLE 6.7: Variance reduction factors in the light-tailed case for the modified Charm-Speed approximation

But the importance sampling density of the ACESI algorithm (green line) approximates the optimal importance sampling density (red line) (see Figure 6.6) now much worse than in Figure 6.4. The importance sampling density of the AOISSI algorithm (blue line) is again a good approximation to the optimal importance sampling density (red line), see Figure 6.6.

6.2 Heavy - tailed case

In this section we test the heavy-tailed case, i.e. the case when the underlying changes in risk factors are assumed to have a multivariate Student t-distribution, see Assumption 3.2.

6.2.1 Setup of the example

For the heavy-tailed case we basically use the same portfolios as given in Section 6.1.1 and the same choices of the parameters except some of them, on which we comment in

	x	$\widehat{l}$	ISS	ACESI (no z)	ACESI	AOISI	AOISSI
(a.1)	130	5.0%	28.9	15.5	19.5	38.5	38.7
	196	1.1%	71.2	47.8	61.8	112.2	112.5
	260	0.3%	182.7	145.0	152.5	342.8	343.5
(a.2)	120	5.3%	85.8	15.7	299.8	118.1	304.3
	185	1.0%	289.1	60.0	695.5	331.5	644.3
	208	0.5%	484.3	106.4	1121.5	548.0	1014.4
(a.3)	136	1.0%	60.9	56.5	182.8	94.2	145.1
(a.4)	153	1.0%	241.2	78.8	459.7	181.4	340.3
(a.5)	141	4.7%	13.7	16.1	17.5	27.3	27.4
	207	1.1%	28.8	44.0	44.8	66.8	66.8
	236	0.5%	45.7	76.8	77.5	116.2	116.2
(a.6)	545	1.0%	47.4	59.1	60.6	67.9	68.3
(a.7)	1827	1.0%	24.1	37.5	107.1	208.2	209.3
(b.1)	265	1.0%	162.5	65.6	169.3	198.9	225.4
(b.2)	308	1.1%	45.5	36.8	81.5	84.5	97.6
(b.3)	248	1.1%	17.3	31.3	38.5	58.5	59.6
(b.4)	308	1.0%	8.4	22.7	23.0	49.5	49.5
(b.5)	771	1.1%	29.7	29.0	29.2	32.1	32.2
(b.6)	165	1.0%	1.0	1.5	1.5	10.8	10.8

TABLE 6.8: Variance reduction factors in the light-tailed case for improved approximation

the following.

Here, again, all assets have initial values of 100 and are modeled as geometric Brownian motions with an annual volatility of $\sigma = 0.3$. In order to make it consistent with Assumption 3.2, the standard deviation of the changes in risk factors $\Delta\mathbf{S}_j$ should be given by $\mathbf{S}_j\sigma\sqrt{\Delta t}$, which gives the scaled Student t-distribution:

$$\Delta\mathbf{S} \sim \sqrt{\frac{\nu-2}{\nu}}t_d(\nu, \mathbf{0}, \mathbf{\Sigma}_S),$$

under the corresponding choice of the covariance matrix $\mathbf{\Sigma}_S$, similar to the light-tailed case from Section 6.1.1.

The total sample size N is taken now to be 40000, as motivated by the same choice in the numerical study of Glasserman et al. (2002).

The options are valued by the Black-Scholes formulas and their extensions (see Appendix C). There is, of course, inconsistency in using these formulas for the case of the Student t-distribution, but this is the choice in Glasserman et al. (2002), and as mentioned there, it is very common in practice.

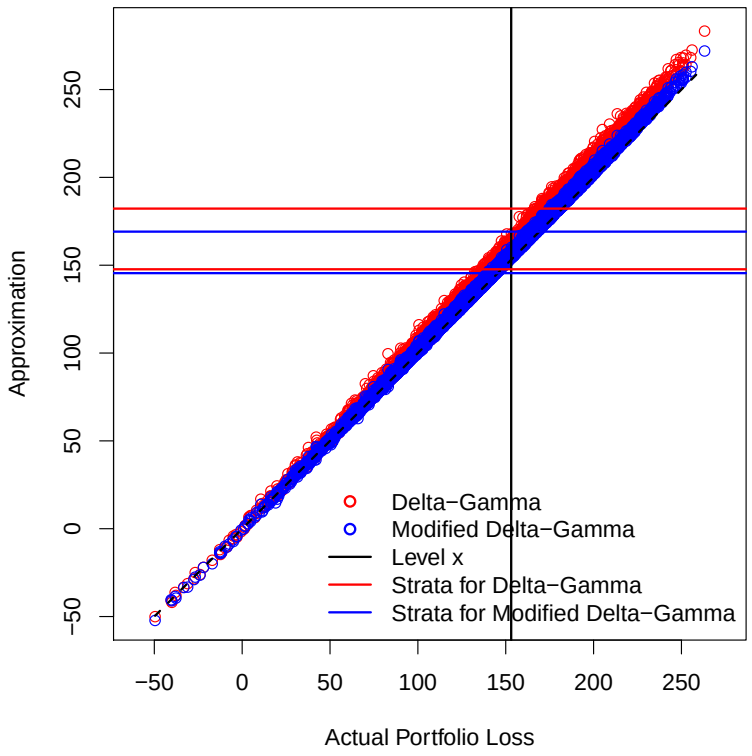


FIGURE 6.3: Quality of the modified Delta-Gamma approximation in 10000 randomly generated scenarios for the portfolio (a.4) in the light-tailed case (for details on the portfolio see Section 6.1.1).

6.2.2 Choice of the parametric family for CE based algorithms

As already mentioned in Section 6.1.2, the correct choice of the parametric family for CE based algorithms plays a key role in achieving the best performance with the smallest time costs. In Section 6.1.2 we discussed the main principles in the choice of the parametric family on the example of normally distributed changes in risk factors. In this section we give a similar discussion on the example of the multivariate Student t-distributed changes in risk factors, i.e. under Assumption 3.2.

We first note, that according to Section 6.2.1, we model the changes in risk factors according to the scaled multivariate Student t-distribution $\Delta\mathbf{S} \sim \sqrt{\frac{\nu-2}{\nu}} t_d(\nu, \mathbf{0}, \mathbf{\Sigma}_S)$. Similarly to Section 6.1.2, we are interested in making the system of equations (2.13) decoupled. As one could understand from Section 6.1.2, it can be done by expressing the changes in risk factors $\Delta\mathbf{S}$ via a random vector with independent marginals. In the case of the scaled Student t-distribution it can be done by using the expression:

$$\Delta\mathbf{S} = \sqrt{\frac{\nu-2}{\nu}} \frac{\mathbf{CZ}}{\sqrt{Y/\nu}} = \frac{\mathbf{CZ}}{\sqrt{Y/(\nu-2)}},$$

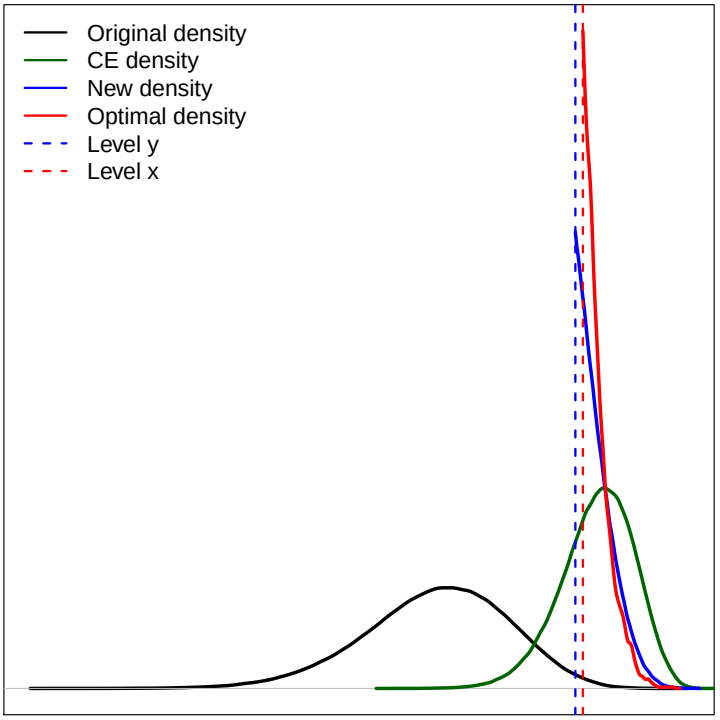


FIGURE 6.4: Comparison of ACESI and AOISSI importance sampling densities in 500000 randomly generated scenarios for the portfolio (a.4) in the light-tailed case (for details on the portfolio see Section 6.1.1).

where $\mathbf{C}\mathbf{C}^T = \mathbf{\Sigma}_S$ and $Y \sim \chi^2_\nu$ is a chi-squared distributed random variable with ν degrees of freedom and which is independent of the vector $\mathbf{Z}$. This gives us the opportunity to define the parametric family in a similar way to Section 6.1.2, parametrized by the d-dimensional mean vector $\boldsymbol{\mu}$ and the variance vector $\boldsymbol{\sigma}^2$, as well as the parameter ν.

The first improvement can be done by extending the family of densities for the random variable Y, which is chi-squared distributed with ν degrees of freedom. This is a special case of the gamma distributed random variable. More precisely, the scaled chi-squared distributed random variable with ν degrees of freedom, $\frac{Y}{\nu-2}$, has the gamma distribution $\frac{Y}{\nu-2} \sim \Gamma\left(\frac{\nu}{2}, \frac{2}{\nu-2}\right)$ with the shape parameter $\nu/2$ and the scale parameter $2/(\nu-2)$. This procedure extends the parametric family by the shape and scale parameters of the gamma distribution, which substitute the parameter ν. Along this section we denote by s the shape parameter and by θ the scale parameter of the corresponding gamma distribution. The original changes in risk factors correspond then to the shape parameter $s = \nu/2$ and the scale parameter $\theta = 2/(\nu-2)$.

If we simplify the notations similarly to Section 6.1.2 by applying the superposition of the portfolio loss and approximation functions with the function $\Delta\mathbf{S}(\mathbf{Z}, Y) = \frac{\mathbf{C}\mathbf{Z}}{\sqrt{Y}}$,

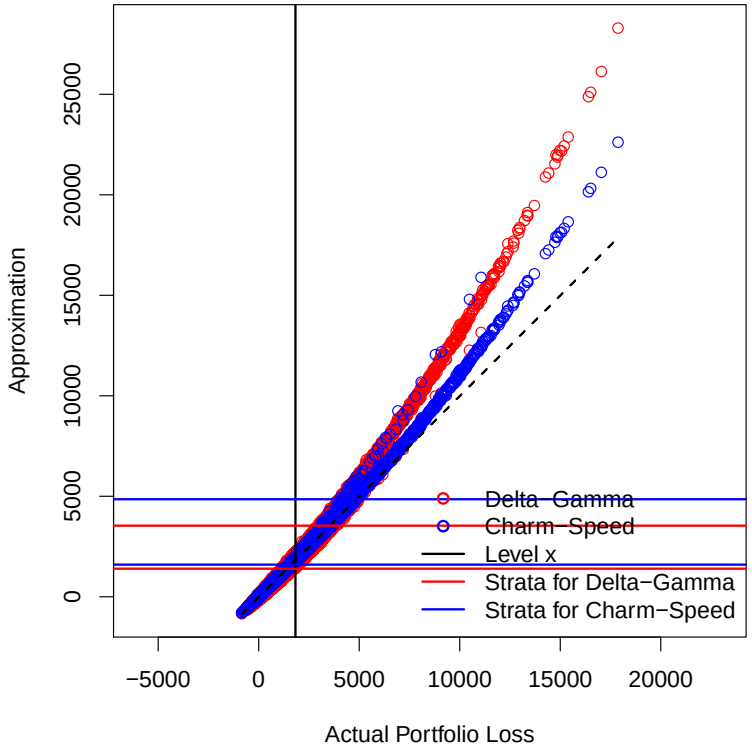


FIGURE 6.5: Quality of the Charm-Speed approximation in 10000 randomly generated scenarios for the portfolio (a.7) in the light-tailed case (for details on the portfolio see Section 6.1.1).

where $\mathbf{Z} \sim N_d(\boldsymbol{\mu}, \mathrm{diag}(\boldsymbol{\sigma}^2))$ and $Y \sim \Gamma(s, \theta)$, then the random vector $\mathbf{Z}$, extended by the independent random variable Y, can be treated as the changes in risk factors.

We further provide the exact solution to the system of equations (2.13) under the above defined choice of the parametric family, where the parameter vector $\mathbf{v}$ is a concatenation of the vectors $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^2$ and the parameters s and θ. Moreover, due to the choice of the diagonal covariance matrix and the independent variable Y, the joint density $f(\mathbf{X}_i, Y_i; \mathbf{v})$ is a product of the marginal densities $N_1(\boldsymbol{\mu}_j, \boldsymbol{\sigma}_j^2)$ for $j = 1, \ldots, d$ and $\Gamma(s, \theta)$. The system (2.13) then becomes decoupled and its first d pairs of equations read as follows (the same as in Section 6.1.2):

$$\begin{cases} \boldsymbol{\mu}_j = \frac{\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i,Y_i))W(\mathbf{X}_i,Y_i;\mathbf{u},\mathbf{w})\mathbf{X}_{ij}}{\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i,Y_i))W(\mathbf{X}_i,Y_i;\mathbf{u},\mathbf{w})}, \\ \boldsymbol{\sigma}_j^2 = \frac{\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i,Y_i))W(\mathbf{X}_i,Y_i;\mathbf{u},\mathbf{w})(\mathbf{X}_{ij}-\boldsymbol{\mu}_j)^2}{\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i,Y_i))W(\mathbf{X}_i,Y_i;\mathbf{u},\mathbf{w})}, \end{cases}$$

for $j = 1, \ldots, d$.

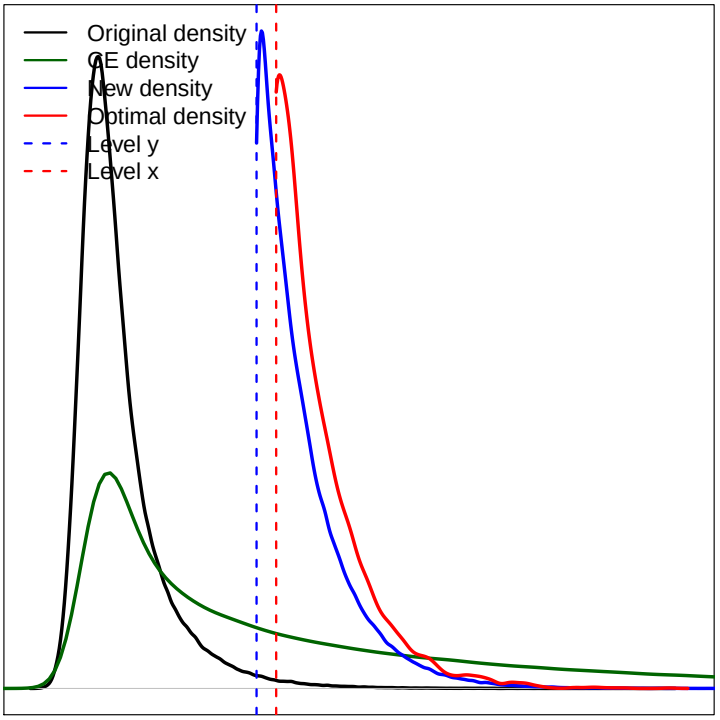


FIGURE 6.6: Comparison of ACESI and AOISSI importance sampling densities in 500000 randomly generated scenarios for the portfolio (a.7) in the light-tailed case (for details on the portfolio see Section 6.1.1).

The last pair of equations for the marginal Y reads as follows:

$$\begin{cases} \frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i,Y_i))W(\mathbf{X}_i,Y_i;\mathbf{u},\mathbf{w})\frac{\partial \log f(Y_i;s,\theta)}{\partial s} = 0, \\ \frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i,Y_i))W(\mathbf{X}_i,Y_i;\mathbf{u},\mathbf{w})\frac{\partial \log f(Y_i;s,\theta)}{\partial \theta} = 0 \end{cases}$$

where $f(\cdot;s,\theta)$ is the density of the gamma distribution $\Gamma(s,\theta)$, parametrized by the shape parameter s and the scale parameter θ.

The next step is to derive the partial derivatives $\frac{\partial \log f(y;s,\theta)}{\partial s}$ and $\frac{\partial \log f(y;s,\theta)}{\partial \theta}$, which is a standard material of any statistical book:

$$\frac{\partial \log f(y;s,\theta)}{\partial s} = \log\left(\frac{y}{s}\right) - \frac{\Gamma'(s)}{\Gamma(s)},$$

$$\frac{\partial \log f(y;s,\theta)}{\partial \theta} = \frac{y-s\theta}{\theta^2},$$

where $\Gamma(s)$ means here the gamma function and $\Gamma'(s)$ denotes its derivative. Note, that the function $\frac{\Gamma'(s)}{\Gamma(s)}$ is called the digamma function.

The solution to the above system of equations after plugging in the partial derivatives

can be achieved through solving the second equation w.r.t. the product $s\theta$ and then finding a zero of the first equation by means of optimization tools:

$$\begin{cases} s\theta = \frac{\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i,Y_i))W(\mathbf{X}_i,Y_i;\mathbf{u},\mathbf{w})Y_i}{\frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i,Y_i))W(\mathbf{X}_i,Y_i;\mathbf{u},\mathbf{w})}, \\ \frac{1}{M}\sum_{i=1}^{M} I_{(\gamma,+\infty)}(G(\mathbf{X}_i,Y_i))W(\mathbf{X}_i,Y_i;\mathbf{u},\mathbf{w})\left(\log\left(\frac{Y_i s}{s\theta}\right) - \frac{\Gamma'(s)}{\Gamma(s)}\right) = 0, \\ \theta = \frac{s\theta}{s}. \end{cases}$$

6.2.3 Results

	x	$\widehat{l}$	IS	ACE	AICE	ISS	ACESI	AICESI	AOISSI
(a.1)	204	5.0%	5.6	11.5	10.6	11.8	30.4	28.4	36.9
	469	0.9%	21.2	58.6	56.6	34.9	113.9	143.1	148.9
	719	0.3%	49.6	154.2	158.7	69.8	382.2	381.8	298.9
(a.2)	143	5.1%	6.2	8.9	8.8	13.4	200.2	242.2	213.6
	311	1.0%	25.2	44.3	43.9	45.1	414.0	342.5	379.4
	497	0.3%	74.6	145.5	126.0	124.4	677.0	681.4	766.6
(a.3)	149	0.9%	37.5	31.7	29.3	91.5	35.1	18.4	49.8
(a.4)	145	1.0%	24.7	40.2	42.2	43.1	261.3	193.4	249.7
(a.5)	276	5.0%	5.6	12.4	13.5	11.4	30.1	32.0	35.8
	617	1.0%	18.8	59.4	53.9	29.4	107.6	111.6	77.2
	1009	0.3%	53.6	190.0	188.1	71.3	412.4	401.7	259.7
(a.6)	5112	1.0%	24.5	92.8	97.6	77.4	370.5	412.4	249.7
(a.7)	5441	1.0%	16.8	50.7	56.4	30.2	95.5	234.5	299.5
(b.1)	415	1.0%	22.6	38.2	38.7	38.4	108.4	138.8	196.5
(b.2)	482	0.9%	37.2	37.4	33.7	57.2	147.8	134.9	22.1
(b.3)	518	1.0%	38.3	45.2	29.7	65.3	89.4	81.6	6.7
(b.4)	800	1.0%	31.4	9.4	31.4	56.8	35.2	39.3	42.4
(b.5)	835	1.0%	21.6	5.4	9.6	28.2	9.3	12.3	2.7
(b.6)	345	1.0%	37.6	5.3	4.7	52.6	5.9	4.9	35.1

TABLE 6.9: Variance reduction factors in the heavy-tailed case for the Delta-Gamma approximation

In this section we present the results for the heavy-tailed example from Section 6.2.1 and discuss them in the following. The process of testing is the same as in the light-tailed case in Section 6.1.3. We first start from the analogue to Table 6.1. More precisely, in Table 6.9 we collect the variance reduction factors of all variance reduction algorithms based on the Delta-Gamma approximation.

The first part (portfolios with continuous pay-offs (a.1)-(a.7)) of Table 6.9 looks expected. The variance reduction factors of the ACE and AICE algorithms are approximately the

same, always outperforming the variance reduction factors of the IS algorithm (even without using improved approximation), except only for the portfolio (a.3). For the improved approximation, namely the modified Delta-Gamma approximation, the corresponding variance reduction factors of the ACE and AICE algorithms become 41.2 and 42.5 respectively, that is, larger than the variance reduction factor 37.5 of the IS algorithm. This advantage of our new importance sampling algorithms can be explained by the wider parametric family of densities among which an importance sampling density is searched. In the case of the IS algorithm this family is parametrized only by the twisting parameter θ, whereas in the case of the ACE and AICE algorithms, the family is parametrized by the mean vector, the diagonal covariance matrix and the shape and scale parameters.

We further note that, as mentioned Glasserman et al. (2002), an extra stratification has a smaller effect on the variance reduction rate in this heavy-tailed case than in the light-tailed case. This can be also seen from Table 6.9, and can be explained by the extra dependency between the marginal changes in risk factors, even if they are not correlated. This problem has, however, vanished in the case of our new algorithms due to the wide enough choice of the parametric family (extra variance reduction achieved by the special stratification is often significant, see Table 6.9). Therefore, our new algorithms outperform the ISS algorithm for all portfolios with continuous pay-offs even more than in the light-tailed case and even without any improved approximation.

Much more interesting is the second part (portfolios with discontinuous pay-offs (b.2)-(b.6)) of Table 6.9, which has become a surprise for us. It is clear, that for these portfolios, the Delta-Gamma approximation does not fit the portfolio loss function good enough. But the IS and ISS algorithm showed a quite good performance (at least in comparison to the light-tailed case, where it could not give any variance reduction for the portfolio (b.6), see Table 6.1). This is not crucial for us, because for the improved approximation our new algorithms still outperform the IS and ISS algorithms (see Table 6.10), but it is important to understand the reason of such an unexpected behavior. In order to do this we should have a look at the corresponding importance sampling densities.

As a test example we chose the portfolio (b.5) for which our new algorithms show the worst performance for both Delta-Gamma approximation and the improved one (modified Delta-Gamma approximation in this case). The importance sampling densities for the ACESI and AOISSI algorithms are given in Figures 6.7 and 6.8 for the Delta-Gamma and the modified Delta-Gamma approximations correspondingly. In Figure 6.7 one can see, that neither the importance sampling density of the AOISSI algorithm (blue line), nor the importance sampling density of the ACESI algorithm (green line), approximate

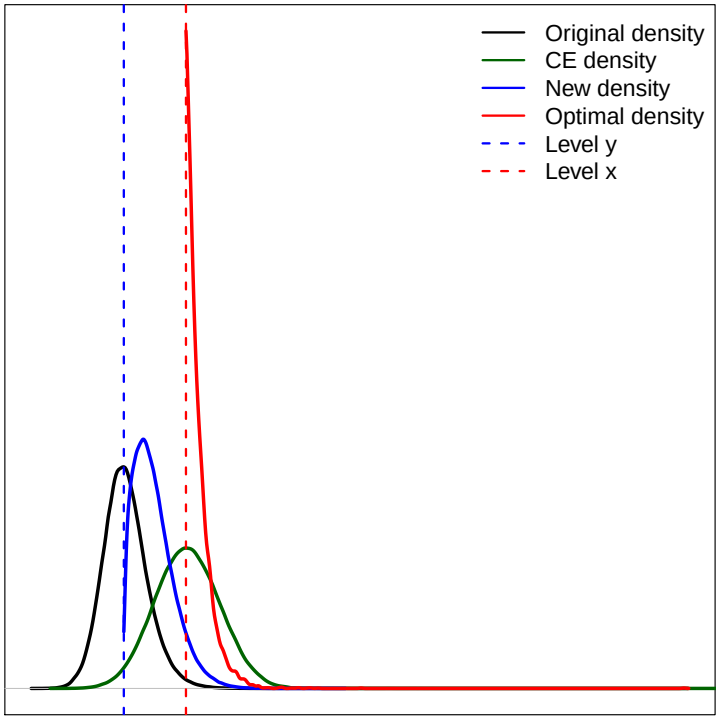


FIGURE 6.7: Comparison of ACESI and AOISSI importance sampling densities in 500000 randomly generated scenarios for the portfolio (b.5) in the heavy-tailed case for Delta-Gamma approximation (for details on the portfolio see Section 6.1.1).

the optimal importance sampling density (red line) sufficiently good. Although the mode (peak of the density) of the importance sampling density of the ACESI algorithm is around the level x, the strata levels are far from this level. This might be a reason for the poor performance of the ACESI algorithm based on the Delta-Gamma approximation. The situation with the importance sampling density of the AOISSI algorithm is even worse, again because the lower level y of the special stratification is too far from level x, since the Delta-Gamma approximation is not good enough.

We look next at Figure 6.8 with the improved approximation, namely the modified Delta-Gamma approximation. Things get slightly better, but, as Table 6.10 shows, the performance of the AOISSI algorithm is still not satisfying, whereas the ACESI algorithm performs significantly better. In order to improve these algorithms one can either find a better approximation, or change the parametric family of densities, or increase the samples involved in the optimal reference parameter update.

We finally use the main advantage of the new algorithms, namely their flexibility in working with any approximation function, and present Table 6.10 in the heavy-tailed case, similar to Table 6.8 in the light-tailed case, which compares the ISS algorithm with our new algorithms based on the improved approximation (the one which gives the

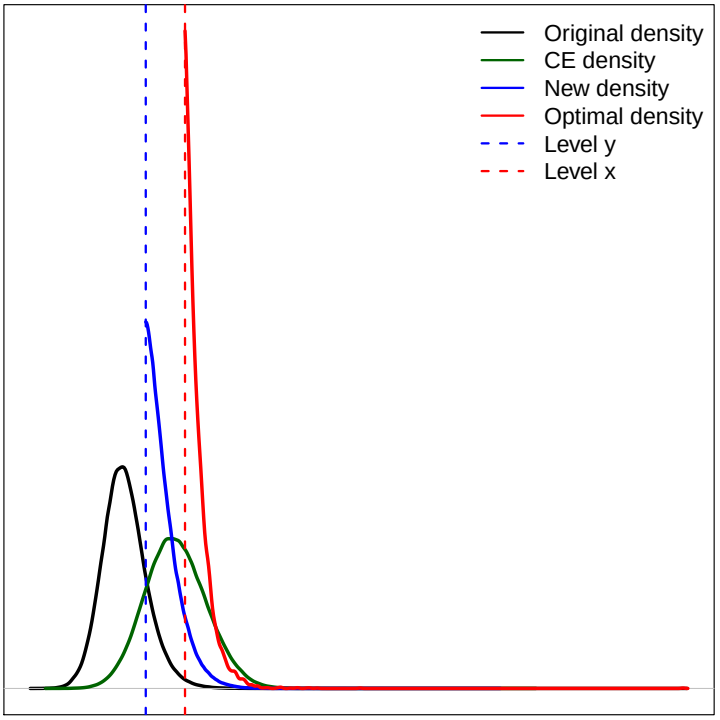


FIGURE 6.8: Comparison of ACESI and AOISSI importance sampling densities in 500000 randomly generated scenarios for the portfolio (b.5) in the heavy-tailed case for modified Delta-Gamma approximation (for details on the portfolio see Section 6.1.1).

best variance reduction rate for a particular portfolio and algorithm among four recently mentioned approximations). The approximation used for the ISS algorithm is again the Delta-Gamma approximation.

Similarly to the light-tailed case in Table 6.8, our new algorithms can be significantly improved by using some improved approximation.

We skip the rest of the tests performed in Section 6.1.3 on the comparison of approximate and full algorithms, influence of the upper level z of the special stratification on the performance, since all the comments given there, basically, hold true here. The multipliers achieved by the ACES, AICES, AOIS and AOISS algorithms are even smaller here than in the light-tailed case.

6.3 Results discussion

In this section we summarize the numerical results in the light- and heavy-tailed cases, as well as simple properties, advantages and disadvantages of all variance reduction algorithms presented in Chapter 4.

	x	$\widehat{l}$	ISS	ACESI (no z)	ACESI	AOISI	AOISSI
(a.1)	204	5.0%	11.8	22.2	33.5	43.7	46.3
	469	0.9%	34.9	96.7	122.1	132.7	148.9
	719	0.3%	69.8	285.1	382.2	264.5	298.9
(a.2)	143	5.1%	13.4	17.9	200.2	106.3	213.6
	311	1.0%	45.1	86.2	414.0	393.0	458.1
	497	0.3%	124.4	297.5	695.2	941.3	1030.0
(a.3)	149	0.9%	91.5	44.9	121.1	65.0	94.8
(a.4)	145	1.0%	43.1	71.8	449.4	217.1	448.7
(a.5)	276	5.0%	11.4	30.3	53.9	43.7	51.2
	617	1.0%	29.4	105.0	140.9	110.0	124.7
	1009	0.3%	71.3	341.5	412.4	306.6	326.0
(a.6)	5112	1.0%	77.4	143.7	383.7	184.0	309.0
(a.7)	5441	1.0%	30.2	119.7	455.6	226.8	522.2
(b.1)	415	1.0%	38.4	75.4	109.5	168.9	196.5
(b.2)	482	0.9%	57.2	51.4	153.2	78.5	114.1
(b.3)	518	1.0%	65.3	88.8	188.9	97.7	138.1
(b.4)	800	1.0%	56.8	79.6	119.9	52.4	66.0
(b.5)	835	1.0%	28.2	19.7	20.4	4.6	5.0
(b.6)	345	1.0%	52.6	24.3	27.0	34.2	35.1

TABLE 6.10: Variance reduction factors in the heavy-tailed case for the improved approximation

We first summarize that applying the variance reduction procedure based on the importance sampling together with stratification is worth it. As the numerical results show, even for very simple and highly nonlinear portfolios, the variance reduction algorithms reduce the variance by one to three orders of magnitude. Besides, the effect of the variance reduction only increases, when the probability of interest decreases.

Let us assume for now, that we are interested only in the models of changes in risk factors given by either Assumption 3.1 or by Assumption 3.2, because they are easy to use in practice. Under both these assumptions there exists already the quite easy and efficient ISS algorithm, which speeds up the MC simulation of the market risk measures. It is highly restricted to the knowledge of the distribution of the Delta-Gamma approximation, which is found under these both assumptions in, e.g., Glasserman (2004). In such a situation, why would one be interested in one of the new, more complicated, algorithms given in Chapter 4? First of all, the ACE and AICE algorithms are also quite easy to understand and implement, and they can be improved in the same way by applying the original stratification as in the ISS algorithm (under Assumptions 3.1 and 3.2). The extra variance reduction, achieved by applying the same stratification, should

be approximately the same. But the ACE and AICE algorithms can be improved by choosing a wider parametric family of densities among which the importance sampling density is searched. For example, the performance of the ACE and AICE algorithms is significantly higher in the heavy-tailed case, see Table 6.10. This gives an obvious advantage to our new algorithms even in a such restricted world as the one given by Assumptions 3.1 and 3.2.

Now assume, that we are not satisfied anymore with modeling the changes in risk factors by Assumptions 3.1 and 3.2. Then the ISS algorithm does not work. Up to our knowledge, there is no other research regarding the variance reduction algorithms as efficient as ISS in the context of market risk estimation. Then, there is an obvious value of our new algorithms, that can be applied in such a situation without any complicated adaptations.

We further guide the user to choose one particular algorithm among those given in Chapter 4. According to the performance, it should be either ACES or AOISS in the case of VaR estimation and either ACES with special stratification without the level z or AOIS in the case of CVaR estimation. As the numerical study showed, none of these algorithms dominates the other (their performance is proportional to the performance of the ACESI, AOISI and AOISSI algorithms). There are portfolios where one algorithm is better and portfolios where it is worse than other one.

In such a situation we suggest a user to use a combination of both algorithms, ACES and AOISS, in the case of VaR estimation and ACES with special stratification without the level z and AOIS in the case of CVaR estimation. This can be easily done due to their construction. They both are CE based algorithms (the choice of the importance sampling part) together with the special stratification (either with level z or without). Therefore, the importance sampling change of measure can be found by both ACE and AOIS algorithms and then the best of them is chosen (the one whose Kulback-Leibler cross entropy to the optimal importance sampling is smaller). And then the special stratification can be performed in a standard way.

The above stated only confirms all advantages of our new algorithms:

1. It is a highly flexible approach:
    - it can be applied for any type of distribution of the risk factors, no matter light- or heavy-tailed, with very little change in the corresponding code,
    - any approximation of the portfolio loss function can be used without any extra adaptations of the code (the user simply changes one function by the other),

- original stratification still can be applied in cases when it is possible (the distribution of the approximation is easy to evaluate),
- a combination of the algorithms can be easily achieved,
- the choice of the parametric family can be easily adjusted.

2. It is highly efficient:

- it outperforms the ISS algorithm in most examples and often significantly,
- alternatives to the Delta-Gamma approximation can significantly increase the variance reduction rate,
- any improvement of the ISS algorithm can also be applied here, hence, generally, it has at least the same performance,
- the improvement of the choice of the parametric family of densities can improve the performance significantly,
- the combination of the algorithms can give the best performance between two algorithms, by taking advantages of both of them.

Chapter 7

Conclusions

Calculating the probabilities of high portfolio losses is very important in real financial markets. The standard MC method, unfortunately, converges very slowly. Therefore, there is a big demand for variance reduction procedures for such an estimation.

This work guides the reader through the process of constructing new efficient and easy algorithms for reducing the variance of the MC method for the estimation of the probability of high portfolio losses. The idea is taken from Glasserman (2004), Glasserman et al. (2000a,b, 2002), namely to apply an importance sampling together with stratification procedure.

The easy and efficient ISS algorithm given by these authors is developed only for a few models of changes in risk factors, namely for normal and Student t-distributed changes. This does not allow the user of MC simulation to use this variance reduction algorithm in cases of alternative assumptions, which are, actually, the main advantage of the MC simulation algorithm.

The main accent in constructing a new algorithm is, therefore, made on the flexibility of the algorithm. The new variance reduction procedures work, without any hard modifications of the algorithms, for any distribution of the changes in risk factors, no matter if light- or heavy-tailed. The recently introduced, but already successfully applied in many applications, CE importance sampling algorithm has become a starting point in our investigation. It is easy to understand and implement, and it covers all possible models of changes in risk factors.

Based on the CE importance sampling procedure, together with the specially constructed stratification, we have introduced the ACES and AOISS algorithms and their modifications. The numerical study has shown a better performance for both these algorithms in most test cases, compared to the ISS algorithms. This fact makes these algorithms even

more attractive to the user. The same numerical study, however, did not give us a unique winner in performance among them. Therefore, we have suggested to use a combination of both algorithms, which can be easily achieved by modifying the importance sampling part. The appendixes give necessary tools in order to make the implementation process of these algorithms easier.

As we already mentioned, the probability of high portfolio losses is an important quantity in real financial markets, however, there are a few even more common measures of market risk, namely VaR and CVaR. Up to our knowledge, there is no unique literature on how to use the variance reduction algorithms for the estimation of the probability of high portfolio losses in the context of VaR and CVaR estimation. Therefore, we also describe the process of transformation of the variance reduction procedures for the estimation of the probability of high portfolio losses to the problem of VaR and CVaR estimation. This gives the user fast simulation procedures for the popular market risk measures.

Finally, we have investigated the asymptotics of all these variance reduction estimators of VaR and CVaR. We have shown the consistency and the asymptotic normality of each of them. It helped us to construct asymptotically valid confidence intervals for VaR and CVaR estimators in the context of variance reduction.

Appendix A

Sampling from Conditional Distribution

This appendix is related to the methods of sampling from the conditional distribution. Generally, when the conditional distribution is not known explicitly, special procedures are necessary. We further present two methods of sampling from the conditional density, which became useful in the variance reduction algorithms from Chapter 4.

We first define the problem. We remind, that the random vector of interest $\mathbf{X}$ is absolutely continuous (see Assumption 2.1) and has the density denoted by $f(\mathbf{x};\mathbf{u})$. In this appendix there is no need to keep the track of the parameter $\mathbf{u}$, since no change of measure is performed. Therefore, we omit it in the notations and denote the density of the vector $\mathbf{X}$ simply by $f(\mathbf{x})$.

For some real valued function $G(\mathbf{x})$ of the vector $\mathbf{x}$, of the same dimension as the random vector of interest $\mathbf{X}$, and some subset A of the real line, such that $\mathbb{P}(G(\mathbf{X}) \in A) > 0$, we are interested in methods of sampling from the conditional density of $\mathbf{X}$ given $G(\mathbf{X}) \in A$.

Such a conditional density can be written as:

$$\frac{I_A(G(\mathbf{x}))f(\mathbf{x})}{\mathbb{P}(G(\mathbf{X}) \in A)},$$

which is precisely the optimal importance sampling density under Remark 2.2. Therefore, we often do not know the normalizing constant $\mathbb{P}(G(\mathbf{X}) \in A)$, but all methods below do not involve this constant.

A.1 Brute-force acceptance-rejection method along the lines

We start with the following assumption:

Assumption A.1. We assume, that the sampling procedure from the density $f(\mathbf{x})$ is easy and known.

In this chapter we present an easy method for simulating from the conditional density of $\mathbf{X}$ given $G(\mathbf{X}) \in A$, if the simulation from the unconditional one is possible (see Assumption A.1). This method is presented in, e.g., Glasserman (2004), and is used by Glasserman et al. in the ISS algorithm (see Algorithms 4.7 and 4.8).

The idea of the method is very simple and is given in the following algorithm:

Algorithm A.1 (Brute-force acceptance-rejection method)**.**

1. *Sample* $\mathbf{X}$ *from the density* $f(\mathbf{x})$ *and evaluate* $G(\mathbf{X})$*.*
2. *If* $G(\mathbf{X}) \in A$*, return* $\mathbf{X}$*, else reiterate from step 1.*

This algorithm is a degenerate form of the classical acceptance-rejection method (see Glasserman (2004)) with the deterministic rejection decision instead of the randomized. The candidate $\mathbf{X}$ is accepted precisely if $G(\mathbf{X}) \in A$. The probability of acceptance on each attempt is, therefore, $\mathbb{P}(G(\mathbf{X}) \in A)$. The main disadvantage of this method is that, if the probability $\mathbb{P}(G(\mathbf{X}) \in A)$ is small, many simulations are wasted in order to achieve just one sample from the conditional distribution. This drawback makes this method inapplicable in many situations.

This method is of a particular value for the ISS algorithm (see Algorithms 4.7 and 4.8), since it is used in some modified way, that makes it more practical. More precisely, for some disjoint partition of the real line $A_1, \ldots, A_K$, the samples from all conditional densities of $\mathbf{X}$ given $G(\mathbf{X}) \in A_i$, for $i = 1, \ldots, K$ are necessary.

Assume that we are interested in samples of sizes $N_1, \ldots, N_K$ for each conditional density of $\mathbf{X}$ given $G(\mathbf{X}) \in A_i$, $i = 1, \ldots, K$, correspondingly. The brute-force acceptance-rejection method of achieving such samples reads as follows:

Algorithm A.2 (Brute-force acceptance-rejection method for disjoint partition)**.**

1. *Initialize stratum counters* $n_i = 0$*,* $i = 1, \ldots, K$*.*
2. *Sample* $\mathbf{X}$ *from the density* $f(\mathbf{x})$ *and evaluate* $G(\mathbf{X})$*.*

3. *Find j such that $G(\mathbf{X}) \in A_j$.*

4. *If $n_j < N_j$, increase the counter $n_i \leftarrow n_i + 1$ and set $\mathbf{X}_{jn_j} = \mathbf{X}$, else proceed with step 5.*

5. *If all $n_i = N_i$, $i = 1, \ldots, K$, return the samples $\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iN_i}$ for all $i = 1, \ldots, K$, else reiterate from step 2.*

The main difference of Algorithm A.2, compared to Algorithm A.1, is that we discard the generated vector $\mathbf{X}$ only if it fell in the stratum which is already full. Therefore, e.g. in the case of equiprobable strata:

$$\mathbb{P}(G(\mathbf{X}) \in A_1) = \ldots = \mathbb{P}(G(\mathbf{X}) \in A_K) = \frac{1}{K},$$

with equal sample sizes $N_1 = N_2 = \ldots = N_K$, only few generated vectors $\mathbf{X}$ are wasted. This problem is discussed in more details in Glasserman et al. (2000b).

A.2 Hit - and - Run sampler with Ratio - of - Uniforms

As we already mentioned in Appendix A.1, the brute-force acceptance-rejection Algorithm A.1 can not be used to sample from the conditional density with small probability of the event $\mathbb{P}(G(\mathbf{X}) \in A)$. In particular, in order to implement the ICE based algorithm AICE, we need to sample from the optimal importance sampling density of the form (see Algorithm 4.5):

$$g^*(\mathbf{x}) = \frac{I_{(x,+\infty)}(Q(\mathbf{x}))f(\mathbf{x};\mathbf{u})}{\mathbb{P}(Q(\Delta\mathbf{S}) > x)},$$

where the event of conditioning $Q(\Delta\mathbf{S}) > x$ has the probability usually around 0.01. This means that, on average, only one out of a hundred generated values is accepted, according to Algorithm A.1. Other more efficient method has to be introduced in order to handle such situations.

The special combination of Hit-and-Run sampler (see e.g. Gilks et al. (1995)), together with Ratio-of-Uniforms method (see e.g. Gilks et al. (1995)), is introduced by the authors in Karawatzki et al. (2005). The corresponding algorithm is called HITRO and is defined in some variants. In the paper by Karawatzki et al. (2005) is shown, that this algorithm is well suited to generate high dimensional vectors $\mathbf{Y}$ from quite arbitrary densities $g(\mathbf{x})$ (it is developed not only to sample from the conditional densities). Already implemented and tested libraries in different programming languages by Leydold et al. (2005) make the usage of this method even more attractive. In the numerical study in

Chapter 6 we have used the corresponding package in the mathematical programming language R Project.

Now we are going to present the HITRO algorithm from Karawatzki et al. (2005) in more details. Since the MCMC methods are not the matter of our study, we present here only the most important information in order to get a first impression about the method used. An interested reader can appeal to the original paper by Karawatzki et al. (2005) and the literature cited therein, as well as to Gilks et al. (1995), to achieve a deeper knowledge about MCMC methods, including HITRO algorithm.

Since the HITRO algorithm is based on the Hit-and-Run sampler together with the Ratio-of-Uniforms method, it is useful first to define both these algorithms. We start from the Hit-and-Run sampler which is an iterative algorithm used to sample d-dimensional random vectors with density $g(\mathbf{x})$ in some fixed, but arbitrary, bounded open set S (see Gilks et al. (1995), Karawatzki et al. (2005)). At each iteration j, the Hit-and-Run algorithm first randomly samples the direction $\mathbf{d_j}$ (a unit d-dimensional vector). Then, the next value $\mathbf{Y}_{j+1}$ is chosen according to the full conditional distribution along the straight line passing through $\mathbf{Y}_j$ in the direction $\mathbf{d_j}$, more precisely:

Algorithm A.3 (Hit-and-Run)**.**

1. *Choose a starting d-dimensional vector $\mathbf{Y}_0 \in S$ and set $j = 0$.*
2. *Generate a random direction $\mathbf{d_j}$ uniformly on the $(d+1)$-sphere.*
3. *Generate $\lambda_j \in \Lambda_j = \{\lambda : \mathbf{Y}_j + \lambda \mathbf{d}_j \in S\}$ from the density $g_j(\lambda)$ defined by:*

$$g_j(\lambda) = \frac{g(\mathbf{Y}_j + \lambda \mathbf{d}_j)}{\int_{\Lambda_j} g(\mathbf{Y}_j + u\mathbf{d}_j)} du.$$

4. *Set $\mathbf{Y}_{j+1} = \mathbf{Y}_j + \lambda_j \mathbf{d}_j$ and $j = j + 1$.*
5. *If next value is necessary, reiterate from the step 2, else stop.*

The Ratio-of-Uniforms algorithm (see Gilks et al. (1995), Karawatzki et al. (2005)) is also introduced to generate d-dimensional random vectors with density $g(\mathbf{x})$. It first generates a $(d+1)$-dimensional vector uniformly distributed over the region:

$$A(g) = A_{r,\mathbf{m}}(g) = \{(\mathbf{y}, z) : 0 < z < \sqrt[rd+1]{g(\mathbf{y}/z^r + \mathbf{m})}\},$$

for some $r > 0$ and a d-dimensional vector $\mathbf{m}$. It can be done by sampling from some minimal bounding rectangular which contains $A(g)$ and then accept the generated vector if it belongs to $A(g)$. Otherwise, the next vector should be generated. Therefore,

it is important, that the region $A(g)$ is bounded. The achieved generated vector is recomposed accordingly to construct the random vector from the density $g(\mathbf{x})$. More precisely, the algorithm reads as follows:

Algorithm A.4 (Ratio-of-Uniforms)**.**

1. *Fix input parameters* $r > 0$ *and* $\mathbf{m}$ *(d-dimensional vector).*
2. *Generate* $(d+1)$*-dimensional vector* $(\mathbf{U}, V)$ *uniformly in* $A_{r,\mathbf{m}}(g)$*.*
3. *Set* $\mathbf{Y} = \mathbf{U}/V^r + \mathbf{m}$*.*

These both methods perform very slow in high-dimensional cases and for some complex density functions. The combination of these methods with some modifications is, therefore, introduced by Karawatzki et al. (2005) in order to construct a fast and easy sampling procedure from basically every density function. The following version of this method, namely the HITRO-plate (see Karawatzki et al. (2005)), is of particular interest for us (we call it simply HITRO):

Algorithm A.5 (HITRO)**.**

1. *Fix input parameters:* $g(\mathbf{x})$ *- the density function to sample from, its mode* $\mathbf{m}$ *(d-dimensional vector),* $r > 0$ *and the sample size* N *(number of generated vectors).*
2. *Compute* $g_m = g(\mathbf{m})$*. Set* $(\mathbf{U_0}, V_0) = (\mathbf{0}, 0.5)$ *and* $j = 0$*.*
3. *Generate a random direction* $\mathbf{d_j} = (\mathbf{d}_u, d_v)$ *uniformly on the* $(d+1)$*-sphere. Set* $\lambda_{j0} = -|v/d_v|$ *and* $\lambda_{j1} = |(1-v)/d_v|$*.*
4. *Generate* λ_j *uniformly distributed on the interval* $(\lambda_{j0}, \lambda_{j1})$*.*
5. *Set* $(\mathbf{U_{j+1}}, V_{j+1}) = (\mathbf{U_j}, V_j) + \lambda_j \mathbf{d_j}$*.*
6. *Set* $\mathbf{Y}_{j+1} = \mathbf{U_{j+1}}/V_{j+1}^r + \mathbf{m}$*.*
7. *If* $(rd+1)\log(V_{j+1}) \geq \log(g(\mathbf{Y}_{j+1})/g_m)$*, then set* $\lambda_{j0} = \lambda_j$ *if* $\lambda_j < 0$ *or* $\lambda_{j1} = \lambda_j$ *otherwise and reiterate from step 4, else set* $j = j+1$*.*
8. *If* $j < N$*, then reiterate from step 3, else stop.*

Remark A.1. *The density function* $g(\mathbf{x})$ *need not be normalized. This means, that any positive multiple of the density can be used in Algorithm A.5.*

Remark A.2. *The mode* $\mathbf{m}$ *in the input parameters can be replaced by any point* $\mathbf{c}$*, which represents the center of the distribution. This can be seen as choosing the vector* $\mathbf{c}$ *such that* $g(\mathbf{c})$ *is close to the maximum of the density.*

Since the algorithm has no restriction on the density function $g(\mathbf{x})$, it can be used for the conditional density of interest, namely:

$$g(\mathbf{x}) = \frac{I_A(G(\mathbf{x}))f(\mathbf{x})}{\mathbb{P}(G(\mathbf{X}) \in A)}.$$

More precisely, according to Remark A.1, we do not need to know the constant $\mathbb{P}(G(\mathbf{X}) \in A)$ and can use for sampling the simpler function $I_A(G(\mathbf{x}))f(\mathbf{x})$. In the numerical study in Chapter 6 we use Algorithm A.5 with $r = 1$ and the center instead of the mode of the density (see Remark A.2), in order to sample from the conditional densities.

As authors in Karawatzki et al. (2005) conclude, the HITRO algorithm is simple and easy to implement, relatively fast, and works for almost any arbitrary density function. It is proven there, that the HITRO algorithm works optimal for distributions whose region of acceptance is convex. This is, in particular, the case for all log-concave distributions (see An (1995)) under the choice of the parameter $r = 1$. Log-concave distributions include a wide class of distributions. Among them are the Uniform, Gaussian, Exponential, Gamma, Weibull and many others. The Student t-distribution does not belong to this family, however, it does not mean that the HITRO algorithm does not work in a such case. The numerical study in the multidimensional Gaussian, Cauchy and Student t-cases, given in Karawatzki et al. (2005), shows a great performance. Even in dimension 100 the number of iterations in Algorithm A.5 stays below 7, which is important due to the heavy revaluation of the portfolio loss function.

Appendix B

Simulated annealing algorithm for multi-dimensional optimization

In this appendix we present a method for solving the multi-dimensional optimization problem in a large space. A class of simulated annealing algorithms given by Belisle (1992) shows a great practical value in solving some optimizations in the numerical study in Chapter 6. This class of algorithms belongs to the class of stochastic global optimization methods and is a version of the Monte Carlo technique. It uses only function values and works also for non-differentiable functions. Further, we briefly present the simulated annealing algorithm. For more details we refer an interested reader to Belisle (1992) and the literature cited therein.

Assume that we are interested in minimizing the real-valued function V. The simulated annealing algorithm is a MCMC algorithm, which generates the states $(\mathbf{X}_1, \ldots, \mathbf{X}_N, \ldots)$ in parallel to the random sequence of candidate points $(\mathbf{Y}_1, \ldots, \mathbf{Y}_N, \ldots)$ and non random sequence of positive temperatures $(T_1, \ldots, T_N, \ldots)$. The minimum of the function V is then calculated as a limit of the sequence $(V(\mathbf{X_i}), i \geq 0)$.

If the sequences $(\mathbf{X}_1, \ldots, \mathbf{X}_N)$, $(\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$ and $(T_1, \ldots, T_N)$ are already constructed, the next values of $\mathbf{X}_{N+1}$, $\mathbf{Y}_{N+1}$ and T_{N+1} are set as follows:

$$\mathbf{X}_{N+1} = \begin{cases} \mathbf{Y}_{N+1}, & \text{with probability } p(\mathbf{X}_N, \mathbf{Y}_N, T_N) \\ \mathbf{X}_N, & \text{with probability } 1 - p(\mathbf{X}_N, \mathbf{Y}_N, T_N) \end{cases},$$

where $\mathbf{Y}_{N+1}$ is simulated from the Gaussian Markov kernel $R(\mathbf{X}_N, \cdot)$ ($R(\cdot, \cdot)$ is a Markov kernel if $R(\mathbf{x}, \cdot)$ is a probability measure and $R(\cdot, A)$ is a measurable function) with scale proportional to the actual temperature T_N, and the probability $p(\mathbf{X}_N, \mathbf{Y}_N, T_N)$ is defined

by:

$$p(\mathbf{x},\mathbf{y},t) = \begin{cases} \exp\left[-\frac{V(\mathbf{y})-V(\mathbf{x})}{t}\right], & \text{if } V(\mathbf{y}) > V(\mathbf{x}) \\ 1 & \text{if } V(\mathbf{y}) \leq V(\mathbf{x}) \end{cases},$$

and the temperature are decreased according to the logarithmic cooling schedule given by Belisle (1992):

$$T_{N+1} = \frac{T_0}{\log\left(\left[\frac{N}{T_{max}}\right] * T_{max} + e\right)},$$

where $[\cdot]$ denotes the integer part and T_{max} is the maximum number of function evaluations at each temperature given by the user in the input.

This algorithm can be very useful in getting to a good value on a very rough surface. It is a standard theory of the multi-dimensional optimization, which means, that the various mathematical packages already have it implemented. We used it as part of the standard optimization package in the mathematical programming language R Project.

Appendix C

Options pricing formulas

In this Appendix we define the options used in the numerical study in Chapter 6 and present their pricing formulas as given in, e.g., Hull (2012b), Korn and Korn (2001). Since it is not the main purpose of the thesis to discuss the options in details, we present here the material from Hull (2012b), Korn and Korn (2001) only briefly.

A call option gives the holder of the option the right to buy an asset at a certain date (maturity) for a certain price (strike price). The put option is defined in a similar way, but giving the right to sell the asset. By the call and the put options we mean here the European type options, i.e. the options that can be exercised only at the maturity.

For every option contract there are basically two sides, the investor who has taken the long position, i.e. has bought the option, and the investor who has taken the short position, i.e. has sold (written) the option. The call and the put options, depending on the positions have the following payoffs:

Call, long: $\max(S_T - K, 0)$,

Call, short: $-\max(S_T - K, 0) = \min(K - S_T, 0)$,

Put, long: $\max(K - S_T, 0)$,

Put, short: $-\max(K - S_T, 0) = \min(S_T - K, 0)$,

where T denotes the maturity, S_T denotes the asset price at the maturity time T and K denotes the strike price.

No matter which model is used, we assume, that the options are priced according to the standard Black-Scholes formulas and their extensions (this is often done in practice, as

mentioned in Glasserman (2004), Glasserman et al. (2000a,b, 2002)). We further just give the Black-Scholes formulas for the prices of call and put options.

The call price at time t is given by:

$$c(t) = S_0\Phi(d_1(t)) - Ke^{-r(T-t)}\Phi(d_2(t)),$$

and the put price at time t is given by:

$$p(t) = Ke^{-r(T-t)}\Phi(-d_2(t)) - S_0\Phi(-d_1(t)),$$

where S_0 denotes the starting asset price, r denotes the risk-free interest rate, $\Phi(x)$ is the cumulative distribution function of the standard normal distribution $N_1(0,1)$, the value $d_1(t)$ is given by:

$$d_1(t) = \frac{\log(S_0/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}},$$

and the value $d_2(t)$ is given by:

$$d_2(t) = \frac{\log(S_0/K) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} = d_1(t) - \sigma\sqrt{T-t},$$

where σ denotes the annual volatility.

We next discuss the down-and-out call, which is a particular type of the knock-out barrier option, i.e. it is a regular call option, that ceases to exist when the underlying asset price reaches a certain barrier level H, which is below the initial asset price. If $H \leq K$, then the down-and-out call price at time t is given by:

$$c_{do}(t) = c(t) - [S_0(H/S_0)^{2\lambda}\Phi(y(t)) - Ke^{-r(T-t)}(H/S_0)^{2\lambda-2}\Phi(y(t) - \sigma\sqrt{T-t})],$$

where the value λ is given by:

$$\lambda = \frac{r + \sigma^2/2}{\sigma^2},$$

and the value $y(t)$ is given by:

$$y(t) = \frac{\log(H^2/(S_0K))}{\sigma\sqrt{T-t}} + \lambda\sigma\sqrt{T-t}.$$

Finally, the cash-or-nothing put is a simple example of a binary option, i.e. it pays off Q, if the asset price is below the strike price K and nothing if it is above the strike price

K. The value of the cash-or-nothing put at time t is, therefore, given by:

$$p_{co}(t) = Qe^{-r(T-t)}\Phi(-d_2(t)),$$

with the same value $d_2(t)$ as for regular call and put.

For other types of options and their pricing formulas we refer an interested reader to Hull (2012b), Korn and Korn (2001).

Bibliography

C. Albanese, K. Jackson, and P. Wiberg. Dimension Reduction in the Computation of Value-at-Risk. *The Journal of Risk Finance*, 2002.

C. Alexander. *Market Risk Analysis: Quantitative Methods in Finance*, volume I. John Wiley & Sons, Ltd., Chichester, 2008a.

C. Alexander. *Market Risk Analysis: Practical Financial Econometrics*, volume II. John Wiley & Sons, Ltd., Chichester, 2008b.

C. Alexander. *Market Risk Analysis: Pricing, Hedging and Trading Financial Instruments*, volume III. John Wiley & Sons, Ltd., Chichester, 2008c.

C. Alexander. *Market Risk Analysis: Value-at-Risk Models*, volume IV. John Wiley & Sons, Ltd., Chichester, 2008d.

M. An. Log-concave Probability Distributions:Theory and Statistical Testing. Working Paper, 1995. URL `http://ssrn.com/abstract=1933`.

T. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Inc., New Jersey, 3rd edition, 2003.

P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9:203–228, 1999.

S. Asmussen and P. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer, 2007.

R. Bahadur. A note on quantiles in large samples. *Annals of Mathematicsl Statistics*, 37:577–580, 1966.

C. Belisle. Convergence Theorems for a Class of Simulated Annealing Algorithms on R^d. *Journal of Applied Probability*, 29(4):885–895, 1992.

P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 3rd edition, 1995.

J. Chan and D. Kroese. Improved cross-entropy method for estimation. *Statistics and Computing*, 22:1031–1040, 2012.

J. Chan, P. Glynn, and D. Kroese. A Comparison of Cross-Entropy and Variance Minimization Strategies. *Journal of Applied Probability*, 48A:183–194, 2011.

F. Chu. *Variance Reduction Techniques for Estimating Quantiles and Value-at-Risk.* PhD thesis, Faculty of New Jersey Institute of Technology, Department of Information Systems, New Jersey, May 2010.

F. Chu and M. Nakayama. Confidence Intervals for Quantiles When Applying Variance-Reduction Techniques. *ACM Transactions On Modeling and Computer Simulation*, 36(Article 7):(25 pages plus 12–page online–only appendix), 2012.

K. Dowd. *An Introduction to Market Risk Measurement.* John Wiley & Sons, Ltd., Chichester, 2002.

L. Ederington and W. Guan. Higher Order Greeks. *The Journal of Derivatives*, 14(3): 7–34, 2007.

J. Ghosh. A new proof of the Bahadur representation of quantiles and an application. *Annals of Mathematicsl Statistics*, 42:1957–1961, 1971.

W. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice.* Chapman & Hall, 1995.

P. Glasserman. *Monte Carlo Methods in Financial Engineering.* Springer-Verlag, Inc., New York, 2004.

P. Glasserman, P. Heidelberger, and P. Shahabuddin. Importance Sampling and Stratification for Value-at-Risk. *Computational Finance 1999 (Proceedings of the Sixth International Conference on Computational Finance), Y.S. Abu-Mostafa, B. LeBaron, A.W. Lo, and A.S. Weigend, eds., MIT Press, Cambridge, Mass.*, pages 7–24, 2000a.

P. Glasserman, P. Heidelberger, and P. Shahabuddin. Variance reduction techniques for estimating Value-at-Risk. *Management Science*, 46:1349–1364, 2000b.

P. Glasserman, P. Heidelberger, and P. Shahabuddin. Portfolio Value-at-Risk with heavy-tailed risk factors. *Mathematical Finance*, 12:239–269, 2002.

P. Glynn. Importance sampling for Monte Carlo estimation of quantiles. *Proceedings of 1996 Second International Workshop on Mathematical Methods in Stochastic Simulation and Experimental Design, St. Petersburg, Russia*, pages 180–185, 1996.

L. Hong and G. Liu. Monte Carlo estimation of Value-at-Risk, Conditional Value-at-Risk and their sensitivities. In J. Himmelspach K. White S. Jain, R. Creasey and M. Fu, editors, *Proceedings of the 2011 Winter Simulation Conference*, pages 95–107, 2011.

J. Hull. *Risk Management and Financial Institutions*. John Wiley & Sons, Inc., Hoboken, New Jersey, 3rd edition, 2012a.

J. Hull. *Options, Futures, and Other Derivatives*. Prentice Hall, Boston, 8th edition, 2012b.

P. Jorion. *Value-at-Risk: the new benchmark for managing financial risk*. McGraw-Hill, New York, 2nd edition, 2001.

R. Karawatzki, J. Leydold, and K. Potzelberger. Automatic Markov chain Monte Carlo procedures for sampling from multivariate distributions. Research Report Series, 2005. URL `http://epub.wu.ac.at/1400/1/document.pdf`.

J. Kiefer. On Bahadur's representation of sample quantiles. *Annals of Mathematicsl Statistics*, 38:1350–1353, 1967.

R. Korn and E. Korn. *Option Pricing and Portfolio Optimization: Modern Methods of Financial Mathematics*, volume 31. Graduate Studies in Mathematics, American Mathematical Society, 2001.

R. Korn and M. Pupashenko. A New Variance Reduction Technique for Estimating Value-at-Risk. *Applied Mathematical Finance*, 2014. URL `http://ssrn.com/abstract=2402602`. Submitted.

R. Korn, E. Korn, and G. Kroisandt. *Monte Carlo Methods and Models in Finance and Insurance*. Taylor and Francis Group, LLC, 2010.

J. Leydold, W. Hörmann, E. Janka, R. Karawatzki, and G. Tirler. UNU.RAN - A Library for Non-Uniform Universal Random Variate Generation, 2005. URL `http://statistik.wu-wien.ac.at/unuran/`.

A. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, 2005.

G. Pflug. *Some remarks on the value-at-risk and conditional value-at-risk*, chapter In Probabilistic Constrained Optimization: Methodology and Applications, edited by S. Uryasev, pages 272–281. Kluwer, Dordrecht, The Netherlands, 2000.

M. Pupashenko. Minimizing Variance and Cross-Entropy Importance Sampling Methods for Rare Event Simulation. Master thesis, University of Kaiserslautern, Kaiserslautern, August 2011.

M. Pupashenko. Variance Reduction Technique for Estimating Value-at-Risk based on the Cross-Entropy. *Journal of Mathematics and System Science*, 4(1):37–48, 2014.

Y. Rakotondratsimba. Modified delta-gamma approximation, April 2009. URL `http://ssrn.com/abstract=1395969`.

R. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000.

R. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26:1443–1471, 2002.

R. Rubinstein and P. Glynn. How to Deal with the Curse of Dimensionality of Likelihood Ratios in Monte Carlo Simulation. *Stochastic Models*, 25(4):547–568, 2009.

R Rubinstein and D. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization Monte Carlo Simultaion, and Machine Learning*. Springer, New York, 2004.

R Rubinstein and D. Kroese. *Simulation and the Monte Carlo Method*. John Wiley & Sons, Inc., 2nd edition, 2008.

R. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc., New York, 1980.

L. Sun and L. Hong. A general framework of Importance Sampling for Value-at-Risk and Conditional Value-at-Risk. In B. Johansson A. Dunkin M. Rossetti, R. Hill and R. Ingalls, editors, *Proceedings of the 2009 Winter Simulation Conference*, pages 415–422, 2009.

L. Sun and L. Hong. Asymptotic representations for importance-sampling estimators of value-at-risk and conditional value-at-risk. *Operations Research Letters*, 38:246–251, 2010.

A. Trindade, S. Uryasev, A. Shapiro, and G. Zrazhevsky. Financial prediction with constrained tail risk. *Journal of Banking & Finance*, 31:3524–3538, 2007.

Wissenschaftlicher Werdegang

09.2000 - 06.2004	Allgemeine Hochschulreife an Kiew Naturwissenschaftliches Gymnasium Nr. 145, Ukraine
09.2004 - 06.2008	Bachelor-Studium an der Nationalen Taras-Schewtschenko-Universität Kiew, Ukraine
09.2008 - 06.2010	Master-Studium an der Nationalen Taras-Schewtschenko-Universität Kiew, Ukraine
10.2009 - 08.2011	Master-Studium an der Technischen Universität Kaiserslautern
10.2011 - 07.2014	Doktorand bei Prof. Dr. Ralf Korn an der Technischen Universität Kaiserslautern

Fachliche Veröffentlichungen

2014 R.Korn, und **M.Pupashenko** - *A New Variance Reduction Technique for Estimating Value-at-Risk*, Applied Mathematical Finance (eingereicht).

2014 **M. Pupashenko** - *Variance Reduction Technique for Estimating Value-at-Risk based on Cross-Entropy*, Journal of Mathematics and System Science, Nr. 4, Seiten 37-48.

2013 P. Ruckdeschel, M. Kohl, N. Horbenko, G. Kroisandt, D. Pupashenko, und **M. Pupashenko** - *RobExtremes: Optimally robust estimation for extreme value distributions (v. 0.9)*, Open Source Software - R-package published on r-forge.

2011 **M. Pupashenko** - *Convergence of reward functionals in a reselling model for a European option*, Theor. Probability and Math. Statist., no.83, pp. 135-148.

2008 **M. Pupashenko**, und A. Kukush - *Reselling of European option if the implied volatility varies as CIR process*, Theory of Stochastic Processes, vol.14 (30), no.3-4, pp.114-128.

2007 A. Kukush, und **M. Pupashenko** - *Bounds for a sum of random variables under a mixture of normals*, Theory of Stochastic Processes, vol.13 (29), no.4, pp.82-97.

2006 J. Dhaene, A. Kukush, und **M. Pupashenko** - *On the characterization of premium principle with respect to pointwise comonotonicity*, Theory of Stochastic Processes, vol.12 (28), no.3-4, pp.26-42.

Scientific Background

09.2000 - 06.2004	High school at the Scientific High School Nr. 145 in Kiev, Ukraine
09.2004 - 06.2008	Bachelor-Studies at the National Taras-Schewtschenko-University Kiev, Ukraine
09.2008 - 06.2010	Master-Studies at the National Taras-Schewtschenko-University Kiev, Ukraine
10.2009 - 08.2011	Master-Studies at the University of Kaiserslautern
10.2011 - 07.2014	PhD student with supervisor Prof. Dr. Ralf Korn at the University of Kaiserslautern

Publications

2014 R.Korn, and **M.Pupashenko** - *A New Variance Reduction Technique for Estimating Value-at-Risk*, Applied Mathematical Finance (submitted).

2014 **M. Pupashenko** - *Variance Reduction Technique for Estimating Value-at-Risk based on Cross-Entropy*, Journal of Mathematics and System Science, Nr. 4, Pages 37-48.

2013 P. Ruckdeschel, M. Kohl, N. Horbenko, G. Kroisandt, D. Pupashenko, and **M. Pupashenko** - *RobExtremes: Optimally robust estimation for extreme value distributions (v. 0.9)*, Open Source Software - R-package published on r-forge.

2011 **M. Pupashenko** - *Convergence of reward functionals in a reselling model for a European option*, Theor. Probability and Math. Statist., no.83, pp. 135-148.

2008 **M. Pupashenko**, and A. Kukush - *Reselling of European option if the implied volatility varies as CIR process*, Theory of Stochastic Processes, vol.14 (30), no.3-4, pp.114-128.

2007 A. Kukush, and **M. Pupashenko** - *Bounds for a sum of random variables under a mixture of normals*, Theory of Stochastic Processes, vol.13 (29), no.4, pp.82-97.

2006 J. Dhaene, A. Kukush, and **M. Pupashenko** - *On the characterization of premium principle with respect to pointwise comonotonicity*, Theory of Stochastic Processes, vol.12 (28), no.3-4, pp.26-42.