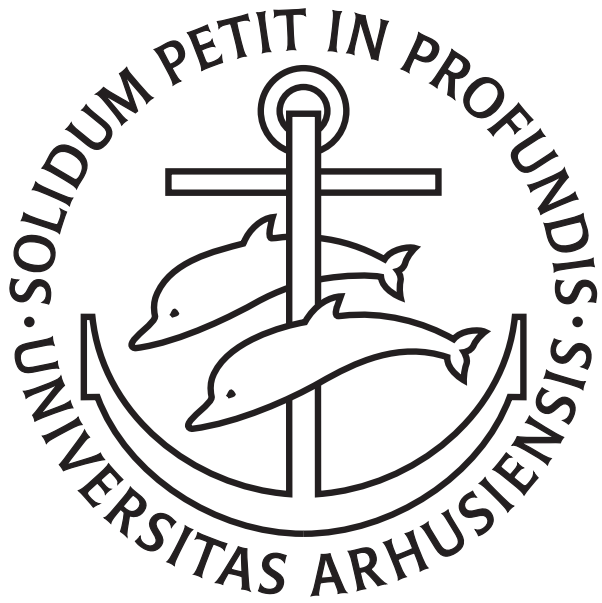
Risk Probabilities: Asymptotics and Simulation

PhD Dissertation



Leonardo Rojas Nandayapa

Advisor: Søren Asmussen

September 2008
Department of Mathematical Sciences
Faculty of Sciences University of Aarhus

To my parents

Contents

Preface

Back in 2004, I approached my master thesis supervisor Mogens Bladt looking for some advice to start a PhD in Probability. After a few months, several emails and a lot of paperwork I found myself packing to go to Denmark to work in Applied Probability under the supervision of Søren Asmussen. This dissertation resumes my work during the past four years as a PhD student at the Department of Mathematical Sciences at the University of Aarhus.

In the span of those four years I have worked on a set of problems presented to me by Søren Asmussen. The common idea was to find approximations – analytic expressions and Monte Carlo estimates – for quantities involving sums of heavy-tailed random variables possibly dependent. Problems of this nature are of major importance in numerous fields, however their solutions are often difficult. Working on them have represented a huge, although very rewarding challenge for me.

The process that led to obtaining the material in this dissertation has been slow – specially in the beginning. However, Søren Asmussen was always there – three times per week except when traveling – to give me invaluable help and advice and encourage me to work harder. By bringing me in contact with the latest advances in the field and prominent researchers, he created the optimal environment for me to realize my PhD studies. Beyond his excellent skills as supervisor, I have learned from him a great deal of things involving cooking and gardening among other stuff. It was always a great joy to visit him and his wife May Lise at their home in Egtved. Therefore, it is a great pleasure for me to express my gratitude for his hospitality, support and supervision during these four years.

I wish to thank José Blanchet for creating an enjoyable atmosphere at the IEOR Department at Columbia University during my visit in the spring of 2008. I have benefited enormously from our discussions which allowed me to gain a deeper understanding of diverse areas. I am also grateful to Hansjörg Albrecher, Sandeep Juneja, Tomasz Rolski, Henning Bunzel and Bent Jesper Christensen for joint work at different stages.

I am deeply indebted to the staff of the Department of Mathematical Sciences at the Aarhus University for all the support in different matters during my time there: Oddbjørg Wethelund who has helped me on uncountable times and situations; Lars "daleif" Madsen who helped me with the design of this document and from whom I have learned LaTeX (although I still cannot read his book in Danish); Randi Mosegaard who reviewed the linguistics in some parts and patiently guided me through the submission process.

I also want to express my gratitude to all my teachers and professors at the IIMAS-UNAM and the ITAM – the institutions in México where I realized my master and undergraduate studies, respectively. Special thanks go to Mogens Bladt, Begoña Fernández, Juan José Fernández-Durán and Alberto Tubilla for encouraging conversations.

I wish to thank all my friends from ITAM, Teknolog and Vennelyst for providing the

ideal atmosphere for the last eleven years. For sharing everyday stories and cheering me up via mail, messenger and phone during the last four years I will always be grateful to Luis, Carlos, Mariana and Marco. Also special thanks go to Victor and Juan from whom I learned the basics of Matlab – which turned out to be extremely useful for doing my research.

Financial support from CONACYT is kindly acknowledged. I also thank the *Thiele Centre* for supporting my traveling.

Finally, I want to thank to my family for the outstanding support and affection shown through all these years in spite of the distance. In particular to my mom who always listened and spirited me up through the phone.

Leonardo Rojas Nandayapa
Aarhus, September 2008

Summary

In the material revised in this dissertation the common and main objects of interest will be tail probabilities of a sum S_n of heavy-tailed random variables. More precisely,

$$\mathbb{P}(S_n > u), \qquad u \to \infty.$$

These probabilities are of a major importance in various branches of Applied Probability, such as Risk Theory, Queueing Theory and Financial Management, and are subject of intense research nowadays. To understand their relevance one just needs to think of insurance companies facing losses due to natural disasters, banks seeking protection against huge losses, failures in expensive and sophisticated systems or loss of valuable information in electronic systems.

The main difficulty when dealing with this kind of problems is the unavailability of a closed analytic expression for the distribution function of a sum of random variables. The presence of heavy-tailed random variables complicates even more the problem. The objective of this dissertation is to provide better approximations by means of sharp asymptotic expressions and Monte Carlo estimators. By doing so, we will obtain a deeper insight into how events involving large values of sums of heavy-tailed random variables are likely to occur.

The objective of the first Chapter is to provide some additional motivation for the study of stochastic models with heavy-tails and introduce the main tools that will be used for its analysis. After establishing some basic concepts and notation, two of the key ingredients in this dissertation are presented and discussed: *Heavy-tailed Distributions* and *Monte Carlo Methods for Rare Events*.

Probabilities of Ruin are studied in the Chapter 2. A brief introduction to *Risk processes* and *Queueing process* is given there making emphasis on how these processes are related to a sum of random variables. This is followed by a a short revision of the literature available on the approximation of *Probabilities of Ruin* – sharp asymptotic expressions and Monte Carlo estimators. Finally, two novel Monte Carlo estimators are proposed for approximating the probability of ruin in a finite time horizon in one of the most important heavy-tailed settings: *Regular Variation*.

Tail probabilities of sums of *correlated lognormal* random variables are analyzed in the Chapter 3. Their joint distribution is derived from an exponential transformations of the components of a multivariate normal vector. Therefore, it is consider an essential model for dependent heavy-tailed phenomena. A sharp asymptotic expression is derived for the tail probability of the sum correlated lognormals – a result which directly relates them with the class of Subexponential distributions in the i.i.d. setting. Moreover, this approach will provide a useful insight into the occurrence of large values of the sum. Two sets of Monte Carlo estimators with excellent numerical and theoretical properties are given next.

Tail probabilities of the sum of the components of a *Log-elliptical* random vector are considered in the Chapter 4. These vectors are obtained by an exponential transformation of random vectors with *elliptical distributions* – a large class of multivariate distributions which includes the *multivariate normal*, the *multivariate-t*, *normal mixtures* and *generalized hyperbolics* among others. After a brief introduction to this class of distributions, a Monte Carlo estimator for the tail probability of the sum is proposed and analyzed. In particular, for the multivariate lognormal case it is shown to have optimal theoretical properties.

A brief analysis of the *small tail probability* $\mathbb{P}(S_n < u)$ involving i.i.d. lognormal random variables is presented in the Chapter 5. After some motivation for the study of this problem, it will be discussed how the *Laplace-Stieltjes transform* is used to solve this kind of problems. A sharp approximation for the *Laplace-Stieltjes transform* is proposed, analyzed and used for the construction of two Monte Carlo estimators.

The methodology of an empirical study of the economical value of *realized covariance* in investment decisions can be found in the Chapter 6. A general overview of *realized volatility* – a volatility estimator designed for high-frequency data – is provide first including a discussion of some of the proposals found in the literature to overcome the problems of non-synchronicity, incomplete data for the whole day and data contaminated by noise which are characteristic of high-frequency financial data. Then it is suggested to merge some of these proposals to obtain a *realized covariance for the whole day* and measure its value in investment decisions.

Some relevant, but standard material has been collected in the first three Appendices. Numerical examples of the methods suggested across the dissertation were collected in the last Appendix, while its discussion was kept in the original chapters.

For an easier identification, those pieces of original research obtained during my time working in this dissertation were labeled either as *Theorems*, *Lemmas* or *Corollaries*. Results taken from the literature or deducted from known results were labeled as *Propositions*. The following is a summary of these results. The Theorems at the end of Chapter 1 are corrected versions of the results which appeared first in Asmussen and Rojas-Nandayapa (2006) and later in Rojas-Nandayapa (2006) (these are extensions of the original work of Asmussen and Binswanger, 1997; Asmussen and Kroese, 2006). The sharp asymptotic expression studied in chapter 2 was derived in Asmussen and Rojas-Nandayapa (2008). The first set of algorithms proposed in Chapter 3 were proposed and analyzed in Asmussen et al. (2008). The second set of algorithms is partly extracted from Asmussen and Rojas-Nandayapa (2006) and represents work in progress. The material in Chapter 4 is joint work with José Blanchet; this will be reorganized and submitted for publication soon. An early report on these results can be found in Blanchet et al. (2008). Chapter 5 is work in progress together with Søren Asmussen and has not been submitted anywhere. The empirical study in the Chapter 6 represents a joint effort with Henning Bunzel and Bent Jesper Christensen and is also work in progress.

Chapter 1

Introduction

The aim of this chapter is to serve as an introduction to the main tools and concepts used in the analysis of tail probabilities

$$\mathbb{P}(S_n > u), \qquad u \to \infty,$$

involving heavy-tailed random variables.

Basic concepts, general assumptions and common notation adopted all over the dissertation will be established in the Section 1.1. Basics on asymptotic analysis and probability theory are the main topics here.

The vague concept of heavy-tailed random variable used so far will be made precise in the Section 1.2. A motivation for the study of stochastic models with heavy-tails will be followed by a short discussion of the alternative definitions of *heavy-tailed random variables* available in the literature. It will be seen that these apparently different definitions lead to almost identical classifications of heavy-tailed distributions. Particular attention will be paid to the rich class of subexponential distributions and it will be discussed how its defining property provides a useful insight into the occurrence of large values of a sum of heavy-tailed random variables.

A general overview of Monte Carlo methods is provided in Section 1.3 with a particular emphasis in the area known as *Rare Event Simulation*. In those cases where no closed analytic expression for the distributions function is readily available, Monte Carlo methods have become one of the most reliable possibilities so far. The notions of *rare event* and *efficient estimator* will be formalized here to provide the proper framework for the analysis of Monte Carlo estimators of tail probabilities.

Later on, the most established tool in a light tailed setting (Siegmund, 1976; Bucklew et al., 1990; Sadowsky and Bucklew, 1990; Sadowsky, 1996) and its limitations in a heavy-tailed setting (Asmussen et al., 2000) will be discussed. The last part is devoted exclusively to the heavy-tailed setting; a recount of methods developed in the literature is given there followed by a more detailed exposition on the conditional estimators proposed by Asmussen and Binswanger (1997); Asmussen and Kroese (2006) in the i.i.d. case. Proofs of efficiency in an extended number of case are the first pieces of original research presented in this dissertation.

1.1 Basic Concepts and Notation

The use of capital letters will be (mostly) reserved for random variables and/or stochastic processes. The *cumulative distribution function* of any given random variable will be

usually represented by F or G and in some parts abbreviated c.d.f. The *tail probability* of a random variable with distribution F is defined as $1 - F(\cdot)$. The standard practice is to use the notation $\overline{F}$. The *tail probability* function is also known in other contexts as *survival function*.

The use of boldface has been kept to denote matrices and vectors which by convention will be in column form. Besides the standard matrix and vector operations we will consider functions of a matrix to be entry-wise. For example, $\log(\mathbf{A})$ should be understood as the matrix obtained by taking the logarithm of the individual entries of the matrix $\mathbf{A}$.

For any given sequence of random variables we will use S_d to represent the d-th partial sum, M_d for the maximum among the first d random variables and $X_{(k)}$ for the k-th order statistic among the first d (implicit) random variables.

We will adopt the practice of inserting the symbol $\widehat{\ }$ for denoting an unbiased estimator of a given quantity or function, say z. Then, we will understand that $\widehat{z}$ is an unbiased estimator of z. When we consider more than one estimator we will distinguish them by adding subindexes.

1.1.1 Laplace-Stieltjes Transform

The Laplace-Stieltjes transform of a function F is defined as the following Lebesgue-Stieltjes integral

$$\mathcal{L}^*F(\theta) := \int_{-\infty}^{\infty} \mathrm{e}^{-\theta t} dF(t), \qquad \theta \in \mathbb{C}.$$

If F is a distribution function and $X \sim F$, then $\mathbb{E}\left[\mathrm{e}^{-\theta X}\right] = \mathcal{L}^*F(\theta)$. The Laplace-Stieltjes transform is closely related to the Laplace transform in the following way, if the function F has a derivative f, then $\mathcal{L}F^*(\theta) = \mathcal{L}f(\theta)$ where

$$\mathcal{L}f(\theta) := \int_{-\infty}^{\infty} \mathrm{e}^{-\theta t} f(t) dt, \qquad t \in \mathbb{C},$$

is the Laplace transform of the function $f(\cdot)$. Note that by dropping the negative sign of the argument θ in the definition of the Laplace transform we would obtain the moment generating function of the random variable.

1.1.2 Asymptotic Notation

We will use the Bachmann-Landou asymptotic notation all over the dissertation (cf. DeBruijin, 1958) which is as follows: Let ϕ and g real functions such that $\phi > 0$ and $x_0 \in \mathbb{R} \cup \{\pm\infty\}$.

- We say that f is *big Oh* of ϕ as $x \to x_0$ and denote it $f = O(\phi)$ if there exists $k > 0$ such that
$$\limsup_{x \to x_0} f(x)/k\phi(x) < 1.$$
- We say that F is *small oh* of ϕ as $x \to x_0$ and denote it $f = o(\phi)$ if
$$\lim_{x \to x_0} f(x)/\phi(x) = 0.$$
- We say that f and ϕ are asymptotically equivalent as $x \to x_0$, denoted $f \sim \phi$ if
$$\lim_{x \to x_0} f(x)/\phi(x) = 1.$$

- We say that f is asymptotically proportional to ϕ as $x \to x_0$ and denote it $f = \Theta(\phi)$ if for some fixed positive value L we have

$$\lim_{x \to x_0} f(x)/\phi(x) = L.$$

Note that $f \sim \phi$ is equivalent to write $f(\cdot) = \phi(\cdot)(1 + o(1))$. Since the symbol $\sim$ is also used to express that a random variable follows a particular distribution the last notation will be prefer to avoid confusions. In a similar fashion, $f = \Theta(\phi)$ is equivalent to write $f(\cdot) = \phi(\cdot)(L + o(1))$ for some fixed positive value L. This notation is introduced in this dissertation to simplify some calculations where the value of L is not relevant for our purposes.

1.1.3 Failure Rate Function

We close this section with another concept widely used in applied probability:

Definition 1.1 (Failure Rate Function)**.** The *failure rate function* or *hazard rate function* of a random variable with distribution function F and density function f is defined

$$\lambda(t) := f(t)/\overline{F}(t).$$

1.2 Heavy-Tails and Subexponentiality

The notion of heavy-tails arises from those phenomena in nature where considerable consequences are result of events with small but significative frequencies. The classical example is that of a natural catastrophe such as an earthquake or a hurricane. The consequences can be dramatically as shown in the example given in Table 1.1 where the ten worst natural disasters in terms of fatalities for the past ten years are listed.

Event	Place	Date	Fatalities
Tsunami	Indian Ocean	Dec 2004	229,866
Cyclone Nargis	Burma	May 2008	140,000
Kashmir earthquake	Pakistan	Oct 2005	86,000
Sichuan earthquake	China	May 2008	69,197
Heat wave	Europe	Sum 2003	37,451
Bam earthquake	Iran	Dec 2003	30,000
Vargas mud slide	Venezuela	Dec 1999	20,006
Gujarat earthquake	India	Jan 2001	20,000
Hurricane Mitch	Central America	Oct 1998	18,277
Izmit earthquake	Turkey	Aug 1999	17,118

Table 1.1: The ten worst natural disasters in the past 10 years in terms of fatalities.

Examples of this type of phenomena abound in areas such as insurance (claim sizes and amount of claims), finance (portfolio returns) and data networks (file sizes, transmission rates and transmission durations). Hence, it is crucial to count with the appropriate framework for the analysis of this kind of phenomena where large amounts

of theoretical results and specialized techniques from applied probability and statistics are required.

Although this example might provide general guidelines on what a heavy-tailed distribution should be, there is no general consensus on it. Throughout this section we will discuss different rules to discriminate *heavy-tails* which are widely accepted in the literature as definitions for heavy-tails. Most of them are related to some extent to the rate of decay of its tail probability. From a practical point of view, these rules are almost equivalent since they provide similar classifications for the most common distributions.

Before we move on, we will try to illustrate with an example what *should not* be regarded as a heavy-tail. Consider the number of gold medals won by a single athlete in the same Olympic Games. We may agree than the possibilities of winning more than one gold medal for any given athlete are very low, so it might reasonable to consider an extreme event the fact that Michael Phelps won 8 gold medals in Beijing 2008. However, if we look back in the history of the modern Olympic Games we discover that several athletes had been close to this mark.

Athlete	Sport	Games	Gold Medals
Michael Phelps	Swimming	Beijing 2008	**8**
Mark Spitz	Swimming	Munich 1972	**7**
Michael Phelps	Swimming	Athens 2004	**6**
Vitaly Scherbo	Gymnastics	Barcelona 1992	**6**
Kristin Otto	Swimming	Seoul 1988	**6**

Table 1.2: Athletes with most gold medals won in the same Olympic Games

In this example, large outcomes form a cluster around the value 8, a feature which marks a notable difference with the example of the number of fatalities where large values are disperse. In fact, the concept of heavy-tail is not simply related with distributions where large values occur with relatively high frequency. More surprisingly, we will discover that not even large dispersions might be enough to consider a distribution to be heavy-tailed.

1.2.1 Infinite Moments and Infinite Moment Generating Function

An implicit idea in the concept of heavy-tails is that of infinite support. Here, we will concentrate exclusively in those distributions with nonnegative values. From the example of the gold medals we saw that simply large values or high dispersion are not the only characteristics we are looking in a heavy-tailed phenomenon. However, we will consider an infinite mean or infinite variance as sufficient conditions for a distribution to be heavy-tailed. For instance, we will accept that a random variable X is heavy-tailed if $\mathbb{E}[X^k] = \infty$ for some $k \in \mathbb{N}$. Moreover, the class defined by this characteristic is properly contained in the next definition.

Definition 1.2 (Heavy-tails with infinite m.g.f.)**.** We say that a random variable X is heavy-tailed if its moment generating function is infinite for any positive value of the argument. That is

$$\mathbb{E}[e^{\theta X}] = \infty, \qquad \theta > 0.$$

Note that a random variable with an infinite moment of order, say k, is contained in this class. However, it is not always true that a random variable with infinite moment generating function possesses an infinite moment. The popular example is that of a lognormal random variable which has finite moments of every order but its moment generating function is infinite for any positive value of the argument.

1.2.2 Mean Excess Function

A more intuitive classification is given by means of the excess function for those distribution with finite first moment. More precisely, the *mean excess function* of a random variable X is given as

$$e(u) := \mathbb{E}\left[X - u | X > u\right], \qquad u \in \text{supp}(X).$$

The *mean excess function* can be interpreted as the expected value of the excess of a random variable over a given threshold u. For a heavy-tail we would expect that this value increases as $u \to \infty$. In the example of the gold medals, a heavy-tailed distribution would imply that the next athlete that exceeds the number of 8 gold medals, most likely will do it for a fairly large amount. More precisely:

Definition 1.3 (Heavy-tails with infinite mean excess function)**.** We say that a random variable X is heavy-tailed if its mean excess function tends to infinity as $u \to \infty$.

An equivalent expression for the *mean excess function* which is useful for testing the *heaviness* of a given distribution is given by

$$e(u) = \frac{1}{\overline{F}(u)} \int_u^\infty \overline{F}(t) dt.$$

1.2.3 Exponential Decay Rate

Another rule for classification which is easy to verify is given in terms of the rate of decay of the tail probability. The common practice is to use the exponential function as a benchmark. That is,

Definition 1.4 (Heavy-tails with low rate of decay)**.** We say that a probability distribution F is *heavy-tailed* if

$$\liminf_{u\to\infty} \frac{\overline{F}(u)}{\mathrm{e}^{-\lambda u}} > 0, \qquad \forall \lambda > 0,$$

or *light tailed* if for some $\lambda > 0$

$$\limsup_{u\to\infty} \frac{\overline{F}(u)}{\mathrm{e}^{-\lambda u}} < \infty.$$

This is one of the easiest criteria for verifying the *heaviness* of a distribution. In fact, it also provides a definition for a *light tail*. It almost immediately follows that the exponential, gamma and normal distributions are examples of light tails. Note that an exponential transformation of *light tailed* random variables might yield to distributions with heavy-tails. In particular, the Pareto, loggamma and lognormal are the respective heavy-tailed distributions obtained from an exponential transformation of the light tailed distributions above.

1.2.4 Subexponentiality

Subexponential theory provides an alternative definition for heavy-tails which is almost equivalent to the previous ones. This class of distributions was introduced in Chistyakov (1964). In particular, it provides a deep insight into the occurrence of events involving large values of the sum. Several useful properties can be derived from the definition making the class of subexponential distribution attractive for applications.

We say that a probability distribution F belongs to the class $\mathcal{S}$ of subexponential distributions $\mathcal{S}$ if the tail probability of the convolution of two i.i.d. random variables with common distribution F is asymptotically equivalent to two times the tail probability of one of them. More precisely:

Definition 1.5. A probability distribution F is subexponential ($F \in \mathcal{S}$) if its support is contained in $(0, \infty)$ and

$$\lim_{u\to\infty} \frac{\mathbb{P}(X_1 + X_2 > u)}{\mathbb{P}(X_1 > u)} = 2, \tag{1.1}$$

where X_1 and X_2 are independent with common distribution F.

In particular, if a probability distribution F belongs to the class $\mathcal{S}$ an equivalent relation holds for the convolution of an arbitrary number of i.i.d. random variables with common distribution F. That is,

$$\lim_{u\to\infty} \frac{\mathbb{P}(X_1 + \cdots + X_n > u)}{\mathbb{P}(X_1 > u)} \to n.$$

In the literature, this equivalent property is often preferred as definition for the class $\mathcal{S}$. In fact, it provides a useful intuition into how large values of the sum are likely to occur. We illustrate this idea below.

Note that the distribution of the maximum M_n of arbitrary i.i.d. random variables (not necessarily subexponential) is given by $F^n(u)$. Then, the following asymptotic equivalent expression is obtained

$$\mathbb{P}(M_n > u) = 1 - F^n(u) = (1 - F(u)) \sum_{k=0}^{n-1} F^k(u) = \overline{F}(u)\big(n + o(u)\big).$$

The last asymptotic expression follows by noting that $F^k(u) \to 1$ for all $k \in \mathbb{N}$ and the well known relation

$$\sum_{k=0}^{n-1} F^k(u) = \frac{1 - F^n(u)}{1 - F(u)}.$$

The conclusion here is that the tail distribution of the maximum of i.i.d. random variables is determined by the tail of a single random variable. Moreover, if F belongs to the class $\mathcal{S}$ then

$$\lim_{u\to\infty} \frac{\mathbb{P}(X_1 + \cdots + X_n > u)}{\mathbb{P}(M_n > u)} = 1.$$

This expression can be rewritten so it reads instead $\mathbb{P}(M_n > u | X_1 + \cdots + X_n > u) \to 1$. Intuitively it says that the sum becomes large due to the contribution of a single random variable. This behavior is completely opposite to that of lighted tails where the sum of i.i.d. random variables becomes large as a consequence of several moderately large

contributions. Hence, it turns out that distributions within the class $\mathcal{S}$ should be appropriate for modeling those phenomena which show some stability through time but eventually are shocked by an extreme event associated with large values.

The property in Definition 1.5 on the preceding page looks rather simple. However many properties can be derived from it, making $\mathcal{S}$ a very flexible class of distributions. Below we list some of the main properties of the class $\mathcal{S}$ which will be useful for our purposes. For doing so, it will be convenient to establish an order relation among distributions with infinite support according to the *heaviness* of its tail.

Definition 1.6. Let F_k and F_ℓ be nonnegative distribution functions with support in $(0, \infty)$. Then we say,

a) F_k has a lighter tail than F_ℓ (or F_ℓ has heavier tail than F_k) if $\overline{F}_k = o\big(\overline{F}_\ell\big)$.

b) F_k and F_ℓ have proportional tails if $F_k = F_\ell(c + o(1))$ for some fixed $c \in (0, \infty)$.

c) F_k and F_ℓ have equivalent tails if $F_k = F_\ell(1 + o(1))$.

Properties of Subexponential Distributions Let $F \in \mathcal{S}$ and F_1, F_2 two nonnegative distribution functions with infinite support. Then

- The following limit holds uniformly on compact sets of $(0, \infty)$.

$$\lim_{u\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1.$$

  This property defines a larger class of distributions commonly known as *long tailed* distributions which is denoted as $\mathcal{L}$.

- The following property justifies the name of the class $\mathcal{S}$.

$$\lim_{u\to\infty} \frac{\overline{F}(x)}{\mathrm{e}^{-\lambda u}} \to \infty.$$

  This property indicates that the class of heavy-tailed distributions we define accordingly to its *decay rate* is contained in $\mathcal{S}$.

- Let a_1, a_2 be positive constants. If $F_1(u) = F(u)(a_1 + o(1))$ and $F_2(u) = F(u)(a_2 + o(1))$ as $u \to \infty$, then

$$F_1 * F_2(u) = F(u)(a_1 + a_2 + o(1)).$$

- Suppose that $F_1, F_2 \in \mathcal{S}$ are such that $F_2(u)$ has lighter tail than $F_1(u)$. Then, $F_1 * F_2 \in \mathcal{S}$ and

$$F_1 * F_2(u) = F_1(u)(1 + o(1)).$$

  Intuitively, this property says that the tail probability of the sum is dominated by the one with the heaviest tail.

More details and examples of distributions with heavy-tails can be found in Embrechts et al. (1997). In this dissertation we will concentrate in the *lognormal* distribution and the class of *regularly varying* distributions with index α. In particular, these

two distributions fulfill the definitions of each of the classes of heavy-tailed distributions listed above. A short exposition on the lognormal distribution is given in A.1.1 on page 87. Since the definition of regularly varying is often regarded to be a definition of heavy-tails we discuss some relevant aspects next. Further details can be found in A.1.2 on page 88.

1.2.5 Regularly Varying Distributions

Definition 1.7. The class $\mathcal{R}$ of random variables with regularly varying tails is defined as the family of nonnegative random variables whose tail probability can be written as

$$\overline{F}(x) = \frac{L(x)}{x^{\alpha}}, \qquad x, \alpha > 0,$$

where $L(x)$ is a *slowly varying function*. That is, $L(x)$ is a measurable function satisfying

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1, \qquad \forall t \in (0, \infty).$$

Equivalently, a distribution is regularly varying if and only if

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = t^{-\alpha}, \qquad \forall t \in (0, \infty).$$

It is also common to partition the class $\mathcal{R}$ accordingly to its index α. Therefore, $\mathcal{R}_\alpha$ represents the class of regularly varying distributions with index α. The notation $\mathcal{R}_0$ is reserved for the class of slowly varying functions. The Pareto, Burr, α-stable and loggamma are typical examples of regularly varying distributions.

The class $\mathcal{R}_\alpha$ is often understood as those distributions with a tail behavior similar to a power function with exponent α where the slowly varying function $L(x)$ acts as a perturbation factor. This class has been largely studied under the more general theory of *regularly varying functions* (cf. Bingham et al., 1987). For now we will just enunciate a well known result which will be useful in this dissertation.

Proposition 1.8 (Karamata's Theorem)**.** *Let $L \in \mathcal{R}_0$ be bounded in $[x_0, \infty)$ and $\alpha > 1$. Then*

$$\int_x^{\infty} \frac{L(t)}{t^{\alpha}} dt = \frac{L(x)}{(\alpha - 1)x^{\alpha-1}}(1 + o(1)) \qquad x \to \infty.$$

This result says that the integrated tail of a regularly varying function with index $\alpha > 1$ will be regularly varying with index $\alpha - 1$. Even more, it says that the slowly varying function is preserved after the integration.

Using *Karamata's Theorem* it is easy to verify that the *mean excess function* $e(u)$ of a regularly varying goes to infinity as $u \to \infty$, recall the alternative representation for $e(u)$. Similarly, all the moments of order large than the index α of a regularly varying distribution are infinite while those of smaller order than α are finite. It is trivially seen that the tail probability decays slower than the exponential and it is also provable that a regularly varying distribution satisfies the characteristic property of subexponential distributions. Hence, the class $\mathcal{R}$ inherits all the properties of the class $\mathcal{S}$.

1.3 Rare Event Simulation

A *rare event* is simply a set $A \in \mathcal{F}$ with such a small probability that severe computational difficulties arise while trying to calculate it. These are usually characterized by inaccurate approximations or excessive running times. (cf. Asmussen and Glynn, 2007; Bucklew, 2004; Juneja and Shahabuddin, 2006). To illustrate this, we consider a set $\{A_u\}$ of events indexed by u such that

$$z(u) := \mathbb{P}(A_u) \to 0, \qquad u \to u_0.$$

The basic algorithm, named *Crude Monte Carlo*, simulates an arbitrary number R of outcomes $\omega_1, \ldots, \omega_R$ according to the probability measure $\mathbb{P}$, and approximates $\mathbb{P}(A)$ with

$$\widehat{z}(u) = \frac{1}{R}\sum_{i=1}^{R} \mathbb{I}(\omega_i \in A_u), \qquad \omega_1 \in \Omega.$$

Note that, since $\mathbb{I}(\omega_i \in A)$ is a Bernoulli random variable with parameter $p = \mathbb{P}(A_u)$ its standard deviation is given by

$$\mathrm{SD}\left[\widehat{z}(u)\right] = \sqrt{\frac{\mathbb{P}(A_u)\big(1-\mathbb{P}(A_u)\big)}{R}}.$$

and it converges to $z(u)$ by the law of large numbers Commonly, we would like to generate an estimation with some degree of precision and for doing so we can use the language of confidence intervals. More precisely, by the central limit theorem we know that an approximate confidence interval at $(1-\alpha)\%$ is given by

$$\widehat{z}(u) \pm \Phi(1-\alpha/2)\sqrt{\frac{\mathbb{P}(A_u)\big(1-\mathbb{P}(A_u)\big)}{R}},$$

where Φ is the cumulative distribution function of a normal standard r.v. This means that the error of the estimate will be smaller than

$$\Phi(1-\alpha/2)\sqrt{\frac{\mathbb{P}(A_u)\big(1-\mathbb{P}(A_u)\big)}{R}},$$

with a confidence of $(1-\alpha)\%$. This can be conveniently translated into the number of replications necessary to achieve a particular precision: How many replications would be necessary to have an error smaller than ϵ with an α confidence level? The answer is

$$R \approx \frac{\Phi^2(1-\alpha/2)\,\mathbb{P}(A_u)\,(1-\mathbb{P}(A_u))}{\epsilon^2}.$$

At first sight, it might appear that the Crude Monte Carlo estimators would deliver accurate results since the number of simulations required for getting a precision ϵ would decrease as $\mathbb{P}(A_u) \to 0$. However, taking a fixed precision ϵ is not the right approach to rare events. The problem here is that the size of the error might be huge compared to the size of the estimation, so we need to set ϵ accordingly to the size of the probability $\mathbb{P}(A_u)$.

For instance, if we set an error size of the order $O(\mathbb{P}(A_u))$ as $u \to u_0$ (this simply means that the size of the error will remain proportional to the size of the estimation)

a few simple operations would indicate that the required number of replications would be $R = \Theta\big(1/\mathbb{P}(A_u)\big)$. Clearly, this quantity will go to infinity as $\mathbb{P}(A_u) \to 0$. The conclusion is that Crude Monte Carlo easily becomes infeasible for estimating small probabilities and the necessity of improved estimators becomes evident. In fact, this is one the main objectives of the area in simulation known as *Variance-Reduction Methods* (cf. Asmussen and Glynn, 2007; Rubinstein and Kroese, 2008). For more details see Appendix B.2 on page 97.

For now, we will just say that a *Variance Reduction method* is simply an algorithm which produces an estimator with smaller variance than Crude Monte Carlo. The quality of a particular method is assessed not only based on the variance reduction itself but also in the amount of computational resources consumed, the theoretical work and the implementation effort invested. Hence, for a particular method to be worthwhile, the variance reduction should be substantial enough to compensate these aspects. It is obvious from the discussion above that the demand of reduction of variance in the presence of rare events is huge.

1.3.1 Efficiency Concepts in Rare Event Simulation

It would be desirable to have an estimator which has a variance reduction factor of the order $O(\mathbb{P}(A_u))$ as $u \to u_0$. In fact, this is a realistic performance which is known in the literature as *bounded relative error* or *strong efficiency* and is properly defined in terms of its variance as follows

$$\limsup_{u\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{z}(u)}{z^2(u)} < \infty.$$

Under some settings, *bounded relative error* is a criterion which turns out to be too demanding. The reason lies in the difficulty of finding an implementable estimator and proving that it has bounded relative error. For this reason, it is usually common to work with a slightly weaker concept known as *logarithmic efficiency* or simply *asymptotic efficiency* which is defined as

$$\limsup_{u\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{z}(u)}{z^{2-\epsilon}(u)} = 0, \qquad \forall \epsilon > 0.$$

From this definition is clear that *bounded relative error* implies *logarithmic efficiency*. From a practical point of view, there is no substantial difference between these two criteria. By contrast, in most applications it is easier to prove *logarithmic efficiency* in part due to the following Proposition which provides a more flexible equivalent criterion.

Proposition 1.9. *A simulation estimator* $\widehat{z}(u)$ *is* logarithmic efficient *if and only if*

$$\liminf_{u\to\infty} \frac{\big|\log \mathbb{V}\mathrm{ar}\,\widehat{z}(u)\big|}{\big|\log z^2(u)\big|} \geq 1.$$

Recently, a stronger efficiency concept has been targeted by several authors. This criterion is known under several names such as *asymptotically zero relative error*, *asymptotically optimal relative error* or *vanishing relative error* and is defined as

$$\limsup_{u\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{z}(u)}{z^2(u)} = 0.$$

This is clearly much stronger than bounded relative error. In fact, when an estimator has *asymptotically zero relative error* it has produced a variance reduction factor of the order $o(\mathbb{P}(A_u))$. It is so strong, that the number of replications necessary to attain an error of order $O(\mathbb{P}(A_u))$ as $u \to u_0$ will decrease as the event becomes rarer until the level where it is only necessary to generate a single replication.

However, it is a difficult task to find an efficient simulation estimator and prove that it verifies any of the conditions above. Therefore, we have defined an efficiency criterion which is weaker than *logarithmic efficiency*. More precisely, for a fixed value $\delta > 0$ we say that an estimator $\widehat{z}(x)$ is δ-*efficient* if

$$\limsup_{u\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{z}(u)}{z^{2-\delta}(u)} < \infty.$$

This means that if we keep the number of replication bounded we will get a variance reduction factor of the order $O\big(\mathbb{P}^{1-\delta/2}(A_u)\big)$. The motivation to define this concept is not merely to fill the gap between *bounded relative error* and the Crude Monte Carlo efficiency but also to provide a tool to detect weakness of an estimator which might be corrected for attaining a stronger efficiency.

L'Ecuyer et al. (2008) introduce new stronger efficiency concepts which take care of the moments of higher order. Hence, these concepts are named *bounded relative error of order k* and *logarithmic efficiency of order k*.

1.3.2 Rare Event Simulation for Sums of Random Variables.

Variance reduction methods are useful tools for constructing efficient estimators. However, it often requires a considerable amount of theoretical work and most often they should be designed for the problem at hand. Therefore, it is relevant to define more precisely the kind of rare events we are interested in.

Generally speaking, we will be interested in those rare events related to a sum of random variables with heavy-tails. More precisely,

$$\mathbb{P}(X_1 + \cdots + X_N > u),$$

where $X_1, X_2, \ldots$ a sequence of random variables and N possibly random. For the rest of this chapter we will discuss some of methods which produce efficient algorithm for the sums of a fixed number of non-negative i.i.d. random variables. In future chapters we will relax these assumptions and specialize in more complex cases which will require further refinements and improvements of the methods presented in this section.

The most established tool for rare event simulation is *Importance Sampling* (see the appendix B.2.2 on page 98). Roughly speaking, for the method to produce a variance reduction it should simulate from a distribution which produces samples in the rare event with higher frequency than the original one. Then, an unbiased estimator comes out as the result of an appropriate weighted average of the outcomes. In fact, theoretically there exists an importance sampling distribution which can produce a *zero variance* estimator. The implementation of such algorithm is typically infeasible since it often requires the knowledge of the probability of interest $\mathbb{P}(A)$. However, it suggests how to select the importance sampling distribution.

For the *light tailed case* the standard tool for estimating $\mathbb{P}(S_n > u)$ is Importance Sampling with an *exponential change of measure*. This simply means that the proposed distribution F_θ is obtained by normalizing the measure $\mathrm{e}^{\theta y}F(dy)$ for some value $\theta \in \mathbb{R}$.

The reason being that if we choose θ to be the solution of $\mathbb{E}_\theta[X_1] = u$, then the distribution associated to the *zero variance estimator* converges to F_θ as the number of summands d increases to infinity. Estimators with excellent efficiency properties can be constructed by choosing a feasible importance sampling distribution which emulates the *optimal change of measure*.

However, we have seen before that a heavy-tailed random variable can be characterized by the nonexistence of the moment generating function. Asmussen et al. (2000) present a number of counterexamples suggesting that different methods should be employed for this kind of distributions. In recent years, several algorithms have been proposed for the efficient estimation of a sum of heavy-tailed random variables. The first logarithmic efficient algorithm was proposed in Asmussen and Binswanger (1997) for the regularly varying case. This algorithm is based on Conditional Monte Carlo and uses order statistics. Asmussen et al. (2000) and Juneja and Shahabuddin (2002) proposed importance sampling algorithms which are logarithmically efficient. A conditional algorithm, similar to that of Asmussen and Binswanger (1997) was proposed in Asmussen and Kroese (2006). That algorithm exploits a symmetry relation of i.i.d. random variables and the conditioning involves the highest order statistic and has bounded relative error in the regularly case. Dupuis et al. (2006) proposed a state-dependent algorithm for the regularly varying case which has bounded relative error. Blanchet and Glynn (2008) proposed the first logarithmically efficient rare-event simulation algorithm for a *GI/G/1* queue for a large class of heavy-tailed distributions.

Here, we will concentrate in the conditional algorithms proposed in Asmussen and Binswanger (1997) and Asmussen and Kroese (2006). At the end of this section we include some extended results. Moreover, several algorithms studied in this dissertation build on these early ideas.

1.3.3 Conditional Monte Carlo Methods Based on Order Statistics

Roughly speaking, a *Conditional Monte Carlo estimator* is a tool which improves the efficiency of an estimator by *extracting* the variability coming from known information. More precisely,

$$z(x) = \mathbb{P}(S_n > x|\mathcal{F}),$$

where $\mathcal{F}$ is a σ-algebra. This estimator is clearly unbiased and by the Rao-Blackwell Theorem it always provides a variance reduction. The problem is to choose a σ-algebra $\mathcal{F}$ such that it is possible to simulate the events in $\mathcal{F}$ and the conditional probability $\mathbb{P}(S_n|\mathcal{F})$ is known in closed form. Obviously it is also desired that the resulting estimator provides a substantial variance reduction.

Recall that a distribution F belongs to the class $\mathcal{S}$ of subexponential distribution if it has the following property

$$\lim_{u\to\infty} \frac{\overline{F}^{d*}(u)}{\overline{F}(u)} = d,$$

where F^{d*} is defined as the d-fold convolution of F. We also discussed that this expression has an useful intuitive interpretation of how large values of the sum occur and it is as follows: the sum of independent subexponential random variables becomes large as the consequence of a single large contribution, namely the maximum.

The main idea in the algorithms designed by Asmussen and Binswanger (1997) and Asmussen and Kroese (2006) is to exploit this intuition by using order statistics. The idea is to calculate the probability that the maximum alone is responsible for the large

value of the sum while the rest of the order statistics behave in a normal way. It was noted in Asmussen and Binswanger (1997) that by using order statistics instead of the *crude* random variables the efficiency was largely improved.

Although the algorithms in Asmussen and Binswanger (1997) and Asmussen and Kroese (2006) are both based on this idea, they differ significatively in the way of conditioning – these differences will be explained later. The modified version in Asmussen and Kroese (2006) provided a significatively more efficient and easier to implement algorithm.

1.3.3.1 Asmussen-Binswanger estimator

The algorithm proposed by Asmussen and Binswanger (1997) suggests to condition with respect to the order statistics. So, we need to simulate the first $d-1$ order statistics out of d. The procedure requires to simulate $X_1, \ldots, X_d$ and simply discard the largest one. Then we use the conditional probability of the maximum given the rest of the order statistics to calculate the probability that the sum is larger than u. This comes out as

$$\mathbb{P}\big(S_d > u \,\big|\, X_{(1)}, \ldots, X_{(d-1)}\big) \;=\; \frac{\overline{F}((u - S_{(d-1)}) \vee X_{(d-1)})}{\overline{F}(X_{(d-1)})},$$

where $S_{(d-1)} = X_{(1)} + \cdots + X_{(d-1)}$. Asmussen and Binswanger (1997) proved that their algorithm is logarithmically efficient in the i.i.d. regularly varying case. Binswanger (1997) proved logarithmic efficiency when the random variables are i.i.d. lognormal distributed.

The algorithm in Asmussen and Binswanger (1997) was proposed for i.i.d. random variables. However, we can easily drop the *identically distributed* from the assumptions. When simulating the order statistics, we just need to keep track of the (random) index of the largest random variable, say K, then the conditioning will deliver instead

$$\mathbb{P}\big(S_d > u \,\big|\, X_{(1)}, \ldots, X_{(d-1)}\big) \;=\; \frac{\mathbb{P}(X_K > (u - S_{(d-1)}) \vee X_{(d-1)})}{\mathbb{P}(X_K > X_{(d-1)})}. \tag{1.2}$$

Algorithmically:

Asmussen-Binswanger Algorithm: Independent Nonidentical r.v.'s.

1. Simulate $X_1, \ldots, X_d$. Register the index of K of the largest random variable and form the order statistics $X_{(1)}, \ldots, X_{(d)}$.

2. Return
$$\widehat{z}_{AB}(u) = \frac{\overline{F}_K\big(X_{(d-1)} \vee (u - S_{d-1})\big)}{\overline{F}_K\big(X_{(d-1)}\big)},$$
where $F_k(\cdot)$ is the distribution function of the k-th random variable.

The claimed efficiency for independent and nonidentical distributed lognormal random variables is given in the following Theorem and its proof can be found on page 16.

Theorem 1.10. *Let $X_1, .., X_n$ be independent lognormal random variables. Then the estimator $\widehat{z}_{AB}(x)$ is logarithmic efficient.*

1.3.3.2 Asmussen-Kroese estimator

A somewhat similar idea was suggested by Asmussen and Kroese (2006). Instead of using the order statistics they conditioned with respect to the event where a particular random variable takes the paper of the maximum. Then, they use the fact that the random variables are i.i.d. to unbias their estimator using the following symmetry argument

$$\mathbb{P}(S_d > x) = d\,\mathbb{P}(S_d > x, X_d = M_d). \tag{1.3}$$

So, this estimator is built by conditioning on $\mathcal{F} = \sigma(X_1, \ldots, X_{d-1})$ and noting that

$$d\,\mathbb{P}(S_d > x, X_d = M_d | X_1, \ldots, X_{d-1}) = d\overline{F}(M_{d-1} \vee (x - S_{d-1})) \tag{1.4}$$

(recall $M_k := \max\{X_1, \ldots, X_k\}$). Asmussen and Kroese (2006) proved that their algorithm has bounded relative error in the regularly varying case. The same conclusion with lognormal marginals was proved independently by Asmussen and Rojas-Nandayapa (2006) and Hartinger and Kortschak (2006). In particular, the last paper went further by noting that it has *asymptotically vanishing relative error*.

As in the case of the algorithm by Asmussen and Binswanger (1997) we can easily drop the *identically distributed* assumption. For doing so we substitute the symmetric argument with

$$\mathbb{P}(S_d > x) = \sum_{k=1}^{d} \mathbb{P}(S_d > x, X_k = M_d).$$

This idea was empirically explored in Asmussen and Rojas-Nandayapa (2006). As a first approach one could try to estimate individually each of this terms conditioning as above. Although the resulting estimator usually has good efficiency properties it requires d times more replications. An alternative approach with slightly better results is described next. Define a random variable K such that under the probability measure $\mathbb{P}$ it is a discrete random variable supported on $\{1, \ldots, d\}$ with $\mathbb{P}(K = i) = p_i$, then

$$\mathbb{P}(S_d > x) = \sum_{k=1}^{d} \mathbb{P}(S_d > x, X_k = M_d) = \mathbb{E}\left[\frac{\mathbb{I}(S_d > x, X_K = M_d)}{p_K}\right].$$

so, if we condition with respect to $\mathcal{F} = \sigma(K, X_1, \ldots, X_{K-1}, X_{K+1}, \ldots, X_d)$ we obtain

$$\mathbb{P}(S_d > x) = \mathbb{E}\left[\frac{\overline{F}_K(M_{-K} \vee (x - S_{-K}))}{p_K} \,\middle|\, \mathcal{F}\right], \tag{1.5}$$

where F_k is the distribution of the k-th random variable and M_{-k} and S_{-k} are defined as the maximum and sum of the X_i's without considering the k-th random variable. The convenient election of the p_k's should deliver a significative variance reduction. Intuitively, this should be minimized if we choose $p_k := P(M_d = X_k | S_d > u)$. That is, the probability that the k-th random variable is larger than the rest conditioned to the rare event. However, this probability is not available beyond the independent case. Our suggestion is to use

$$p_k(u) = \frac{\mathbb{P}(X_k > u)}{\sum_{i=1}^{d} \mathbb{P}(X_i > u)}.$$

In empirical studies we have verified that this proposal is *close* to the optimal selection of the p_k's. In the i.i.d. case this yields to the original proposal by Asmussen and Kroese (2006). In the chapter 3 we will discuss how this estimator looks when we drop the *independence* assumption. The algorithm for estimating $\mathbb{P}(S_n > u)$ is given next.

Asmussen-Kroese Algorithm: Independent nonidentical r.v.'s

1. Simulate K.
2. Simulate $X_1, \ldots, X_{K-1}, X_{K+1}, \ldots, X_d$.
3. Return
$$\widehat{z}_{AK}(u) = \frac{\overline{F}\big(M_{(-K)} \vee (u - S_{-K})\big)}{p_K}.$$

The efficiency for nonidentical and independent random variables in the lognormal and regularly varying is given in the next two Theorems and their proofs can be found on pages 19 and 21 respectively.

Theorem 1.11. *Let $X_1, X_2, \ldots, X_d$ be independent lognormal random variables, K a discrete random variable supported over $\{1, \ldots, d\}$. Then, $\widehat{z}_{AK}(u)$ is an unbiased estimator of $\mathbb{P}(S_d > u)$ with bounded relative error.*

Theorem 1.12. *Let $X_1, X_2, \ldots, X_d$ be independent regularly varying random variables with indexes α_i respectively, K a discrete random variable supported over $\{1, \ldots, d\}$. Then, $\widehat{z}_{AK}(u)$ is an unbiased estimator of $\mathbb{P}(S_d > u)$ with bounded relative error.*

1.4 Concluding Remarks

1.4.1 Numerical Examples

In our numerical examples we have implemented the Asmussen-Kroese algorithm for estimating the tail probability of the sum of independent nonidentically distributed lognormal random variables. The results are shown in Example D.4 on page 107 where it has been compared against other algorithms designed for the correlated case which are discussed in the Chapters 3 and 4. A detailed description of the examples can be found on page 106. For now, we will only concentrate on the results label as **AK** corresponding to the Asmussen-Kroese estimator and its stratified version **S-AK** which makes use of the strategy proposed in Juneja (2007). This strategy consists in dividing the event of interest in two parts, namely,

$$\mathbb{P}(S_n > u) = \mathbb{P}(S_n > u, M_n < u) + \mathbb{P}(M_n > u),$$

estimate the first term with the method used in Asmussen and Kroese (2006) and evaluate the second tern in closed form.

The numerical results indicate that the modified version of the Asmussen-Kroese provides reliable results for estimating the tail probability of a sum of nonidentical random variables. However, it is noted that the stratification algorithm has variance equal to zero. The main problem here is that the algorithm is not efficient for estimating the probability of the event $\{S_n > u, M_n\}$ as documented in Juneja (2007) while the other algorithms shown in that example were designed to provide a better estimate of that region. A logical solution might be to use truncated random variables. However, this is a limited approach since it can not be generalized to the correlated case.

This example shows indirectly shows that $\mathbb{P}(M_n > u)$ provides an excellent approximation of this tail probability (note that the variance of the estimator is zero, so the estimator is only composed of the term $\mathbb{P}(M_n > u)$). In view of this and the discussion

on subexponential theory, I have decided to conduct an additional numerical experiment for approximating the tail probability of a sum of independent lognormal random variables with three different methods:

1. The tail probability of the maximum $\mathbb{P}(M_n > u)$.
2. The subexponential approximation.
3. The sum of the tail probabilities of the individual random variables

$$\sum_{i=1}^{n} \mathbb{P}(X_i > u). \tag{1.6}$$

In the future we will refer to this as the *aggregated tails*.

The results are summarized in the Example D.1 on page 103. The subexponential approximation tends to underestimate the real probability since it ignores all those random variables with tail behavior only slightly lighter than the heaviest one. This suggest considering the adjusted approximation in 1.6 which can be easily proved to be asymptotically equivalent. Moreover, this expression effectively bounds above the tail probability of the sum.

Notes and comments

Useful references for heavy are tails are Adler et al. (1996); Embrechts et al. (1997); Resnick (1987, 2006). A full treatment of regularly varying functions can be found in Bingham et al. (1987). For general topics in simulation we refer to Asmussen and Glynn (2007); Rubinstein and Kroese (2008); Henderson and Nelson (2006). In the case of rare event simulation Asmussen and Glynn (2007); Bucklew (2004); Juneja and Shahabuddin (2006) are useful references.

Theorem 1.8 was proved for the i.i.d. case in Binswanger (1997) and here we have generalized this result for the nonidentical and independent setting. As mention before, Theorem 1.11 was independently proved in Asmussen and Rojas-Nandayapa (2006) and Hartinger and Kortschak (2006). The version suggested here considers the nonidentical and independent setting and corresponds to a corrected version of the statement given in Asmussen and Rojas-Nandayapa (2006). Theorem 1.12 was proved in the original paper by Asmussen and Kroese (2006) and here we have provided an alternative proof which considers the more general setting of nonidentical and independent random variables. Juneja (2007) proposed a stratification strategy which improves significatively the performance of a given algorithm in a subexponential setting.

Recently, Blanchet and Li (2008) proposed an state-dependent importance sampling which is claimed to be strongly efficient under more general settings, assuming only subexponentiality of the increments. For the particular case of i.i.d. lognormal random variables Bee (2008) proposed an importance sampling algorithm based on a mixture of lognormals. The parameters were adjusted using Cross-Entropy and the result tested empirically.

1.5 Proofs

Proof of Theorem 1.10. In order to characterize the dominant tail behavior we define

$$\sigma^2 = \max_{1 \le k \le d} \sigma_k^2, \quad \mu = \max_{k:\, \sigma_k^2 = \sigma^2} \mu_k,$$

and let F be the distribution of a lognormal random variable with parameters μ and σ. Note that the index K is a discrete random variable supported over $\{1,\dots,d\}$, so we can simplify our proof using the following inequality

$$\mathbb{E}\left[\widehat{z}^2_{AB}\right] = \mathbb{E}\left[\frac{\overline{F}^2_K\big(X_{(d-1)} \vee (u - S_{d-1})\big)}{\overline{F}^2_K\big(X_{(d-1)}\big)}\right] \leq \sum_{k=1}^{d} \mathbb{E}\left[\frac{\overline{F}^2_k\big(X_{(d-1)} \vee (u - S_{d-1})\big)}{\overline{F}^2_k\big(X_{(d-1)}\big)}\right].$$

The idea is to obtain an asymptotic upper bound for the expectation for a fixed k. Then we break this expectation in two pieces as follows

$$\begin{aligned}\mathbb{E}\left[\frac{\overline{F}_k{}^2\big((u - S_{(d-1)}) \vee X_{(d-1)}\big)}{\overline{F}_k{}^2(X_{(d-1)})}\right] &= \mathbb{E}\left[\frac{\overline{F}^2_k\big((u - S_{(d-1)}) \vee X_{(d-1)}\big)}{\overline{F}^2_k\big(X_{(d-1)}\big)}; X_{(d-1)} < \frac{u}{d}\right] \\ &\quad + \mathbb{E}\left[\frac{\overline{F}_k{}^2((u - S_{(d-1)}) \vee X_{(d-1)})}{\overline{F}_k{}^2(X_{(d-1)})}; X_{(d-1)} > \frac{u}{d}\right].\end{aligned}$$

The quotient inside the second expectation is always smaller than 1, so we can bound the whole expectation with $\mathbb{P}\big(X_{(d-1)} > u/d\big)$. For the first expectation, it will be useful to note that if $X_{(d-1)} < u/d$ then the following inequalities hold

$$u - S_{(d-1)} \;\geq\; u - (d-1)X_{(d-1)} \;\geq\; u - \frac{d-1}{d}\,u = u/d \;\geq\; X_{(d-1)}.$$

This implies that in the event $\{X_{(d-1)} < u/d\}$, the following inequality holds true as well

$$\overline{F}_k\big((u - S_{(d-1)}) \vee X_{(d-1)}\big) \leq \overline{F}_k(u/d).$$

Inserting these bounds in the expectations we arrive at the following upper bound

$$\mathbb{E}\left[\frac{\overline{F}^2_k(u/d)}{\overline{F}^2_k(X_{(d-1)})}; X_{(d-1)} < \frac{u}{d}\right] + \mathbb{P}\big(X_{(d-1)} > u/d\big). \tag{1.7}$$

We concentrate on the expectation in the last term. Since $X_{(n-1)} < u/d$ we can apply Lemma 1.13 on the next page to get a bound for the quotient in the first expectation to obtain

$$c\,\mathbb{E}\left[\frac{\overline{F}^2(u/d)}{\overline{F}^2(X_{(d-1)})}; X_{(d-1)} < \frac{u}{d}\right] = c\,\overline{F}^2(u/d)\,\mathbb{E}\left[\frac{1}{\overline{F}^2(X_{(d-1)})}; X_{(d-1)} < \frac{u}{d}\right],$$

where c is a constant (recall that F was defined as the distribution with the dominant tail). Letting $F_{(d-1)}$ and $f_{(d-1)}$ be the distribution and density functions of $X_{(d-1)}$ respectively, we rewrite this expectation in integral form and use partial integration to obtain

$$\begin{aligned}\int_0^{u/d} \frac{f_{(d-1)}(y)}{\overline{F}^2(y)}dy &= -\frac{\overline{F}_{(d-1)}(y)}{\overline{F}^2(y)}\bigg|_0^{u/d} + 2\int_0^{u/d}\frac{\overline{F}_{(d-1)}(y)f(y)}{\overline{F}^3(y)}dy \\ &= 1 - \frac{\overline{F}_{(d-1)}(u/2)}{\overline{F}^2(u/2)} + 2\int_0^{u/d}\frac{\overline{F}_{(d-1)}(y)}{\overline{F}^2(y)}\frac{f(y)}{\overline{F}(y)}dy.\end{aligned}$$

We get a new upper bound by just ignoring the negative term. For dealing with integral it will be useful to note that

$$\frac{\overline{F}_{(d-1)}(t)}{\overline{F}^2(t)} \leq \frac{\sum_{i\neq j}\overline{F}_i(t)\overline{F}_j(t)}{\overline{F}^2(t)} = O(1), \qquad (0,\infty). \tag{1.8}$$

This is true since the $\overline{F}$ has the heaviest tail so it dominates all $\overline{F}_k$'s, and the quotient remains bounded as $y \to \infty$. Trivially, the same holds true as $y \to 0$. Then, by a continuity argument this quotient remains bounded all over $(0,\infty)$ by a constant, say $c_1 > 0$. We use this to obtain a new upper bound

$$1 + c_1 \int_0^{u/d} \frac{f(t)}{\overline{F}(t)} dy = 1 - c_1 \log \overline{F}(u/d).$$

Inserting this new bound in (1.5) we have obtained a new bound for $\mathbb{E}\,\widehat{z}^2_{AB}(u)$ which has the following shape

$$c\,\overline{F}^2(u/d)\big[1 - c_1 \log \overline{F}(u/d)\big] + \overline{F}_{(d-1)}(u/d) \leq c_2\,\overline{F}^2(u/d)\big[1 - c_1 \log \overline{F}(u/d)\big],$$

where the last inequality was obtained by using the argument (1.8). So, to prove logarithmic efficiency we need

$$\lim_{u\to\infty} \frac{\mathbb{E}\,\widehat{z}_{AB}(u)}{\mathbb{P}^{2-\epsilon}(S_d > u)} \leq \lim_{u\to\infty} \frac{c_2\,\overline{F}^2(u/d)\big[1 - c_1 \log \overline{F}(u/d)\big]}{\overline{F}^{2-\epsilon}(u)} = 0.$$

Using *Mill's ratio* and some basic calculus it is provable that the last limit is zero for all $\epsilon > 0$. By doing this the proof is complete. □

Lemma 1.13. *Let F_1 and F_2 lognormal distributions such that F_2 has a heavier tail than F_1. Then, there exists $c \in \mathbb{R}$ such that for all $y \leq x$ it holds that*

$$\frac{\overline{F}_1(x)}{\overline{F}_1(y)} \leq c\frac{\overline{F}_2(x)}{\overline{F}_2(y)}.$$

Proof. Let $\lambda_1(x)$, $\lambda_2(x)$ the corresponding failure rate functions of the lognormal distributions F_1 and F_2. First we will prove that there exist constants $c_1 > 0$ and $y_0 > 0$ such that the following inequality is true

$$-\lambda_1(t) \leq -\lambda_2(t) + c_1 \mathbb{I}_{[0,y_0]}(t).$$

For proving this, we will start from the inequality

$$\begin{aligned}[\lambda_1(t) - \lambda_2(t)]^+ &= \lambda_1(t) - \lambda_2(t) + [\lambda_2(t) - \lambda_1(t)]\ \mathbb{I}_{\{t:\lambda_1(t)<\lambda_2(t)\}}(t) \\ &\leq \lambda_1(t) - \lambda_2(t) + \lambda_2(t)\,\mathbb{I}_{\{t:\lambda_1(t)<\lambda_2(t)\}}(t),\end{aligned}$$

from where it follows that

$$-\lambda_1(t) \leq -\lambda_2(t) + \lambda(t)\mathbb{I}_{\{t:\lambda_1(t)<\lambda_2(t)\}}(t).$$

Since $\lambda_2(t)$ is real-valued on closed intervals of the type $[0, y_0]$ it remains bounded in there by continuity. So, it is just necessary to prove that $\{t : \lambda_1(t) < \lambda_2(t)\} \subseteq [0, y_0]$ for some $y_0 \in \mathbb{R}^+$. We consider the two possible cases in which $\overline{F}_1$ has heavier tail than

$\overline{F}_2$. In the first of them we consider $\sigma_1 < \sigma_2$. So we use the tail asymptotic expression for $\lambda(x)$ to obtain

$$\lim_{x\to\infty} \frac{\lambda_1(x)}{\lambda_2(x)} = \lim_{x\to\infty} \frac{\log x/x\sigma_1^2}{\log x/x\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2} > 1,$$

from where the conclusion follows easily. The second case comes when $\sigma_1 = \sigma_2$ and $\mu_1 < \mu_2$. For proving that $\lambda_2(x) \leq \lambda_1(x)$ we will just check that $\lambda(x,\mu)$ is a decreasing of function of μ. The derivative is given as

$$\begin{aligned}\frac{d}{d\mu}\lambda(x,\mu) &= \frac{\dfrac{\log x-\mu}{\sigma^2} f(x,\mu)\overline{F}(x,\mu) - f(x,\mu)\displaystyle\int_x^\infty \frac{\log t-\mu}{\sigma^2} f(t,\mu)dt}{\overline{F}^2(t,\mu)} \\ &= \frac{\log x f(x,\mu)\overline{F}(x,\mu) - f(x,\mu)\displaystyle\int_x^\infty \log t f(t,\mu)dt}{\sigma^2\overline{F}^2(t,\mu)}.\end{aligned}$$

The last expression is verified to be negative from the observation

$$\int_x^\infty \log t f(t,\mu)dt > \log x \int_x^\infty f(t,\mu)dt = \log x\overline{F}(x).$$

Then we just use this intermediate result to prove that

$$\begin{aligned}\frac{\overline{F}_1(x)}{\overline{F}_1(y)} = \exp\left\{-\int_y^x \lambda_1(t)dt\right\} &\leq \exp\left\{-\int_y^x \lambda_2(t)dt + \int_y^x c_1\mathbb{I}_{[0,y_0]}(t)dt\right\} \\ &\leq \exp\left\{-\int_y^x \lambda_2(t)dt + \int_0^{y_0} c_1 dt\right\} \\ &= \exp\left\{\log\frac{\overline{F}_2(x)}{\overline{F}_2(y)} + c_2\right\} = c\frac{\overline{F}_2(x)}{\overline{F}_2(y)}.\end{aligned}$$

□

Proof of Theorem 1.11. Recall that the condition for *asymptotic bounded relative error* is equivalent to

$$\lim_{u\to\infty} \frac{\mathbb{E}\left[\widehat{z}_{AK}^2(u)\right]}{\mathbb{P}(S_d > u)} < \infty.$$

By subexponentiality we have that $\mathbb{P}(X_k > u) = O(\mathbb{P}(S_n > u))$ for all k. Using this relation and the fact that all p_i's are all larger than 0 it will be enough to prove that

$$\limsup_{u\to\infty} \frac{\overline{F}_k^2(M_{-k} \vee (u - S_{-k}))}{\mathbb{P}^2(X_k > u)} < \infty \qquad k = 1,\ldots,d.$$

The idea will be to provide an upper bound where we get rid of the random variable S_{-k} since its distribution is unknown to us. For doing so, we divide the sample space

in two events, namely $A_1 = \{M_{-k} \leq u/2d\}$ and $A_2 = \{M_{-k} > u/2d\}$, and note that in A_1 the following relations hold

$$u - S_{-k} \geq u - nM_{-k} \geq u - u/2 = u/2 > u/2d \geq M_{-k}.$$

Using this we can obtain an upper bound in terms of M_d only

$$\begin{aligned}\frac{\mathbb{E}\left[\overline{F}_k^2(M_{-k} \vee (u - S_{-k}))\right]}{\overline{F}_k^2(u)} &\leq \mathbb{E}\left[\frac{\overline{F}_k^2(u - nM_{-k})}{\overline{F}_k^2(u)}; M_{-k} < u/2d\right] \\ &\quad + \mathbb{E}\left[\frac{\overline{F}_k^2(M_{-k})}{\overline{F}_k^2(u)}; M_{-k} > u/2d\right].\end{aligned}$$

So, with a simple change of variables we can rewrite this expression in integral form as follows

$$\int_0^{u/2} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_{M_{-k}}(y/d)dy + \int_{u/2d}^{\infty} \frac{\overline{F}_k^2(y)}{\overline{F}_k^2(u)} f_{M_{-k}}(y)dy.$$

The advantage of this bound is that the density of M_{-k} is known to us. In fact, this density is always smaller than the sum of the individual densities as can be seen from the following expression

$$f_{M_{-k}}(\cdot) = \sum_{i \neq k} f_i(\cdot) \prod_{j \neq i,k} F_j(\cdot) \leq \sum_{i=1}^{d} f_i(\cdot).$$

Inserting this new bound and taking the sum out of the integral we arrive to the conclusion that the estimator will have bounded relative error if

$$\limsup_{u\to\infty} \int_0^{u/2} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_i(y/d)dy + \int_{u/2d}^{\infty} \frac{\overline{F}_k^2(y)}{\overline{F}_k^2(u)} f_i(y)dy < \infty, \qquad i,k = 1,\ldots,d. \quad (1.9)$$

We prove separately that each of this two integrals remain bounded as $u \to \infty$. The first integral remains bounded due to Lemma 1.14. The second one is the easy since it can be evaluated directly using L'Hopital Theorem,

$$\lim_{u\to\infty} \frac{\int_{u/2d}^{\infty} \overline{F}_k^2(y) f_i(y)dy}{\overline{F}_k^2(u)} = \lim_{u\to\infty} \frac{\overline{F}_k^2(u/2d) f_i(u/2d)}{4d\,\overline{F}_k(u) f_k(u)} \to 0.$$

This limit can be easily verified using *Mill's ratio*. Putting together these results the result follows immediately.

□

Lemma 1.14. *Under the hypothesis of the Theorem 1.11 it holds that*

$$\lim_{u\to\infty} \int_0^{u/2} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_i(y/d)dy < \infty.$$

Proof. Consider

$$\overline{F}_k(u) = \exp\left\{ -\int_0^u \lambda(t)dt \right\},$$

where $\lambda(t)$ is the failure rate of the lognormal distribution and by standard subexponential theory we know that $\lambda(t)$ is asymptotically equivalent to $\frac{\log(u)}{\sigma^2 u}$. By choosing $c > \frac{1}{\sigma}^2$ we obtain that $c\frac{\log t}{t}$ is an asymptotic upper bound for $\lambda(t)$, then

$$\begin{aligned}\frac{\overline{F}_k(u-y)}{\overline{F}_k(u)} &= \exp\left\{ \int_{u-y}^u \lambda(t)dt \right\} < \exp\left\{ c\log u \int_{u-y}^u \frac{1}{t}dt \right\} \\ &= \exp\left\{ c\log u(\log u - \log(u-y)) \right\}.\end{aligned}$$

Using a first order Taylor expansion of $\log(\cdot)$ around $(u-y)$ and the fact that it is a concave function we have that $\log u < \log(u-y) + \frac{y}{u-y}$, so the last expression is bounded by

$$\exp\left\{ c\ \frac{y\log u}{u-y} \right\}.$$

Take $u > 1$. Our claim is that the set $\left\{ y : \log(2y) > \frac{y\log u}{u-y} \right\} = (g(u), u/2)$ for some function $g(u) \to 1/2$. This is true since both functions are increasing and equal when $y = u/2$, but $\log(2y)$ is concave and $\frac{y\log u}{y-u}$ is convex proving that there exists a smaller root than $u/2$. Next we verify that for any value $y_0 > 1/2$ there exists a value u_0 such that for all $u > u_0$ the inequality $\log(2y_0) > \frac{y_0\log u}{u-y_0}$ is fulfilled and therefore $g(u) < y_0$. We use this to get

$$\int_{y_0}^{u/2} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_i(y/d)dy < \int_{y_0}^{\infty} c_1 \exp\{c\log y\} f_i(y/d)dy = \int_{y_0}^{\infty} c_2\, y^c f_i(y/d)dy.$$

Since all the moments of a lognormal random variable are bounded we can conclude that the last expression is also bounded. For $y \in (0, y_0)$ we simply use the fact that a lognormal random variable belongs to the class $\mathcal{L}$, so we obtain

$$\int_0^{y_0} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_i(y/d)dy < \frac{\overline{F}_k^2(u-y_0)}{\overline{F}_k^2(u)} \to 1.$$

□

Proof of Theorem 1.12. Note that in the proof of Theorem 1.11 we did not make use of the hypothesis about the distribution up to (1.9). Hence, we can retake the proof from there so it remains to prove that the same holds for regularly varying distributions. That is

$$\int_0^{u/2} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_i(y/d)dy + \int_{u/2d}^{\infty} \frac{\overline{F}_k^2(y)}{\overline{F}_k^2(u)} f_i(y)dy < \infty,$$

where F_k is a regular varying distribution function with index α_k and f_i are densities of regularly varying random variables with indexes α_i. The first integral can be easy bounded with

$$\frac{\overline{F}_k^2(u/2)}{\overline{F}_k^2(u)} = 2^{-2\alpha_k} + o(1) \qquad u \to \infty.$$

For the second one we can use L'Hopital rule to obtain

$$\frac{\overline{F}_k^2(u/2d)}{\overline{F}_k^2(u)} f_i(u/2d) = (2d)^{2\alpha_k} f_i(u/2d) = o(1).$$

Putting together these two expressions we complete the proof. □

Chapter 2

Ruin Probabilities in a Finite Time Horizon

In this chapter we will study the *ruin probability in finite time* in risk models and its relation with sums of heavy-tailed i.i.d. random variables. In particular, the main objective will be the precise approximation of this probability via Monte Carlo methods.

In the Section 2.1 we provide some motivation for the study of the ruin problem. We characterize the stochastic processes which will be used as risk models and show how this problem is related to a tail probability of a sum of random variables. We continue with a brief exposition on Queueing Theory and comment on how risk processes are directly associated with a workload process in an M/G/1 queue. This relationship has been long understood and it will be useful to translate the results from one area into the other. For the remaining sections we have chosen to adopt the language of risk theory, but it should be understood that the stated results will also hold for the equivalent problems in Queueing Theory.

In the Section 2.2 we will consider the *ruin problem* and review some of the main results in the literature. In particular, we will focus on asymptotic approximations of the probability of ruin in infinite time with the purpose of illustrating the differences between models with light and heavy-tails. Then we will move to the case which interest us the most: *ruin probabilities in a finite time horizon in a heavy-tailed setting*. In there we will comment on the main asymptotic approximations available in the literature for this particular setting. In particular, the most interesting cases are those where the time horizon varies together with the level u so the time $t(u)$ will be a function of the level u.

In Section 2.3 we will study two Monte Carlo estimates proposed in Rojas-Nandayapa and Asmussen (2007) for estimating the ruin probability in finite time in the presence of heavy-tails. The key idea of the algorithms will be the conditional Monte Carlo estimator suggested in Asmussen and Kroese (2006). In particular, we are able to theoretically verify good efficiency properties in the most important case of subexponential distributions: regular variation. The first algorithm will be designed exclusively for the M/G/1–Cramér-Lundberg setting. The second algorithm applies to general Lévy processes and we also empirically demonstrate that it has excellent efficiency and prove bounded relative error under particular conditions on t, though at the moment we have not been able to come up with variance estimates quite as sharp as for the first algorithm in the general setting. Some general ideas into how to improve this algorithm are discussed in the section of Notes and Comments as well as some future work.

2.1 Risk/Queueing Processes

2.1.1 Risk Theory

Risk Theory is commonly associated with the mathematical problem faced by an insurance company which has to decide how much the premiums should cost to the insured in order to cover future claims without falling in bankruptcy. The basic *insurance risk model* is based in the following assumptions

- The i-th claim occur at a random times T_i. The times form a sequence of non-negative random variables satisfying $0 \leq T_1 \leq T_2 \leq \dots$
- The size of the i-th claim is a non-negative random variable U_i. The claim sizes will form a sequence of independent and identically distributed random variables.
- The size claims X_i and the time claims T_i are mutually independent.

We define the *claim arrival process* $(N(t))_{t\geq 0}$ as the total number of claims occurred during the time period $(0,t)$, and the *total claim amount process* $(S(t))_{t\geq 0}$ from adding up the claims derived in the time period $(0,t)$. Formally,

$$N(s) = \sup\{n : T_n \leq s\}, \qquad Y(s) = \sum_{i=1}^{N(s)} U_i,$$

the last with the understanding that $\sum_{i=1}^{0} := 0$.

In the basic model we assume that the interarrival times $\tau_i := T_i - T_{i-1}$ follow an exponential distribution with parameter λ. In fact this assumption fully characterizes the *claim arrival process* $N(t)$ which is turns into the well known *homogeneous Poisson process* and the *total claim amount process* is identified as a *compound Poisson process*. We will refer to this as the *Cramér-Lundberg risk model*. In fact, the more general model which allows the interarrival times $\tau_i = T_i - T_{i-1}$ to be an i.i.d. sequence of random variables is known as the *Renewal risk model*. In fact, this can be studied from the perspective of *Renewal theory* which is a branch of applied probability concerned with the probability of events directly associated with sums of independent random variables (cf. Feller, 1971; Asmussen, 2003).

By including a process $Z(s)$ which describes the dynamics associated with the flow of premiums and the initial capital u we arrive to the basic *risk process*

$$X(s) := u + Z(s) - Y(s) = u + Z(s) - \sum_{i=1}^{N(s)} U_i.$$

In the classical *Cramér-Lundberg–Renewal* setting, the *premium process* has the form $Z(t) = ct$ where $c > 0$ stands for the *premium income rate*. However, in recent years there has been a substantial interest in this model perturbed by a Brownian Motion or an infinite activity jumping part. A Lévy process describe precisely this kind of dynamics and therefore we partially deal with such extension and we will refer to it as the *Lévy risk model*. Our primary object of interest in this chapter will be the approximation of the *ruin probability in finite time* defined as

$$\psi(u,t) := \mathbb{P}\big(X(s) < 0, \text{ for some } s \in (0,t)\big), \qquad u,t \in \mathbb{R}^+.$$

From a theoretical point of view the following definition for infinite time will be also useful. That is, the *ruin probability in infinite time*

$$\psi(u) := \mathbb{P}\big(X(s) < 0, \text{ for some } s > 0\big), \qquad u \in \mathbb{R}^+.$$

2.1.2 Queueing Theory

Queueing Theory is the branch of Applied Probability which deals with problems for waiting lines or *Queues*. In Kendall's notation, queues are represented under the form

$$A/S/NS/NC$$

where A and S are strings describing the distribution of the interarrivals and service times respectively and NS and NC are integer numbers corresponding to the number of servers and the capacity of the queue. The two most common string values for both interarrival and service times are: M for Markovian referring to the exponential distribution, G or GI for General an General Independent. By tradition GI is employed for the interarrival distribution and G for the service times.

The service times $U_1, U_2, \ldots$ will form a sequence of nonnegative i.i.d. random variables and $N(t)$ will denote the total amount of customers which have joined the queue until time t. There exist several stochastic processes associated with a queue. The following ones are among the most studied in the literature.

- The *queue length* at time t or simply the number of customers at time t. It is commonly denoted $\{Q(t)\}_t$.
- The *waiting time* W_k of the k-th customer or simply the time that a given customer spends in a queue until the start of its service.
- The *workload process* $\{W(t)\}_t$ is the total time – counted from t – that the systems needs to clear the entire queue. In particular, it should be noted that $\{W(t)\}$ is not an adapted process for the natural filtration. This simply means that at time t the workload is not observable since it is equal to the aggregation not only of the services times of the customers in the queue at that moment, but also those arriving before the queue is emptied.

Here, we will be interested in the *M/G/1 queue*. More precisely, customers arrive at a single service desk (server) according to a Poisson process and have service times which are independent and identically distributed. In particular, given that the queue has never been empty in the time interval $(0, t)$ we have the following relation

$$V(t) = \sum_{i=1}^{N(t)} U_i - t,$$

where $N(t)$ is a Poisson process, say with intensity λ, and the negative drift $-t$ represents the time passing by as the server deals with the customers.

Asmussen and Petersen (1989) showed the existence of the duality between the workload of an *M/G/1* queue and the probability of ruin in finite time. More precisely,

$$\psi(u, t) = \mathbb{P}(V(s) > u),$$

where $(V(s))_{s\geq 0}$ is the steady–state workload process of an initially empty *M/G/1* queue. This duality will allow us to speak freely of results which are valid for both an *M/G/1* queue and a Cramér-Lundberg risk model.

For more details on Queueing Theory see Asmussen (2003) and Robert (2003).

2.1.3 Conventions and Notation

For the rest of the chapter we will adopt the following notation and conventions. For an arbitrary process $(X(s))_{s\geq 0}$ it will be useful to define the random variable

$$\overline{X}(t_1, t_2) = \sup_{t_1 \leq s < t_2} \{X(s) - X(t_1)\},$$

that is, the record value of the process during the time period (t_1, t_2). We will adopt the common practice of starting the process from 0, reverting the sign, so the new process will be defined as

$$X(t) := \sum_{i=1}^{N(s)} U_i - Z(s).$$

So $\psi(u, t)$ will be equivalent to the probability that the record value in the time period $(0, t)$ of the new process is larger than u. This modification simplifies the exposition of the results and makes evident the relation of risk/workload models with large values of sums of random variables.

We will reserve the use of $\mu_F := \mathbb{E}\, U_i$ for the common expectation of the jumps and let μ be the drift of the process $X(t)$ or equivalently the expected increment of the process per unit of time. In the renewal model this is

$$\mu := \mathbb{E}\, \overline{X}(t+1, t) = \mu_F\, \lambda - c,$$

where $\lambda := \mathbb{E}\, N(1)$ and therefore $\mathbb{E}\, X(t)/t \to \mu$. In fact, the quantity $\rho := \lambda\, \mathbb{E}\, N(1)$ will have a practical significance and it is commonly known as the *traffic intensity* or *expected claim amount per unit of time*. Also note that in the *Cramér-Lundberg* model, λ is the intensity parameter of the Poisson process. In order to simplify the notation we will assume in further sections that the *premium income rate* is equal to one. For our purposes this will not be a problem since it can be easily generalized by a simple time change argument. For instance, in the Cramér-Lundberg model this means that the parameter λ of the Poisson process should be changed by λ/c.

2.2 Ruin Probabilities

From a practical point of view it should be reasonable to require $\psi(u) < 1$. This just says that the insurance company has a positive probability of avoiding bankruptcy. Then it is obvious that the process should have a negative drift – recall that we have inverted the sign of the process. Therefore we should have

$$\mu = \lambda \mu_F - c < 0,$$

or equivalently $\rho/c < 1$. This is known as the *net profit condition* or simply *stability condition*. In fact, it is of a major importance for an insurance company to choose the *premium rate* c that should be charged to the insured in order to ensure the solvency of the company. Moreover, most of the results presented below require this condition

to hold. Therefore, we assume for the rest of the chapter that the *net profit condition* is satisfied except where noted.

The celebrated *Pollaczeck-Khinchine formula* provides an equivalent representation of the probability of ruin in infinite time (cf. Feller, 1971; Asmussen, 2000, 2003). For instance, in the *Cramér-Lundberg* case this is given as

$$\psi(u) = (1-\rho)\sum_{n=0}^{\infty} \rho^n \overline{F}_I^{n*}(u), \tag{2.1}$$

where F_I^{n*} is the n-fold convolution of the distribution defined as

$$F_I(x) = \mu_F^{-1} \int_0^x \overline{F}(y)dy.$$

which is known as the *integrated tail distribution*. This formula is obtained by means of the *Wiener-Hopf factorization* which is a fundamental result in *Random Walk Theory*.

2.2.1 Ruin in the Light Tailed Setting

The use of the Pollaczeck-Khinchine formula is limited for explicit calculations of the ruin probability because of the difficulty of evaluating the n-fold convolution. In consequence there are just a few cases where explicit calculations can be done – the exponential distribution is an example. Nevertheless, the Pollaczeck-Khinchine formula has been used to provide precise asymptotic expressions for $\psi(u)$. In particular, the so called *Cramér-Lundberg estimates* for light-tailed random variables are derived from it. These are given next.

Definition 2.1 (Lundberg exponent)**.** Let F be a distribution function such that the Laplace-Stieltjes transform of its *integrated tail distribution* F_I is defined, i.e.

$$\mathcal{L}F_I^*(\theta) := {\mu_F}^{-1} \int_0^\infty e^{\theta t}\overline{F}(t)dt < \infty.$$

The *Lundberg exponent* γ is defined as the value θ which solves $\mathcal{L}F_I^*(\theta) = c/\rho$.

The following is a classical result in Risk Theory.

Proposition 2.2 (Cramér-Lundberg Theorem)**.** *Consider the Cramér-Lundberg model. If the* Lundberg exponent γ *exists then*

- *The following relation known as* Lundberg's inequality *holds*

$$\psi(u) \leq e^{-\gamma u}.$$

- *Moreover, the last result can be refined into the following asymptotic approximation known as* Cramér approximation*.*

$$\lim_{x\to\infty} \frac{\psi(u)}{e^{-\gamma u}} = C < \infty,$$

where the value of the constant C is given by

$$C := \frac{1-\rho}{\gamma\lambda \int_0^\infty t\mathrm{e}^{\gamma t} F_I(t)dt}.$$

2.2.2 Ruin in the Heavy-Tailed Setting

The most realistic models are those which involve heavy-tailed random variables. In Section 1.2 on page 3 we saw that heavy-tailed random variables can be distinguished for having an infinite moment generating function or equivalently a infinite moment generating function for positive values of the argument. Therefore, the *Cramér-Lundberg estimates* will not be available for this kind of distributions. However, it is precisely the Pollaczceck-Khinchine formula (2.1) from where we derive precise asymptotic expressions for the class $\mathcal{S}$ of subexponential distributions – recall that the defining property of the class $\mathcal{S}$ is given in terms of the n-fold convolution. More precisely:

Proposition 2.3. *Let F be a distribution function such that the corresponding integrated tail distribution $F_I \in \mathcal{S}$ is subexponential. Then, in the renewal model it holds that*

$$\lim_{u\to\infty} \frac{\psi(u)}{\overline{F}_I(u)} \;=\; \frac{\rho}{1-\rho}. \tag{2.2}$$

This results is due to Embrechts and Veraverbeke (1982). However, Mikosch and Nagaev (2001) showed that the rate of convergence of this estimate is rather slow. So, Monte Carlo methods appear an appropriate approach.

So far, we have only consider the probability of ruin in infinite time. A detailed exposition on the probability of ruin in finite time for both light and heavy-tailed settings can be found in Asmussen (2000). Here, we will show some recent results for the heavy-tailed set which are valid for the renewal model and even for the more general case of the *Lévy risk model*. In particular, the more interesting case is when the time horizon t varies together with the level u, so we consider instead $\psi(u, t(u))$ where $t(u)$ is a function of u.

In Asmussen and Kluppelberg (1996), asymptotic expressions for $\psi(u, t(u))$ are given when $t(u)$ is of the same order of the *mean excess function* $e(u) = \mathbb{E}\,[U - u \,|\, U > u]$. In the regularly varying case this means that $t(u)/u \to k \in (0, \infty)$. When $c = 1$, the Corollary 1.6 from the cited reference states that

$$\lim_{u\to\infty} \frac{\psi(u, t(u))}{\psi(u)} \;=\; 1 - (1 + (1-\rho)k)^{-\alpha+1} \;>\; 0\ . \tag{2.3}$$

The same result holds in a general Lévy process, as can be seen from Klüppelberg et al. (2004) after some rewriting. If instead $t(u)/u \to 0$, then Foss et al. (2005); Tang (2004) imply

$$\lim_{u\to\infty} \frac{\psi(u, t(u))}{\lambda\, t(u) \overline{F}(u)} \;=\; 1. \tag{2.4}$$

This result does not require the stability condition $\rho < 1$; the given references only give the result for discrete-time random walks, but it is not difficult to extend it to the M/G/1–Cramér Lundberg setting and even to a general Lévy process by writing $X = Y + Z$ and treating Z as a light-tailed perturbation of Y.

2.3 Monte Carlo Estimation of the Ruin Probability

In the light-tailed case, it has long been understood how to perform efficient simulation of $\psi(u)$ as well as of $\psi(u, t)$. The basic tool is importance sampling (cf. Asmussen, 2000) where the algorithms described there exploit variants of large deviations ideas

(cf. Siegmund, 1976; Bucklew et al., 1990; Anantharam, 1989). In the heavy-tailed case, all algorithms has so far only dealt with the infinite horizon case and the Cramér-Lundberg setting. Here efficient simulation algorithms for $\psi(u)$ have been developed in a number of recent papers, in particular Asmussen and Binswanger (1997); Asmussen et al. (2000); Juneja and Shahabuddin (2002); Asmussen and Kroese (2006); Dupuis et al. (2006). All of these heavily rely on the Pollaczek-Khinchine formula, expressing $\psi(u)$ as the tail probability of a geometric sum. Recently Blanchet and Glynn (2008) proposed an efficient rare-event simulation algorithm for a G/G/1 queue for a large class of sub-exponential distributions. For the finite horizon problem, the only reference we know of is Boots and Shahabuddin (2001). For the algorithm in that paper to be efficient, it is, however, needed that u and t vary together in a specific manner, and that F is not too far from the heavy-tailed Weibull.

The algorithms proposed here are essentially based on the proposal given in Asmussen and Kroese (2006). Let us recall that the estimator in that paper is constructed conditioning to the event that a given random variable takes the maximum and use the i.i.d. hypothesis to unbias the estimator using the following symmetry argument

$$\mathbb{P}(S_d > x) = d\,\mathbb{P}(S_d > x, X_d = M_d), \tag{2.5}$$

so, the estimator takes the form

$$z_{AK}(u) = d\overline{F}(M_{d-1} \vee (u - S_{d-1})).$$

In the first algorithm the role of $X_1, X_2, \ldots$ will be taken by the ladder heights (cf. Asmussen, 2000) and in the second one by the jumps of the compound Poisson part Y. However, both algorithms need substantial modification. In particular, the difficulty lies in the presence of the *premium process* $Z(s)$ which destroys exchangeability. For example, if $Z(s)$ has negative drift, then an early large jump of Y is more likely to cause ruin than a late one, because by the late time the negative drift is likely to have taken X to a smaller value.

For the rest of the chapter we will assume that F is such that its integrated tail distribution F_I belongs to the class of regularly varying distributions $\mathcal{R}$. By Karamata's Theorem this just means that if F is regularly varying with index $\alpha + 1$, then F_I is regularly varying with index α.

2.3.1 Cramér–Lundberg Model

We restrict here our attention to the M/G/1 queue or equivalently the Cramér-Lundberg risk model, which allows us to use the sample path decomposition of the process $\{X(s)\}$ according to ladder steps (cf. Asmussen, 2000). Let the sequence of random vectors $\{(T_i^0, V_i, W_i) : i \in \mathbb{N}\}$ be the lengths of the ladder segments, the ladder heights, and the deficits before ladder epochs as illustrated in Fig. 2.1.

From the Pollaczeck-Khinchine formula it follows that if $\rho < 1$, then the number of ladder steps L is Geometric(ρ), different ladder segments are independent and the joint distribution of (T_i^0, V_i, W_i) is as described in the following Proposition, where parts (i), (ii) are classical and part (iii) is from Asmussen and Kluppelberg (1996):

Proposition 2.4. *Under* $\mathbb{P}^{(0)} = \mathbb{P}(\cdot|\tau(0) < \infty)$, V_i, W_i *have the same joint distribution as the backward and forward recurrence time distribution in a renewal process with interarrival distribution* F. *That is,*

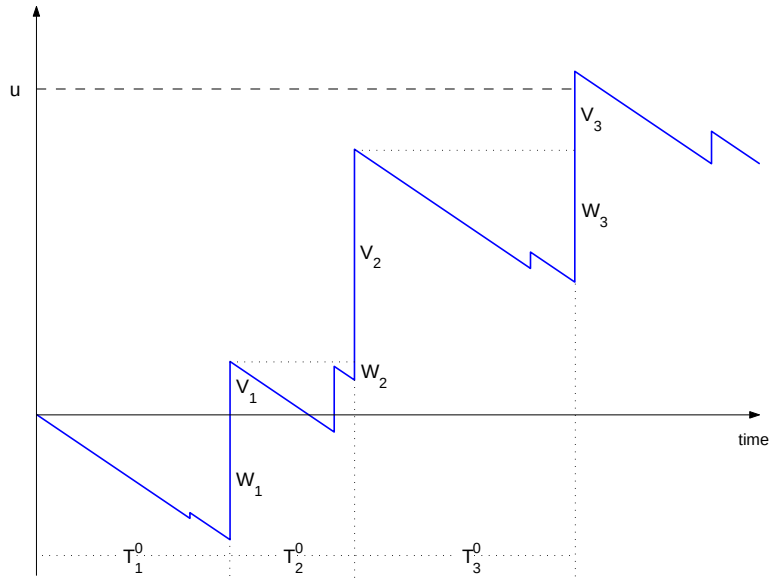


Figure 2.1: The ladder structure

1. *The marginal distributions of* V_i, W_i *are both* F_I*.*

2. *The conditional distribution of* $(W_i|V_i = v)$ *is the same as the distribution of* $(U - v|U > v)$ *where* U *is a r.v. with distribution* F*.*

Moreover,

3. *The conditional distribution of* $(T_i^0|W_i = w)$ *is the same as the unconditional distribution of* $\inf\{t > 0 : R(t) = w\}$*, where the process* R *is given by* $R(t) = t - \sum_{i=1}^{N^0(t)} U_i^0$ *and is an independent copy of* $-S$*.*

In the Cramér-Lundberg model ruin does not occur without ladder steps. So, we can write $\mathbb{P}(\tau(u) < t) = (1-\rho)\,\mathbb{P}(\tau(u) < t|L > 0)$ – recall $L \sim$Geo(ρ). In fact, ruin will occur if the aggregated ladder heights surpass the level u. In that case there will be an index K corresponding to the first partial sum $X_1 + \cdots + X_K$ which surpassed the level u. Hence, ruin will happen before time t if the time of the K-th ladder step is smaller than t. Thus

$$\{\tau(u) < t\} = \left\{V_1 + \cdots + V_L > u,\, T_1^0 + \cdots + T_K^0 < t\right\}.$$

Define $\overline{R}(t) = \overline{R}(0,t)$ and note that the upward skipfree property of R (cf. Asmussen, 2003), implies that the process of first passage time to different levels has stationary independent increments so that

$$\mathbb{P}(\tau(u) < t) \;=\; \mathbb{P}\big(V_1 + \cdots + V_L > u,\, W_1 + \cdots + W_K < \overline{R}(t)\big)\,. \tag{2.6}$$

The estimator $\mathbb{I}\big\{(V_1 + \cdots + V_L > u,\, W_1 + \cdots + W_K < \overline{R}(t)\big\}$ can be viewed as the crude Monte Carlo estimator based on Lemma 2.4, and we proceed to develop some ideas that will reduce this estimator's variance.

Since we have related the event of ruin to the tail probability of sum of independent random variables with heavy-tails we can adapt some of the existing ideas for rare event

simulation. It is well known, from subexponential theory, that ruin happens as a consequence of a single large claim. The conditional algorithms proposed by Asmussen and Binswanger (1997); Asmussen and Kroese (2006) were constructed under this premise. Here we adapt the algorithm proposed Asmussen and Kroese (2006) to the present setting as follows.

The idea of the algorithm is to induce a large jump by defining J to be an independent random variable with a discrete distribution on $\{1, \dots, L\}$. The random variable J is slightly artificial in this setting but it can be interpreted as the (random) index of the variable which becomes the maximum. In Rojas-Nandayapa and Asmussen (2007) this random variable was taken to be discrete uniform, but it is clear that a late large jump could occur out of the time horizon, so it might produce better results to choose a different configuration for this random variable which favors early larger jumps.

We replace the symmetry argument in (2.5) by

$$\mathbb{P}(\tau(u) < t) = \mathbb{E}\left[L; \tau(u) < t,\, V_J = \max\{V_i : i \leq L\},\, L > 0\right]$$

and compute our estimator as the conditional probability of (2.6) given

$$\mathcal{F} = \sigma\big(L, J, \{V_j : j \neq J\}, \{R(s) : s \leq t\}\big)\,,$$

the σ-field containing the information on the random variable J, the process $\{R(s)\}$ up to time t and the number and sizes of the ladder heights V_i except of the J-th. From (2.6) this conditional probability comes out as

$$\mathbb{P}\big(\tau(u) < t|\mathcal{F}\big) = L\,\mathbb{P}\big(W_1 + \cdots + W_K < \overline{R}(t),\, V_J > H \,\big|\, \{V_j : j \neq J\}, J\big),$$

where

$$H = \Big(u - \sum_{j \neq J} V_j\Big) \vee \sup\big\{V_j : j \neq J\big\}.$$

Moreover, since it is easy to simulate random variables conditioned to some interval, we can use a conditional argument over the event $\mathbb{I}(V_J > H)$ to rewrite the last probability as

$$L\,\overline{F}_I(H)\mathbb{P}\big(W_1 + \cdots + W_K < \overline{R}(t) \,\big|\, \{V_j : j \neq J\},\, V_J > H\big). \tag{2.7}$$

However, since the size of the J-th ladder step is unknown it turns out that the random variable K is not measurable with respect the σ-field generated by $\big\{\{V_j : j \neq J\},\, \mathbb{I}(V_J > H)\big\}$ and therefore (2.7) should be calculated via Monte Carlo. We let Q be the probability measure under which the variable V_J has the conditional distribution of $V_J|(H, \infty)$ under the original measure and $\mathbb{E}^{\,Q}$ the corresponding expectation. Then

$$\begin{aligned}
&\mathbb{P}\big(W_1 + \cdots + W_K < \overline{R}(t) \,\big|\, \{V_j : j \neq J\},\, V_J > H\big) \\
&\quad = \mathbb{E}^{\,Q}[\mathbb{P}\big(W_1 + \cdots + W_K < \overline{R}(t) \,\big|\, \{V_j : j \neq J\},\, V_J\big) \,|\, \{V_j : j \neq J\}, J].
\end{aligned}$$

The last step of the algorithm is to reduce the variance coming from the set $\{W_i\}$ as follows: Since $\overline{R}(t) < t$, we have for $i \leq K$ that $\tau(u) \geq t$ if $W_i \geq t$. So, we let $\widetilde{Q}$ to be the probability measure under which each of the W_i's have the conditional distribution of $W_i|\{V_i, W_i < t\}$ under the measure Q. Then

$$\begin{aligned}
&\mathbb{P}\big(W_1 + \cdots + W_K < \overline{R}(t) \,\big|\, \{V_j : j \neq J\},\, V_J^*\big) \\
&\qquad = \widetilde{Q}(W_1 + \cdots + W_K < \overline{R}_t) \prod_{i=1}^{K} \mathbb{P}(W_i < t|V_i).
\end{aligned}$$

Conditional Algorithm: Ruin in the Cramér-Lundberg setting

1. Simulate L, J, $\{V_i : i \neq J\}$ with distributions as above (in particular, $\mathbb{P}(L = n) = (1-\rho)\rho^{n-1}$, $n = 1, 2, \dots$). To simulate R, generate $M = N_t^0$ as Poisson(λt), the jumps $U_1^0, \dots, U_M^0$ as i.i.d. with distribution F, and the jump times $\pi_1, \dots, \pi_M$ as the order statistics from the uniform distribution on $(0, t)$. Let $\pi_{M+1} = t$, $U_0 = 0$ and
$$\overline{R}(t) = \max_{1 \leq k \leq M+1} \{\pi_k - \sum_{i=0}^{k-1} U_i^0\}.$$
2. Calculate H and simulate the random variable $V_J|(H, \infty)$.
3. Identify K and simulate the random variables $\{W_i|\{V_i, W_i < t\} : i \leq K\}$.
4. Return the estimator
$$\widehat{\psi}_A(u) = (1-\rho)\, L\, \overline{F}_I(H)\, \mathbb{I}(W_1^* + \cdots + W_K^* < \overline{R}(t)) \prod_{i=1}^{K} \mathbb{P}(W_i < t|V_i).$$

If we consider the time horizon t to be fixed it is easy to prove that the algorithm above has bounded relative error. The interesting part comes when we consider a finite time horizon which goes to infinity together with the level u. That is we assume that $t = t(u)$ to be a function of u and consider the limit $u \to \infty$.

Intuitively, if the growth rate of $t(u)$ is large with respect to u, the probability $\psi(u, t(u))$ will look more like the ruin probability in an infinite time horizon. By slowing the growth rate, the event will become rarer and more interesting for our purposes. The following Theorem establishes the asymptotic efficiency of this estimator and its proof can be found on page 35.

Theorem 2.5. *Assume $\alpha > 1$ and*
$$\mathbb{P}(U - v < t|U > v) \;=\; O(1)\frac{t}{v}. \tag{2.8}$$
where the $O(1)$ is uniform in $0 < t < t_0$ for any $t_0 < \infty$. Then the estimator $\widehat{\psi}_A(u, t)$ has bounded relative error when $t \to \infty$ with u in such a way that $t/u \to k \in [0, \infty)$.

Note that (2.8) is very weak. It holds, for example, if F has a density $f(x)$ satisfying $f(x) \sim \ell_1(x)/x^\beta$ with ℓ_1 slowly varying (integration shows that one must have $\beta = \alpha+1$, $\ell_1(x) \sim (\alpha+1)\ell(x)$, as is seen by straightforward calculus).

This estimator can in a straightforward way be generalized to a discrete time random walk $X_N = Y_1 + \cdots + Y_N$ with increment distribution F. For example, one replaces N_t by t, T_i by i, $\overline{X}(t_1, t_2)$ by $\max_{i_1 \leq i < i_2} X_i$ and $\sup_{i \neq J} U_i$ by $\max_{j \leq n, j \neq J} Y_j$. Also the derivation of the variance estimates is entirely similar.

2.3.2 Lévy Model

Going beyond the M/G/1–Cramér-Lundberg setting, we assume that the process X is a general Lévy process. As is well known, X can be written as the independent sum of a Brownian motion and a pure jump part, whose construction may involve compensation.

All that will matter for us is the upper tail, and we use therefore the alternative decomposition $X = Y + Z$ where Y is compound Poisson with positive jumps bounded

below say by 1, so that $Y(s) = \sum_1^{N(s)} U_i$, where $\{U_i : i \in \mathbb{N}\}$ are i.i.d. with common distribution F supported over $\mathbb{R}^+$, $\{N(s)\}$ is an independent Poisson process with rate λ and interarrival times $\{T_i : i \in \mathbb{N}\}$. In the Cramér-Lundberg risk model, $Z(s) = -cs$ (here the upward jumps correspond to claims), but it should be noted that the representation $X = Y + Z$ is completely general and can be achieved for any Lévy process by just letting Y be the sum of jumps of size, say at least 1 and $Z = X - Y$ (then Z is the sum of a Brownian component, possibly with drift, and a jump part with jumps bounded by 1).

When simulating X, the compound Poisson part Y is of course straightforward to generate. The remaining Lévy component Z may require more sophistication, and we refer to Asmussen and Glynn (2007) for a survey of methods. Note, however, that Y and Z need not be simulated separately, but X can be simulated and next Y extracted as the jumps > 1.

It follows from Asmussen and Kluppelberg (1996); Foss et al. (2005); Tang (2004) that ruin occurs with high probability as consequence of one big jump. The idea here

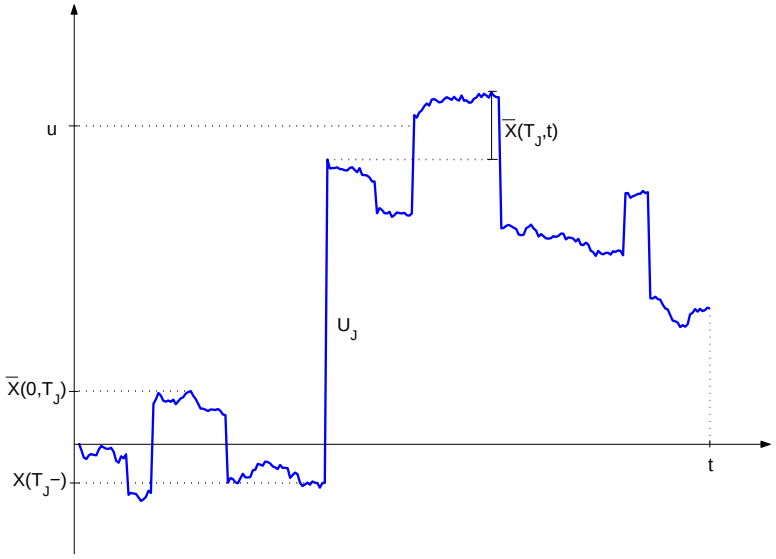


Figure 2.2: A sample path of X with a big jump U_J

is similar to that of the previous section in order to adapt the conditional algorithm in Asmussen and Kroese (2006). First we define a r.v. J which given $\{N(s)\}_{s\leq t}$ has a discrete uniform distribution over $\{1, \ldots, N(t)\}$. Then, we replace the symmetry argument in (2.5) by

$$\begin{aligned}\mathbb{P}(\tau(u) < t) = \mathbb{E}\left[N(t);\, \tau(u) < t,\, U_J = \max\{U_i : i \leq N(t)\},\, N(t) > 0\right] \\ + \mathbb{P}(\tau(u) < t, N(t) = 0).\end{aligned}$$

Note that in this case, if no jumps of the Poisson process occur, ruin still can happen as a consequence of the Lévy process alone as indicated in the last term. Next, we compute the estimator as the conditional probability of $\{\tau(u) < t\}$ given

$$\mathcal{F} \;=\; \sigma\big(J,\, \{N(s)\}_{s\leq t},\, \{U_j : j \neq J\},\, \{Z(s)\}_{s\leq t}\big),$$

the σ-field containing all information about the r.v. J and the process X except the size of jump J. This conditional probability comes out as

$$\mathbb{P}\big(\tau(u) < t \,\big|\, \mathcal{F}\big) = \begin{cases} \mathbb{I}(\overline{Z}(0,t) > u) & N(t) = 0 \\ N(t)\,\mathbb{P}\big(\tau(u) \le t,\, U_J = \max\{U_i : i \le N(t)\}\,|\mathcal{F}_{\mathrm{A}}\big) & N(t) > 0. \end{cases}$$

If $N(t) > 0$, ruin occurs typically as consequence of the big jump U_J or possibly if the value of the process after jump J exceeds u. That is, if

$$X(T_J-) + U_J + \overline{X}(T_J, t) > u,$$

where $X(T_J-)$ is the value of the process just before jump J. However, we also have to take into account the possibility that ruin occurs before the time of jump J (i.e., that $\tau(u) < T_J$). These arguments are illustrated in Fig. 2.2. In conclusion,

$$\mathbb{P}\big(\tau(u) < t,\, U_J = \max\{U_i : i \le N(t)\}\,|\mathcal{F}\big) = \mathbb{P}(U_J > W),$$

where

$$W \;:=\; \sup\{U_i : i \ne J\} \vee \big[\big(u - X(T_J-) - \overline{X}(T_J, t)\big)\mathbb{I}(\tau(u) \ge T_J)\big].$$

Here we used the $\mathcal{F}$-measurability of $\mathbb{I}(\tau(u) \ge T_J)$, $\overline{X}(T_J, t)$ and $X(T_J-)$ to compute the conditional probability $\mathbb{P}(\tau(u) < t|\mathcal{F})$.

Conditional Algorithm: Probability of Ruin in the Lévy Model

1. Simulate the Poisson process $\{N(s)\}_{s\le t}$ by generating $N(t) = N_t$ as Poisson(λt) and given that $N_t \ge 1$, the jump times $T_1 < T_2 < \cdots < T_{N_t}$ as the order statistics from the $(0,t)$-uniform distribution.
2. Simulate the whole of $\{Z(s) : s \le t\}$ (we again refer to Asmussen and Glynn (2007) for this step). If $N_t = 0$ return $\mathbb{I}(\overline{Z}(0,t) > u)$, else go to 3.
3. Simulate J as a discrete uniform r.v. over $\{1, \ldots, N_t\}$ and the U_i, $i \ne J$, from F.
4. Calculate W and return
$$\widehat{\psi}_B(u,t) \;=\; N_t\overline{F}(W)$$

We state the main theoretical results on this algorithm where we assume $t = t(u)$ to be a function of u and consider the limit $u \to \infty$.

Theorem 2.6. *Assume that $t/u \to 0$. Then for each $\epsilon > 0$, one has*

$$\mathbb{V}\mathrm{ar}\,\widehat{\psi}_B(u,t) \;=\; t^{2(\alpha+1)}\overline{F}^2(u)\,O(\mathrm{e}^{\epsilon u})\,.$$

where $f = O(\phi)$ means that $|f(x)| < A\phi(x)$ for some constant A and all x.

The proof of this Theorem can be found on page 36. Note that this result does not require $\alpha > 1$ or $\mu < 0$. The result is, however, somewhat weaker than logarithmic efficiency: then the power of t should have been 2. The bound provided by Theorem 2.6 is sharpest when t grows relatively slowly with u. In particular, if $t = u^\beta$ with $0 \le \beta < 1$, then Theorem 2.6 guarantees that $\widehat{\phi}_B$ provides an improvement over the crude Monte Carlo algorithm only when $\beta < \alpha/(2\alpha+1)$ (but of course, the bound of Theorem 2.6 could be too rough). However:

Theorem 2.7. *Assume $\alpha > 1$ and $t = u^\beta$ with $0 < \beta < \alpha/(1+\alpha)$. Then the estimator $\widehat{\psi}_2(u,t)$ has bounded relative error.*

The proof of this Theorem can be found on page 38.

2.4 Concluding Remarks

2.4.1 Numerical Examples

In this section, we discuss the numerical examples contained in Examples D.7 to D.10 on pages 111–114 to illustrate the performance of our algorithms. In the four examples we considered the M/G/1–Cramér-Lundberg setting taking the jumps U_i of the compound Poisson process to be Pareto with parameter α. That is, the distribution function is given by $\overline{F}(x) = (1+x)^{-\alpha}$.

For each of the examples we have tested the two estimators and have included the asymptotic approximations 2.3 to 2.4 on page 28 for comparison purposes. The parameters where chosen in order to test our algorithms in a variety of scenarios involving short and long time horizons with moderately heavy and very heavy random variables. The conclusion of the numerical examples is that both algorithms appear to give excellent results. It is notable, however, that according to the Time-Relative Errors the algorithm designed for the Lévy model has a substantially better performance than the one designed for the Cramér-Lundberg model. Of course, this can be implementation-dependent.

Notes and comments

A further improvement of the algorithms proposed above is to choose a different discrete distribution for the index random variable of the forced largest jump. The idea is that early jumps are most likely to cause ruin because in late times the process is more likely to have smaller values due to the negative drift of the process, but also because the finite horizon time.

There is a growing interest in *adaptive algorithms* for estimating the probability of a rare event. Roughly speaking, an adaptive algorithm is characterized by updating and learning steps that will improve the efficiency of the final algorithm. Blanchet and Glynn (2008) proposed the first efficient rare-event adaptive algorithm for a $G/G/1$ queue for a large class of subexponential distributions (see also Dupuis and Wang, 2004; Dupuis et al., 2007; Blanchet et al., 2007). Currently, in a joint effort with Søren Asmussen and José Blanchet there is work in process to adapt the ideas of such algorithm to the finite horizon problem described here.

2.5 Proofs

Proof of Theorem 2.5. For an upper bound of the variance of the estimator B we have

$$\begin{aligned}\mathbb{V}\mathrm{ar}\,[\widehat{\psi}_A(u)] &\leq \mathbb{E}\,[L^2\overline{F}_I^2(H)\mathbb{P}^2(W < t(u)|V = V_{\max})] \\ &= \mathbb{E}\,[L^2\overline{F}_I^2(H)\mathbb{P}^2(U - V_{\max} < t(u)|U > V_{\max})],\end{aligned}$$

where $V_{\max}$ is the the largest value among $V_1, \ldots, V_K$. Next observe that the process is simulated in such way that ruin occurs with probability 1, then it is not difficult to verify that $H > u/L$ and $V_{\max} > u/L$. Since $H > u/L$, the stated hypothesis on the overshoot distribution gives the following upper bound for $\mathbb{V}\mathrm{ar}\,[\widehat{\psi}_A]$

$$\begin{aligned}&\mathbb{E}\,[L^2\overline{F}_I^2(u/L)\mathbb{P}^2(U - u/L < t(u)|U > u/L)] \\ &\quad = \; O(1)\frac{t^2}{u^2}\mathbb{E}\,[L^4\overline{F}_I^2(u/L);\, L \leq u] \,+\, \mathbb{E}\,[L^2; L > u]\,. \qquad (2.9)\end{aligned}$$

Now comparing (2.2) with (2.3) when $k > 0$ and with (2.4) when $k = 0$ shows that $\psi(u,t)$ is always of order $t\,\ell(u)/u^\alpha$. Since $k < \infty$, the second term in (2.9) is therefore $o\big(\psi(u,t)^2\big)$. Dividing the first term by $\psi(u,t)^2$, we obtain

$$O(1)u^{2\alpha-2}\mathbb{E}\left[L^4\frac{\overline{F}_I^2(u/L)}{\ell(u)^2};\, L \le u\right] \;=\; O(1)\mathbb{E}\left[L^{2\alpha+4}\frac{\ell(u/L)^2}{\ell(u)^2};\, L \le u\right]. \tag{2.10}$$

The proof will therefore be completed if we can show that the r.h.s. of (2.10) remains bounded as $u \to \infty$. Less restrictively, it is easy to see from the above analysis that it suffices to show this assertion with the qualifier $L \le u$ replaced by $L \le au$ for some $a > 0$. To this end, write

$$\frac{\ell(u/L)}{\ell(u)} \;=\; \frac{\ell(u/L)}{\ell(u/(L-1))}\cdot\frac{\ell(u/(L-1))}{\ell(u/(L-2))}\cdots\frac{\ell(u/2)}{\ell(u)}.$$

From the uniform convergence theorem for slowly varying functions (see the Appendix of Embrechts et al., 1997), it follows that given $\epsilon > 0$, there exists u_0 such that

$$\frac{\ell(ut)}{\ell(u)} \;\le\; 1+\epsilon \quad \text{for all } 1/2 \le t \le 1 \;\text{ and all } u \ge u_0\,.$$

For $u \ge u_0$, the r.v. $L^{2\alpha+4}\ell(u/L)^2/\ell(u)^2\cdot\mathbb{I}\{L \le u/u_0\}$ is therefore bounded by $L^{2\alpha+4}(1+\epsilon)^{2L}$, which is integrable if ϵ is so small that $\rho(1+\epsilon)^2 < 1$. Since $L^{2\alpha+4}\ell(u/L)^2/\ell(u)^2\cdot\mathbb{I}\{L \le u\} \to L^{2\alpha+4}$ a.s. by the definition of a slowly varying function, dominated convergence therefore gives

$$\mathbb{E}\left[L^{2\alpha+4}\frac{\ell(u/L)^2}{\ell(u)^2};\, L \le u/u_0\right] \;\to\; \mathbb{E}\,L^{2\alpha+4} \;<\; \infty\,,$$

and the proof is complete. □

Proof of Theorem 2.6. We start with an upper bound for the variance of the estimator.

$$\begin{aligned}\mathbb{V}\mathrm{ar}\,[\widehat{\psi}_B(u,t)] \le \mathbb{E}\,[\widehat{\psi}_2^2(u,t)] &= \mathbb{E}\left[\widehat{\psi}_B^2(u,t)\,\mathbb{I}(\overline{Z}(0,t) < \sqrt{u})\,\right]\\ &\quad + \mathbb{E}\left[\widehat{\psi}_B^2(u,t)\,\mathbb{I}(\overline{Z}(0,t) \ge \sqrt{u})\,\right].\end{aligned}$$

Since $\widehat{\psi}_B(u,t) \le 1$ a.s. the second term is smaller than $\mathbb{P}(\overline{Z}(0,t) \ge \sqrt{u})$ which is bounded above by $e^{-\gamma\sqrt{u}}$ for some $\gamma > 0$ by Lemma 2.8 on page 38. To get an upper bound for the first term it will be useful to rewrite $\widehat{\psi}_B^2$ as follows

$$\begin{aligned}\widehat{\psi}_B(u,t) = \mathbb{I}(N_t = 0, \overline{Z}(0,t) > u) &+ \mathbb{I}(N_t > 0, \tau(u) < T_J)\,N_t\,\overline{F}(U_{\max})\\ &+ \mathbb{I}(N_t > 0, \tau(u) \ge T_J)\,N_t\,\overline{F}\big(U_{\max} \vee (u - X(T_J-) - \overline{X}(T_J,t))\big),\end{aligned}$$

where $U_{\max} = \sup\{U_i : i \ne J\}$. The first indicator describes the event where the process $\{Y(s) : s < t\}$ has no jumps but the process $\{Z(s)\}$ reached u before time t. The second term corresponds to the case when ruin happens before time T_J, however we still require U_J to be larger than $U_{\max}$. The third term is the complement of the other two: ruin should happen after time T_J, so U_J is required both to be larger than $U_{\max}$ and to make the process reach the level u in the time interval $[T_J, t]$.

Since the events involved are disjoint, by taking the square of $\widehat{\psi}_B(u,t)$ we are left with the sum of the square of each term. Next we multiply by $\mathbb{I}(\overline{Z}(0,t) < \sqrt{u})$ and analyze each term separately. That is $\widehat{\psi}_2^B(u,t)\,\mathbb{I}(\overline{Z}(0,t) < \sqrt{u})$ is equivalent to

$$\mathbb{I}\Big(N_t = 0, \overline{Z}(0,t) > u,\, \overline{Z}(0,t) < \sqrt{u}\Big) \tag{2.11}$$

$$+\,\mathbb{I}\Big(N_t > 0, \tau(u) < T_J,\, \overline{Z}(0,t) < \sqrt{u}\Big)\, N_t^2\, \overline{F}^2(U_{\max}) \tag{2.12}$$

$$+\,\mathbb{I}\Big(N_t > 0, \tau(u) \geq T_J,\, \overline{Z}(0,t) < \sqrt{u}\Big) \times N_t^2 \overline{F}^2\Big(U_{\max} \vee (u - X(T_J-) - \overline{X}(T_J,t))\Big). \tag{2.13}$$

When $u > 1$ the events $\{\overline{Z}(0,t) < \sqrt{u}\}$, $\{\overline{Z}(0,t) > u\}$ are disjoint, so (2.11) is 0. Next, consider the case where $J = 1$ and the process $\{X(s)\}$ reaches level u before the time of the first jump, then $\overline{Z}(0,t) > u$ and the corresponding term (2.12) is 0 when $u > 1$. Therefore the indicator in (2.12) becomes

$$\begin{aligned}\mathbb{I}\big(N_t \geq 1, \tau(u) < T_J, \overline{Z}(0,t) < \sqrt{u}\big) &= \mathbb{I}\big(1 < J \leq N_t, \overline{X}(0,T_J) > u, \overline{Z}(0,t) < \sqrt{u}\big) \\ &\leq \mathbb{I}\big(1 < J \leq N_t, \overline{Y}(0,T_J) > u - \sqrt{u}\big) \\ &\leq \mathbb{I}\big(1 < J \leq N_t, S_{-J} > u - 2\sqrt{u}\big), \end{aligned} \tag{2.14}$$

where $S_{-J} = \sum_{i \neq J} U_i$. Now, if $J > 1$ and $S_{-J} > u - 2\sqrt{u}$ there exists at least one jump larger than $(u - \sqrt{u})/N_t$ and the following relation remains true

$$U_{\max} > \frac{u - 2\sqrt{u}}{N_t}.$$

Hence $\mathbb{I}\big(U_{\max} \geq (u - 2\sqrt{u})/N_t,\, N_t > 1\big)$ is an upper bound of (2.14) and it follows that (2.12) is smaller than

$$\mathbb{I}\big(1 < N_t, U_{\max} \geq (u - 2\sqrt{u})/N_t\big) N_t^2 \overline{F}^2(U_{\max}) \leq \mathbb{I}\big(1 < N_t\big) N_t^2 \overline{F}^2((u - 2\sqrt{u})/N_t).$$

We move to the term (2.13). Observe that if $\overline{Z}(0,t) < \sqrt{u}$ we have that

$$X(T_J-) + \overline{X}(T_J,t) \leq \overline{X}(0,T_J) + \overline{X}(T_J,t) < S_{-J} + 2\overline{Z}(0,t) < S_{-J} + 2\sqrt{u}.$$

Thus we obtain the following upper bound for (2.13):

$$\mathbb{I}(N_t \geq 1) N_t^2 \overline{F}^2\Big(U_{\max} \vee \Big(u - 2\sqrt{u} - S_{-J}\Big)\Big). \tag{2.15}$$

Now, if $U_{\max} < z/N_t$ for a fixed value $z > 0$ then it follows that

$$z - S_{-J} \geq z - (N_t - 1)U_{\max} > z/N_t.$$

This relation implies that

$$U_{\max} \vee \Big(u - 2\sqrt{u} - S_{-J}\Big) > \frac{u - 2\sqrt{u}}{N_t}.$$

So, the term (2.13) is bounded above by $\mathbb{I}(N_t \geq 1) N_t^2 \overline{F}^2\big((u - 2\sqrt{u})/N_t\big)$. Taking expectation and putting all the terms together we have obtained that

$$\mathbb{V}\mathrm{ar}\,\widehat{\psi}_2(u,t) \leq e^{-\gamma\sqrt{u}} + 2\,\mathbb{E}\left[N_t^2 \overline{F}^2\left(\frac{u - 2\sqrt{u}}{N_t}\right); N_t \geq 1\right], \qquad u > 1.$$

Divide the l.h.s. by $\overline{F}^2(u)$, take the limit as $u \to \infty$ and rewrite it as

$$\lim_{u\to\infty} \frac{\mathrm{e}^{-\gamma\sqrt{u}}}{\overline{F}^2(u)} + \frac{\overline{F}^2(u-2\sqrt{u})}{\overline{F}^2(u)} 2\,\mathbb{E}\left[N_t^2 \frac{\overline{F}^2\big((u-2\sqrt{u})/N_t\big)}{\overline{F}^2(u-2\sqrt{u})}; N_t \geq 1\right].$$

Since $\overline{F}$ is regularly varying the first limit is 0 and the second is the limit of

$$\lim_{u\to\infty} \mathbb{E}\left[N_t^2 \frac{\overline{F}(u/N_t)^2}{\overline{F}(u)^2}; N_t \geq 1\right].$$

We split this expectation into two parts corresponding to $1 \leq N_t \leq u$ or $N_t > u$. Recalling $\overline{F}(x) = \ell(x)/x^\alpha$, and since ℓ is slowly varying, $\ell(x)$ and $1/\ell(x)$ are both is $O(x^{\epsilon/4})$, so the first part is bounded by

$$\mathbb{E}\left[N_t^{2+2\alpha} \frac{\sup_{1\leq v\leq u} \ell(v)^2}{\ell(u)^2}; 1 \leq N_t \leq u\right] \;=\; O(u^\epsilon)\mathbb{E}\, N_t^{2+2\alpha} \;=\; O(u^\epsilon) O(t^{2+2\alpha}).$$

Using $\overline{F}(x) \leq 1$, the second part is bounded by

$$\mathbb{E}\left[N_t^2 \frac{u^{2\alpha}}{\ell(u)^2}; N_t > u\right] \;=\; \frac{u^{2\alpha}}{\ell(u)^2} O(\mathrm{e}^{-\delta u}) \;=\; O(\mathrm{e}^{-\delta u/2})$$

for some $\delta > 0$ where we used $t/u \to 0$ and easy tail estimates in the Poisson distribution. Putting these estimates together completes the proof. □

Lemma 2.8. *Let $t = t(u)$ such that $t/u \to 0$ and $\overline{Z}(0,t) = \sup\{Z(s) : s \leq t\}$. Then $\mathbb{P}(\overline{Z}(0,t) \geq u) \leq e^{-\gamma u}$ for some $\gamma > 0$.*

Proof. Let $Z_1(t) = Z(t) - bt$ where u is so large that $\mathbb{E}\, Z_1(1) < 0$. Then Lundberg's inequality (Asmussen, 2000) gives that $\mathbb{P}(\overline{Z}_1(0,\infty) \geq u) \leq e^{-\gamma_1 u}$ for some $\gamma_1 > 0$ (note that all exponential moments of $Z_1(1)$ exist). It follows that

$$\mathbb{P}(\overline{Z}(0,t) \geq u) \;\leq\; \mathbb{P}(\overline{Z}_1(0,\infty) \geq u - bt) \;\leq\; e^{-\gamma_1(u-bt)}.$$

From this the result follows, since $t/u \to 0$. □

Proof of Theorem 2.7. The idea of the proof will provide new bounds for expressions (2.12) and (2.13) in the proof of Theorem 2.6. Recall that the indicator function in (2.12) was bounded by (2.14). Using this we can get the next bound

$$\mathbb{I}\big(S_{-J} > u - 2\sqrt{u}\big) N_t^2 \overline{F}^2(U_{\max}) < \mathbb{I}\Big(S_{-J} > \big(u-2\sqrt{u}\big)^\eta\Big) N_t^2 \overline{F}^2(U_{\max})$$

for all $u > 3 + 2\sqrt{2}$ and any value of $0 < \eta < 1$. We move to the term (2.13). Observe that for any $0 < \eta < 1$ and $u > 3 + 2\sqrt{2}$ the expression (2.13) – which is an upper bound for (2.13), can be bounded above by

$$\begin{aligned}
&\mathbb{I}\Big(\, S_{-J} \leq \big(u-2\sqrt{u}\big)^\eta\Big) N_t^2 \overline{F}^2\big(u - 2\sqrt{u} - S_{-J}\big) + \mathbb{I}\Big(S_{-J} > \big(u-2\sqrt{u}\big)^\eta\Big) N_t^2 \overline{F}^2(U_{\max}) \\
&\qquad \leq\; N_t^2 \overline{F}^2\big(u - 2\sqrt{u} - (u-2\sqrt{u})^\eta\big) + \mathbb{I}\Big(S_{-J} > \big(u-2\sqrt{u}\big)^\eta\Big) N_t^2 \overline{F}^2(U_{\max}).
\end{aligned}$$

Taking expectation and putting all the terms together we have obtained that

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\, \widehat{\psi}_2(u,t) \leq\; & e^{-\gamma\sqrt{u}} + \mathbb{E}\,[N_t^2]\; \overline{F}^2\big(u - 2\sqrt{u} - (u-2\sqrt{u})^\eta\big) \\
& + 2\,\mathbb{E}\left[N_t^2 \overline{F}^2\big(U_{\max}\big); S_{-J} > (u-2\sqrt{u})^\eta\right]
\end{aligned}$$

when $u > 3 + 2\sqrt{2}$. Divide the l.h.s. by $\lambda^2 u^{2\beta}\overline{F}^2(u)$, take the limit as $u \to \infty$ and rewrite it as

$$\lim_{u\to\infty} \frac{\mathrm{e}^{-\gamma\sqrt{u}}}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)} + \frac{\mathbb{E}\,[N_t^2]\overline{F}^2(u - 2\sqrt{u} - (u - 2\sqrt{u})^\eta)}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)} \tag{2.16}$$

$$+ \frac{2\,\mathbb{E}\left[N_t^2\overline{F}^2\big(U_{\max}\big); S_{-J} > (u - 2\sqrt{u})^\eta\right]}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)}. \tag{2.17}$$

Since $\overline{F}$ is regularly varying the first limit is 0 and the second is 1 (the later also follows since $N_t \sim \text{Poisson}(\lambda u^\beta)$). We claim that the last limit is 0 under the hypothesis of the Theorem. For proving so we need Lemma 2.9 on the next page and the following technical condition. Choose $k \in \mathbb{N}$ and δ_k in such way that

$$\frac{1}{k+2}\Big(k\frac{\beta}{\alpha} + 2\Big) < \delta_k < 1 - \beta.$$

It is not difficult to verify that the last is equivalent to have $\beta < \left(\frac{k+2}{k} + \frac{1}{\alpha}\right)^{-1}$. Hence, the existence of δ_k follows from the hypothesis of the Theorem $0 < \beta < (1 + 1/\alpha)^{-1}$ by choosing k large enough (observe that $(\frac{k+2}{k} + \frac{1}{\alpha})^{-1} \nearrow (1 + 1/\alpha)^{-1}$).

Choose δ_1 according to Lemma 2.9 on the following page and such that $\delta_k < \delta_1 < 1$. We bound the last term in (2.17) with

$$\lim_{u\to\infty} \frac{2\,\mathbb{E}\,[N_t^2; N_t \le k]}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)} + \frac{2\,\mathbb{E}\left[N_t^2\overline{F}^2\big(U_{\max}\big); N_t > k, S_{-J} > (u - 2\sqrt{u})^\eta\right]}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)}.$$

Observe that $\mathbb{E}\,[N_t^2; N_t \le k] < k^2\mathbb{P}(N_t \le k)$ goes to 0 at an exponential rate and therefore the first term goes to 0. For the second we can break it in several pieces and bound it as follows

$$\begin{aligned}
\lim_{u\to\infty} & \frac{2\,\mathbb{E}\left[N_t^2\overline{F}^2\big(U_{\max}\big); N_t > k, U_{\max} \ge u^{\delta_1}\right]}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)} \\
& + \frac{2\,\mathbb{E}\left[N_t^2\overline{F}^2\big(U_{\max}\big); N_t > k, U_{(N_t-k+1)} \ge u^{\delta_k}\right]}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)} \\
& + \lim_{u\to\infty} \frac{2\,\mathbb{E}\left[N_t^2; N_t > k, S_{-J} > (u - 2\sqrt{u})^\eta, U_{(N_t-k+1)} < u^{\delta_k}, U_{\max} < u^{\delta_1}\right]}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)}.
\end{aligned} \tag{2.18}$$

By Lemma 2.9 on the next page and the choices of k and δ_k the first two terms are 0. For the third one observe that if $U_{(N_t-k+1)} < u^{\delta_k}$ and $U_{\max} < u^{\delta_1}$ then $S_{-J} < N_t u^{\delta_k} + k u^{\delta_1}$ and therefore

$$\mathbb{I}\Big(S_{-J} > (u - 2\sqrt{u})^\eta, U_{(N_t-k+1)} < u^{\delta_k}, U_{\max} < u^{\delta_1}\Big) < \mathbb{I}\Big(\frac{(u - 2\sqrt{u})^\eta - k u^{\delta_1}}{u^{\delta_k}} < N_t\Big)$$

Remember that the result is valid for any value of $\eta < 1$ and we have chosen $\delta_k < 1 - \beta$. Hence, we choose $\eta > \delta_1$ and close enough to 1 such that $\eta - \delta_k > \beta$. Then we use the

last argument to bound (2.18) with

$$\lim_{u\to\infty} \frac{2\,\mathbb{E}\left[N_t^2; N_t > u^{\eta-\delta_k}\right]}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)} \leq \lim_{u\to\infty} \frac{2\sqrt{\mathbb{E}\left[N_t^4\right]\mathbb{P}(N_t > u^{\eta-\delta_k})}}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)} = 0.$$

In the first inequality we have used Cauchy inequality and in the second the fact that $\mathbb{E}\left[N_t^4\right] \sim (\lambda t)^4$ and that $\mathbb{P}(N_t > u^{\eta-\delta_k})$ goes to 0 at an exponential rate since $\eta - \delta_k > \beta$. □

Lemma 2.9. *Let $k \in \mathbb{N}$. If $0 < \delta_k < 1$ is such that $\frac{1}{k+2}\left(k\frac{\beta}{\alpha} + 2\right) < \delta_k$, then*

$$\lim_{u\to\infty} \frac{\mathbb{E}\left[N_t^2 \overline{F}^2\big(U_{\max}\big); N_t > k, U_{(N_t-k+1)} \geq u^{\delta_k}\right]}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)} = 0,$$

where $U_{(k)}$ is the k-th order statistic from the set $\{X_k : k \neq J\}$.

Proof. Observe that since $\beta/\alpha < 1$ it is always possible to choose $0 < \delta_k < 1$ as in the hypothesis. Hence, we bound the given expression with

$$\begin{aligned}
&\lim_{u\to\infty} \frac{\overline{F}^2\big(u^{\delta_k}\big)\mathbb{E}\left[N_t^2; N_t > k, U_{(N_t-k+1)} \geq u^{\delta_k}\right]}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)} \\
&\quad = \lim_{u\to\infty} \frac{\overline{F}^2\big(u^{\delta_k}\big)\mathbb{E}\left[N_t^2\,\mathbb{P}(U_{(N_t-k+1)} \geq u^{\delta_k} | N_t > k)\right]}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)} \\
&\quad \leq \lim_{u\to\infty} \frac{\overline{F}^{2+k}\big(u^{\delta_k}\big)\mathbb{E}\left[N_t^{2+k}\right]}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)} \\
&\quad = \lim_{u\to\infty} \frac{\overline{F}^{2+k}\big(u^{\delta_k}\big)(\lambda u^{\beta})^{2+k}}{\lambda^2 u^{2\beta}\,\overline{F}^2(u)}.
\end{aligned}$$

We used that $\mathbb{P}(U_{(N-k+1)} > u) \leq N^k \overline{F}^k(u)$ and $\mathbb{E}\left[N_t^k\right] \sim (\lambda t)^k$ for $k \in \mathbb{N}$. It can be easily proved that we can rewrite $\overline{F}^{2+k}(u^{\delta_k}) = \ell_1(u)/u^{\delta\alpha(2+k)}$ and $\overline{F}^2(u) = \ell_2(u)/u^{2\alpha}$ with $\ell_1(u)$ and $\ell_2(u)$ slowly varying. Since $\ell_1(u)$ and $1/\ell_2(u)$ are both $O(u^{\epsilon/2})$ for $\epsilon > 0$ the last expression is equivalent to

$$\lambda^k \lim_{u\to\infty} \frac{\ell_1(u)}{\ell_2(u)} u^{2\alpha+k\beta-(2+k)\alpha\delta_k} < \lambda^k \lim_{u\to\infty} u^{2\alpha+k\beta-(2+k)\alpha\delta_k+\epsilon}.$$

By the election of δ_k in the hypothesis of the Lemma it is straightforward to verify that $2\alpha + k\beta - (2+k)\alpha\delta_k < 0$ and therefore it is possible to choose $\epsilon > 0$ small enough such that the last limit is 0. □

Chapter 3

Tail Probabilities for the Multivariate Lognormal

We consider the problem of approximating tail probabilities of sums of random variables which are nonindependent and possess heavy-tails. As a motivating example, one could consider the problem of computing the probability of large losses or high returns on a portfolio of correlated asset prices. A very popular model in the financial literature is the so-called Black-Scholes model in which stock prices follow lognormal distributions which are usually considered to have significative correlations. Motivated by these types of financial risk problems, we shall concentrate on the approximation of tail probabilities of sums of dependent lognormals. The tail probability of sums of lognormals also appears in other disciplines as physics, electronics and telecommunications (cf. Szyszkowicz and Yanikomerloglu, 2007; Rossberg, 2008; Vanduffel et al., 2008, and references therein).

More precisely, we are considering a random vector such that its logarithmic transformation $\log \mathbf{X}$ follows a multivariate normal distribution. Here we are interested in approximating the probability $z(u) = \mathbb{P}\big(S_d > u\big)$ as $u \to \infty$ which will be done via an asymptotic equivalent expression and complemented by Monte Carlo estimators.

Note that most of the literature has focused on sums of independent random variables. In contrast, we consider a problem that involves the sum of dependent increments which makes the available results for estimating the tail probabilities of sums of heavy-tailed increments difficult to apply in our current setting because they rely heavily on the i.i.d. assumption.

In the Section 3.1 we present an asymptotic approximation for the tail probability $\mathbb{P}(S_d > u)$ which was originally given in Asmussen and Rojas-Nandayapa (2008). In particular, this result shows that the asymptotic behavior of the sum of correlated lognormal random variables is similar to that of the independent case which is characteristic of subexponential distributions. We further analyze those events where the sum becomes large under the guidelines of classic subexponential theory. The intuitive insight gained from it will serve as the key idea for the development of Monte Carlo estimators.

In Section 3.2 we analyze the Monte Carlo estimators for $\mathbb{P}(S_n > u)$ proposed in Asmussen et al. (2008) and Asmussen and Rojas-Nandayapa (2006). The main idea of the first set of estimators is Importance Sampling where the change of measure involves scaling the covariance matrix. Furthermore, Cross-Entropy and Stratification strategies are used to improve the efficiency of the estimator which can be proved to have *vanishing relative error* as shown in Asmussen et al. (2008). In the second part

of Section 3.2, we study a couple of Monte Carlo estimators where the key idea is Conditional Monte Carlo. This estimators are multivariate extensions of the estimators proposed by Asmussen and Binswanger (1997); Asmussen and Kroese (2006) and were empirically studied in Asmussen and Rojas-Nandayapa (2006). Here we extend some of these results and suggest some strategies for improving the efficiency of these estimators.

We close this chapter with a section of numerical comparisons of both the approximation and Monte Carlo estimates.

3.1 Asymptotic Tail Probabilities for Sums of Lognormals

In order to state our results we must introduce some notation. Let $\mathbf{X} = (X_1, X_2, \ldots, X_d)^T$ be a d-dimensional vector such that $\log \mathbf{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and set $S_d = X_1 + \cdots + X_d$. We shall write $\sigma_i^2 = \boldsymbol{\Sigma}_{i,i}$, $\sigma_{i,j} = \boldsymbol{\Sigma}_{i,j}$ for $i \neq j$ and $\rho_{i,j} = \sigma_{i,j}/(\sigma_i \sigma_j)$; these three notions correspond to the variance of the i-th Gaussian component and the covariance and correlation between the i-th and j-th components respectively. We define

$$\sigma^2 = \max_{1 \leq k \leq d} \sigma_k^2, \quad \mu = \max_{k:\, \sigma_k^2 = \sigma^2} \mu_k, \quad m_d := \#\big\{k :\, \sigma_k^2 = \sigma^2,\, \mu_k = \mu\big\}.$$

The parameters σ^2 and μ allow to characterize the dominant tail behavior among the X_j's. In order to see this, let us recall the following well known asymptotic relation often referred to as Mill's ratio (cf. Resnick, 1992; Mikosch, 2003); if $Y_i \sim N(\mu_i, \sigma_i^2)$ then as $y \to \infty$

$$\mathbb{P}(Y_i > y) = \frac{\sigma_i}{(2\pi)^{1/2}(y - \mu_i)} \exp\left\{ - \frac{(y - \mu_i)^2}{2\sigma_i^2} \right\}(1 + o(1)). \tag{3.1}$$

With a simple exponential transformation we can obtain the equivalent asymptotic expression for a lognormal random variable. Note that, in particular, the approximation (3.1) indicates that $\mathbb{P}(X_j > u) = o\big(\mathbb{P}(X_i > u)\big)$ if $\sigma_i^2 > \sigma_j^2$, or $\sigma_i^2 = \sigma_j^2$ and $\mu_i > \mu_j$. We shall introduce the following assumption:

Assumption A: Suppose that $\rho_{k\ell} < 1$ whenever $\sigma_k^2 = \sigma_\ell^2$.

The assumption A is equivalent to say that the correlation between any two lognormals is strictly less than 1 which corresponds to a degenerate case. Otherwise, we would have $X_k = aX_\ell$. If the problem at hand involves S_d we can still fulfill Assumption A by taking

$$X_k + X_\ell = (1 + a)X_k.$$

In the independent case, it is provable that $\mathbb{P}(M_d > u) \sim m_d\, \mathbb{P}(\mu + \sigma Z > \log u)$ as $u \to \infty$ where Z stands for a normal standard random variable. Therefore, we obtain that

$$\mathbb{P}(X_k > u \big| M_d > u) \longrightarrow \begin{cases} 1/m_d & \sigma_k^2 = \sigma^2, \mu_j = \mu \\ 0 & \text{otherwise} \end{cases} \qquad u \to \infty.$$

saying that the probability that a specific random variable is large given that the maximum is large is significative only if it has the dominant tail. The same holds true for correlated normal random variables:

Proposition 3.1. *Let* $\mathbf{Y} = (Y_1, \ldots, Y_d) \sim N\big(\boldsymbol{\mu}, \boldsymbol{\Sigma}\big)$, $X_j = \exp\big(Y_j\big)$ *and put* $M_d = \max\{X_k : 1 \leq k \leq d\}$. *Then*

$$\lim_{u\to\infty} \frac{\mathbb{P}(M_d > u)}{\mathbb{P}(\mu + \sigma Z > \log u)} = m_d.$$

The previous asymptotic result is a consequence of a basic feature of Gaussian distributions, namely the *tail independence* and its proof can be found on page 53. The following is the main result in Asmussen and Rojas-Nandayapa (2008).

Theorem 3.2 (Asmussen and Rojas-Nandayapa (2008))**.** *Suppose* $\gamma(u) \to \gamma^* \in (0, \infty)$ *as* $u \to \infty$. *Define* $\mathbf{Y}(u) = \big(Y_1(u), \ldots, Y_d(u)\big) \sim N(\boldsymbol{\mu}, \gamma(u)\boldsymbol{\Sigma})$, $X_j(u) = \exp\big(Y_j(u)\big)$ *and put* $S_d(u) = X_1(u) + \cdots + X_d(u)$. *Then, if Assumption A on the preceding page holds, we have*

$$\lim_{u\to\infty} \frac{\mathbb{P}(S_d(u) > u)}{\mathbb{P}(\mu + \sigma\gamma(u) Z > \log u)} = m_d,$$

The previous result in the case in which $\gamma(u) = 1$ is proved in Asmussen and Rojas-Nandayapa (2008). The extension to the situation $\gamma(u) \to \gamma^* \in (0, \infty)$, which is required in our future development, follows exactly as in that paper and its proof is given on page 53. In Geluk and Tang (2008) it is claimed that this result is a consequence of their Theorem 3.2 . However, the hypothesis of that Theorem are not satisfied in the case of correlated lognormals.

Asymptotically speaking, Theorem 3.2, says that tail probability of the sum is exclusively determined only by the tail probabilities of the random variables with the heavier marginal tails. In fact, this is an extension of the subexponential property for sums of i.i.d. lognormal random variables which states that

$$\mathbb{P}(S_d > u) = \sum_{j=1}^{d} \mathbb{P}(X_j > u)\big(1 + o(1)\big) = d\,\mathbb{P}(X_j > u)\big(1 + o(1)\big), \qquad u \to \infty.$$

Indeed, it follows from (3.1) that if $\mathbf{Y}(u) \sim N(\boldsymbol{\mu}, \gamma(u)\boldsymbol{\Sigma})$, then

$$\sum_{j=1}^{d} \mathbb{P}\big(S_j(u) > u\big) = m_d \mathbb{P}\big(\mu + \sigma\gamma(u) Z > \log u\big)\big(1 + o(1)\big).$$

The following Corollary follows trivially from Proposition 3.1 and Theorem 3.2.

Corollary 3.3. *Let* $\mathbf{Y} \sim N\big(\boldsymbol{\mu}, \boldsymbol{\Sigma}\big)$, $X_j = \exp(Y_j)$ *and put* $M_d = \max\{X_k : 1 \leq k \leq d\}$, *then*

$$\lim_{u\to\infty} \frac{\mathbb{P}(M_d > u)}{\mathbb{P}(S_d > u)} = 1.$$

A direct consequence of this corollary is that $\mathbb{P}(M_d > u | S_d > u) \to 1$. In turn, we can intuitively interpret just as in the independent case, namely, that the sum of correlated lognormal is large due to the contribution of a single large increment.

A common intuition is that the result in Theorem 3.2 is a consequence exclusively of the tail independence of the Gaussian distribution, which in fact is not true. The proof exploits properties inherent to the Gaussian distribution which cannot be deduced from the copula exclusively. Moreover, Albrecher et al. (2006) present an example of a bivariate and tail independent distribution with lognormal marginals which fails to have this asymptotic behavior.

3.2 Monte Carlo Estimation

First we will analyze two importance sampling estimators for $\mathbb{P}(S_n > u)$ proposed in Asmussen et al. (2008). The first estimator is closely related to the use of an appropriate exponential change-of-measure, *not* directly the one for the underlying Gaussian distributions, but to the scaling of the covariance matrix by a factor that grows at a suitable slow speed as $u \to \infty$. Since the sampler involves a simple scaling, the estimator is straightforward to implement and it can be shown to be asymptotically optimal as $u \to \infty$. Furthermore, Cross-Entropy methods are then used for finding the best tuning for the importance distribution. The second improvement takes advantage of the fact that the largest of the increments dominates the large deviations behavior of the sums of correlated lognormals. The strategy is to decompose the tail event of interest in two contributions, a dominant piece corresponding to the tail of the maximum and a remaining contribution. The dominant contribution is analyzed by means of an estimator with *vanishing relative error* for the maximum of multivariate Gaussians and the remaining contribution is independently handled using the importance sampling strategy utilized in the design of the first estimator. We show that under mild conditions, this estimator actually possesses *asymptotically vanishing relative error*.

In the second part of this section we analyze the two estimators which were empirically studied in Asmussen and Rojas-Nandayapa (2006). The key idea is conditional Monte Carlo based on order statistics. These estimators are multivariate extensions of the estimators proposed in Asmussen and Binswanger (1997) and Asmussen and Kroese (2006) respectively. The main feature of these algorithms is that they are applicable to a wider class of multivariate distributions. The only additional requirement for its implementation is the knowledge of the conditional distribution of every single component given the rest of the vector. The second algorithm fails to be efficient in its original form, however we conjecture that this algorithm attains efficiency with a combination of importance sampling involving an appropriate variance scaling.

3.2.1 Importance Sampling via Variance Scaling

The notion of tail behavior is commonly associated to that of variance, which typically are two non-related characteristics of a distribution. However, in the case of Gaussian random variables we can see that the variance, in fact, controls the tail behavior as discussed in 3.1 by means of the Mill's ratio. Using this principle a natural importance sampling strategy to consider for computing $\mathbb{P}(S_n > u)$ is one that induces high variances. This motivates considering as importance sampler a distribution such as

$$N\big(\boldsymbol{\mu}, \boldsymbol{\Sigma}/(1-\theta)\big), \qquad 0 < \theta < 1.$$

In other words, we just inflate the covariance matrix by the factor $1/(1-\theta)$. We denote the probability measure induced by the Importance Sampling as $\mathbb{P}_\theta(\cdot)$ and we use the notation $\mathbb{E}_\theta(\cdot)$ for the associated expectation operator. The estimator induced by this simple strategy is

$$\widehat{z}_1(u) = \mathbb{I}(S_d > u)\frac{\exp\big\{-\theta(\mathbf{Y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})/2\big\}}{(1-\theta)^{d/2}}.$$

The next Lemma summarizes a useful representation for the second moment of $\widehat{z}_1(u)$ under the importance sampling distribution and its proof can be found on page 55.

However, in order to state such representation we introduce another family of probability measures (in addition to the $\mathbb{P}_\theta$'s), which we shall denote by $(Q_\theta : 0 \le \theta \le 1)$. We use $Q_\theta(\cdot)$ to denote a probability measure under which $\mathbf{Y}$ is $N\big(\boldsymbol{\mu}, \boldsymbol{\Sigma}/(1+\theta)\big)$.

Lemma 3.4.

$$\mathbb{E}_\theta \widehat{z}_1(u)^2 = (1-\theta^2)^{-d/2} Q_\theta(S_d > u). \tag{3.2}$$

As an immediate consequence of the previous result we obtain that the estimator $\widehat{z}_1(\theta)$ is logarithmic efficient if one chooses $\theta(u) \to 1$ at an appropriate speed. We wish to select θ close to unity because under $Q_\theta(\cdot)$ the variances are multiplied by the factor $1/(1+\theta)$ and, to obtain logarithmic efficiency, we need to match the rate of decay of $z(u)^2$ which is determined by the factor one half times the largest variance parameter. The proof of the following Theorem is given on page 56.

Theorem 3.5. *Suppose that $\psi(u) := 1 - \theta(u) = o(1)$, then for $\epsilon \ge 0$*

$$\frac{\mathbb{E}_{\theta(u)} \widehat{z}_1(u)^2}{z(u)^{2-\epsilon}} = \Theta\bigg((\log u)^{1-\epsilon}\, \psi(u)^{-d/2}\, \exp\bigg\{ - \frac{(\epsilon - \psi(u))\big(\log u - \mu\big)^2}{2\sigma^2} \bigg\} \bigg). \tag{3.3}$$

In particular, if $1/\psi(u) = o\big(e^{p(\log u)^2}\big)$ for some $p > 0$, then $\widehat{z}_1(u)$ is logarithmically efficient.

One can choose $\theta(u)$ in many ways which are consistent with the condition that $(1-\theta(u))^{-1} = o\big(e^{p(\log u)^2}\big)$ for some $p > 0$ as $u \to \infty$. One of them involves finding $\theta(u)$ that minimizes the asymptotic expressions for the second moment of the estimator given by (3.4). A simpler approach is to find the unique positive root $\theta(u)$ (which exists for u large enough) to the equation $\mathbb{E}_{\theta(u)} S_d = u$. This root-finding procedure does not contribute significantly to the computational cost of the algorithm because it is done just once. The next lemma, whose proof can be found on page 56 shows that using the root-finding procedure we obtain $1 - \theta(u) = \Theta\big((\log u)^{-1}\big)$ as $u \to \infty$.

Lemma 3.6. *The function $\theta(u)$ given as the unique root of the equation*

$$\mathbb{E}_{\theta(u)} S_d = \mathrm{e}^{\mu_1 + \sigma_1^2/2(1-\theta(u))} + \cdots + \mathrm{e}^{\mu_d + \sigma_d^2/2(1-\theta(u))} = u$$

is such that

$$1 - \theta(b) = \frac{\sigma^2}{2} + o(1).$$

Algorithmically this estimator comes out as.

Algorithm: Importance Sampling via Variance Scaling

1. Find $\theta := \theta(u)$ which as the root of the equation

$$\mathbb{E}_{\theta(u)} S_d = \mathrm{e}^{\mu_1 + \sigma_1^2/(2(1-\theta(u))^2)} + \cdots + \mathrm{e}^{\mu_d + \sigma_d^2/(2(1-\theta(u))^2)} = u.$$

2. Sample $\mathbf{Y} \sim N\big(\boldsymbol{\mu}, \boldsymbol{\Sigma}/(1-\theta)\big)$.

3. Return

$$\widehat{z}_1(u) = \mathbb{I}(S_d > u) \frac{\exp\big(- \theta\, (\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})/2 \big)}{(1-\theta)^{d/2}}.$$

As a corollary to the analysis in Theorem 3.5 and Lemma 3.6 we obtain the following result whose proof can be found on this page.

Corollary 3.7. *The estimator $Z_1(u)$ given by Algorithm 1 satisfies*

$$\frac{Var_\theta \widehat{z}_1(u)}{z(u)^2} = O(u^{1/4} \log u^{d/2+1}).$$

Although the estimator $\widehat{z}_1(u)$ possesses two very convenient features, namely, is very easy to implement and is asymptotically optimal, it also has the disadvantage that the premultiplying factor in the asymptotic variance expression (3.5) might grow substantially involving a factor such as $O(u^{1/4} \log u^{d/2+1})$. So, for moderate values of u and d, the variance performance of the estimator might degrade in a significant way. To cope with this problem one can introduce additional variance reduction techniques, such as stratified sampling or conditional Monte Carlo. Another alternative that takes advantage of the intuitive interpretation given by Corollary 3.3 and that achieves bounded relative error will be studied later, but first, we shall provide another interpretation of the change-of-measure behind $\widehat{z}_1(u)$ using Cross-Entropy ideas.

3.2.1.1 Cross-Entropy Implementation

The Cross-Entropy can be used to provide an answer on how to select $\theta(u)$ within the class of importance sampling distributions given by $\mathbb{P}_{\theta(u)}$ (cf. Rubinstein and Kroese, 2004). The Cross-Entropy method is an iterative procedure which, in principle, improves the estimator in every step. In this section we shall explore an implementation of Cross-Entropy that starts with a choice of $\theta(u)$, based on the solution to the equation $E_{\theta(u)} S_d = u$, that, as we saw previously, can be shown to be asymptotically optimal. Consequently, the application of the Cross-Entropy method is intuitively expected to improve the variance performance of the corresponding estimator.

For our first algorithm in the previous section, we considered $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ fixed and we draw samples from

$$N\big(\boldsymbol{\mu}, \boldsymbol{\Sigma}/(1-\theta)\big).$$

Here we will use instead a larger, but still simple family of parametric multivariate distributions. Our proposal is to take $\boldsymbol{\Sigma}$ fixed and consider

$$N(\widetilde{\boldsymbol{\mu}},\, \boldsymbol{\Sigma}/(1-\theta)) \qquad \widetilde{\boldsymbol{\mu}} \in \mathbb{R}^d \quad \theta \in \mathbb{R}^+.$$

Here we provide directly the expression for the parameters omitting the details of the calculation. For more details on the Cross-Entropy method we refer to Rubinstein and Kroese (2004).

The parameters for the k-th iteration of the Cross-Entropy method are described as follows. First, we sample r i.i.d. r.v.'s $\big(\mathbf{Y}_{i,k} : 1 \le i \le r\big)$ such that

$$\mathbf{Y}_{i,k} \sim N(\widetilde{\boldsymbol{\mu}}_{k-1}, \boldsymbol{\Sigma}/(1-\theta_{k-1})).$$

Given $\mathbf{Y}_{i,k} = \mathbf{y}_{i,k}$ we compute

$$\widetilde{\boldsymbol{\mu}}_k := \frac{\sum_{i=1}^{r} w_{i,k}\, \mathbf{y}_{i,k}}{\sum_{i=1}^{r} \mathbf{y}_{i,k}} \qquad \frac{1}{1-\theta_k} := \frac{\sum_{i=1}^{r} w_{i,k}\, (\mathbf{y}_{i,k} - \widetilde{\boldsymbol{\mu}}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_{i,k} - \widetilde{\boldsymbol{\mu}}_k)}{r \sum_{i=1}^{r} w_{i,k}}, \tag{3.4}$$

where the weights $w_{i,k}$ are given by

$$w_{i,k} := (1-\theta_k)^{d/2} \frac{\exp\big(-(\mathbf{y}_{i,k}-\widetilde{\boldsymbol{\mu}}_k)^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}_{i,k}-\widetilde{\boldsymbol{\mu}}_k)\big)}{\exp\big(-(\mathbf{y}_{i,k}-\widetilde{\boldsymbol{\mu}}_{k-1})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}_{i,k}-\widetilde{\boldsymbol{\mu}}_{k-1})\big)} \mathbb{I}(S_{d,i} > u).$$

It is an easy calculus exercise to verify that this expressions satisfy the conditions of the Cross-Entropy method. One could try to choose a larger family of importance sampling distributions to provide better estimates, however, the expressions can quickly become complicated and more difficult to implement. We performed numerical experiments and noted that the algorithm converges in a few iterations suggesting that our initial distribution is not that far from the optimal distribution within the new family. The precise description of the algorithm is given below.

Algorithm: Cross-Entropy via Variance Scaling.

1. Let $k = 1$ and $\widetilde{\boldsymbol{\mu}}_0 := \boldsymbol{\mu}$. Define $\theta_0 := \theta(u)$ as the solution of

$$\mathrm{e}^{\mu_1+\sigma_1^2\theta(u)/2} + \cdots + \mathrm{e}^{\mu_d+\sigma_d^2\theta(u)/2} = u.$$

2. Simulate a sequence of random vectors r i.i.d. r.v.'s $\big(\mathbf{Y}_{i,k} : 1 \le i \le r\big)$ with common distribution $N_d(\widetilde{\boldsymbol{\mu}}_k, \boldsymbol{\Sigma}/(1-\theta_k))$ and calculate $\widetilde{\boldsymbol{\mu}}_{k+1}$ and θ_{k+1} as given in (3.4). If the new parameters satisfy a convergence criteria go to 3 (see our comments below for a convergence criteria that we used in our numerical examples). Else make $k := k+1$ and repeat 2.

3. Return

$$\widehat{z}_1(u) := (1-\theta_k)^{d/2} \frac{\exp\big(-(\mathbf{Y}_{i,k}-\widetilde{\boldsymbol{\mu}}_k)^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_{i,k}-\widetilde{\boldsymbol{\mu}}_k)\big)}{\exp\big(-(\mathbf{Y}_{i,k}-\widetilde{\boldsymbol{\mu}}_{k-1})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_{i,k}-\widetilde{\boldsymbol{\mu}}_{k-1})\big)} \mathbb{I}(S_{d,i} > u).$$

We might choose several criteria in Step 2 above. However, since we are interested in the relative error we will stop iterating when the absolute difference between the empirical coefficient of variation between the $w_{k,i}$'s (for $1 \le i \le r$) and that of the $w_{k-1,i}$'s is smaller than $\alpha \cdot 100\%$ the empirical coefficient of variation of the $w_{k-1,i}$'s.

3.2.1.2 Stratification Strategy

In our intuitive discussion leading to Corollary (3.3) we observed that large values of S_d happen due to the contribution of a single large jump. On the other hand, in the previous section, we constructed a weakly efficient estimator using an importance sampler based on the fact that the variances dictate the tail behavior of S_d. The idea in this section is to combine these two intuitive observations in order to produce a strongly efficient importance sampling estimator. First, note that

$$\mathbb{P}(S_n > u) = z_M\big(u\big) + z_R\big(u\big),$$

where

$$z_M(u) = \mathbb{P}\big(\max_{1\le i\le d} X_i > u\big), \qquad z_R(u) = \mathbb{P}\big(S_d > u, \max_{1\le i\le d} X_i \le u\big).$$

In view of Theorem 3.2 on page 43 we must have that $z_R(u) = o\big(z_M(u)\big)$ as $u \to \infty$, so the most important contribution comes from the term $z_M\big(u\big)$. We shall refer to $z_R(u)$ as the *residual probability* .

The strategy is to design independent and unbiased estimators, say $\widehat{z}_R(u)$ and $\widehat{z}_M(u)$. This idea has been exploited previously in the literature in the context of i.i.d. increment distributions (cf. Juneja, 2007). The gain comes if $\widehat{z}_M(u)$ is strongly efficient for $z_M\big(u\big)$ even if $\widehat{z}_R(u)$ has a coefficient of variation of order $O\big(z(u)/z_R(u)\big)$ as $u \to \infty$. In other words, $\widehat{z}_R(u)$ may not be strongly efficient for $z_R(u)$, but its coefficient of variation could grow slowly enough so that the combined estimator $\widehat{z}_2(u) = \widehat{z}_M(u) + \widehat{z}_R(u)$ for $z(u)$ is strongly efficient.

For $z_R(u)$ we propose to use $\mathbb{P}_\theta$ as our importance sampling distribution, i.e. $\mathbf{Y} \sim N\big(\boldsymbol{\mu}, , \boldsymbol{\Sigma}/(1-\theta)\big)$. The corresponding estimator takes the form

$$\widehat{z}_R(u) = \mathbb{I}\Big(S_d > u, \max_{1\le i\le d} X_i \le u\Big) \frac{\exp\big(-\theta(u)(\mathbf{Y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})/2\big)}{(1-\theta(u))^{d/2}}.$$

The reason for using $\mathbb{P}_\theta$ as importance sampler is that for estimating $\alpha_2\big(u\big)$ one must induce the underlying rare event $\{S_d > u, \max_{1\le i\le d} X_i \le u\}$ by means of more than one large component (which might be achieved by inflating the variances) as opposed to inducing a single large jump as suggested by Theorem 3.2 on page 43. Just as before we conclude that

$$E_{\theta(u)}\widehat{z}_R(u)^2 = \big(1-\theta(u)^2\big)^{-d/2} Q_{\theta(u)}\big(S_d > u, \max_{1\le i\le d} X_i \le u\big). \tag{3.5}$$

The following result, whose proof can be found on page ?? provides the necessarily elements to analyze $E_{\theta(u)}Z_{2,2}(u)^2$.

Theorem 3.8. *Suppose that Assumption A is in force and that* $\psi(u) := 1-\theta(u) = o(1)$. *Then,*

$$\frac{\mathbb{E}\,\widehat{z}_R^2(u)}{z^2(u)} = O\bigg(\frac{\psi(u)^{-d/2}(\log u)^2}{u^{1-\beta}}\bigg), \qquad u \to \infty.$$

In particular, if $\psi(u)^{-d/2} = o\big(u^{1-\beta}/\log(u)^2\big)$, *and we use an unbiased estimator* $\widehat{z}_M$ *for* $z_M(u)$ *then*

$$\limsup_{u\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{z}_2(u)}{z(u)^2} = \limsup_{u\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{z}_R(u)}{z(u)^2}.$$

Hence, the estimator $\widehat{z}_2(u)$ will inherit the efficiency properties of the estimator $\widehat{z}_M(u)$ that we decide to choose. That is, an unbiased estimator for

$$z_M(u) = \mathbb{P}\Big(\max_{1\le j\le d} Y_j > \log u\Big).$$

Notice that, what we have now is an estimation problem in a setting involving light tails. A possibility is to use the algorithm proposed in Adler et al. (2008) for estimating the probability $\mathbb{P}(M_n > u)$ which in fact has *vanishing relative error*. Part of future work will be to provide an alternative algorithm for estimating the same probability.

3.2.2 Conditional Monte Carlo via Order Statistics

In this section we study the dependent versions of the conditional algorithms proposed by Asmussen and Binswanger (1997); Asmussen and Kroese (2006). Recall from Section 1.3.3 on page 12 that this algorithms were constructed by conditioning with respect

to the order statistics. In fact the second algorithm provided an easier implementation by exploiting a symmetry argument coming from the i.i.d. assumption.

We discussed how to drop the *identically distributed* assumption. Moreover, we proved that these algorithm kept their efficiency properties in the lognormal and regularly varying cases. Moreover, the construction of the dependent version of these algorithms follows exactly in the same manner. The only difference being the form of the conditional distribution due to the dependence. We discuss these algorithms below.

3.2.2.1 Asmussen-Binswanger Estimator: Dependent version

For extending the algorithm of Asmussen and Binswanger (1997) to the multivariate setting we need to calculate the conditional probability that $\{S_n > u\}$ given the first $d-1$ order statistics $X_{(1)}, \ldots, X_{(d-1)}$. However, since the random variables are no longer i.i.d. we also need to know the keep track of the original indexes. If all this information is contained in the σ-algebra $\mathcal{F}$, then we obtain

$$\begin{aligned}\mathbb{P}\big(S_d > u \,\big|\, \mathcal{F}\big) &= \frac{\mathbb{P}(X_K > (u - S_{(d-1)}) \vee X_{(d-1)})}{\mathbb{P}(X_K > X_{(d-1)})} \\ &= \frac{\overline{G}_K\big((u - S_{(d-1)}) \vee X_{(d-1)}, \mathbf{X}_{-k}\big)}{\overline{G}_K\big(X_{(d-1),\mathbf{X}_k}\big)},\end{aligned}$$

where $G_k(\cdot, \mathbf{x}_{-k})$ is the conditional distribution of $\big(X_k | \mathbf{X}_{-k} = \mathbf{x}_{-k}\big)$. Algorithmically we have:

Asmussen-Binswanger Algorithm: Dependent Random Variables.

- Simulate $X_1, \ldots, X_d$. Register the index of K of the largest random variable and form the order statistics $X_{(1)}, \ldots, X_{(d)}$.
- Return
$$\widehat{z}_{AB}(u) = \frac{\overline{G}_K\big((u - S_{(d-1)}) \vee X_{(d-1)}, \mathbf{X}_{-k}\big)}{\overline{G}_K\big(X_{(d-1),\mathbf{X}_k}\big)}.$$

We discuss the characteristics of this algorithm together with the dependent version of the algorithm of Asmussen and Kroese (2006) which is explained next.

3.2.2.2 Asmussen-Kroese Estimator: Dependent version

We define a random variable K such that under the probability measure $\mathbb{P}$ it follows a discrete distribution supported on $\{1, \ldots, d\}$ with $\mathbb{P}(K = i) = p_i$. So, we make

$$\mathbb{P}(S_d > x) = \sum_{k=1}^{d} \mathbb{P}(S_d > x, X_k = M_d) = \mathbb{E}\big[\frac{\mathbb{I}(S_d > x, X_K = M_d)}{p_K}\big],$$

and condition with respect to $\mathcal{F} = \sigma(K, X_1, \ldots, X_{K-1}, X_{K+1}, \ldots, X_d)$ to obtain

$$\begin{aligned}\mathbb{P}(S_d > x) &= \mathbb{E}\big[\frac{\mathbb{P}\big(X_K > M_{-K} \vee (x - S_{-K})\big)}{p_K} \,\big|\, \mathcal{F}\big] \\ &= \mathbb{E}\big[\frac{\overline{G}_K\big(M_{-K} \vee (x - S_{-K}), \mathbf{X}_{-K}\big)}{p_K}\big],\end{aligned}$$

where $G_k(\cdot, \mathbf{x}_{-k})$ is the conditional distribution of $\big(X_k | \mathbf{X}_{-k} = \mathbf{x}_{-k}\big)$ and M_{-k} and S_{-k} are defined as the maximum and sum of the X_i's without considering the k-th random variable. As before we suggest to use

$$p_k(u) = \frac{\mathbb{P}(X_k > u)}{\sum_{i=1}^{d} \mathbb{P}(X_i > u)}.$$

The algorithm is as follows:

Asmussen-Kroese Algorithm: Dependent version.

- Simulate K.
- Simulate $X_1, \ldots, X_{K-1}, X_{K+1}, \ldots, X_d$.
- Return
$$\widehat{z}_{AK}(u) = \frac{\overline{G}\big(M_{(-K)} \vee (u - S_{-K}, \mathbf{X}_{-K})\big)}{p_K}.$$

These algorithms were empirically studied in Asmussen and Rojas-Nandayapa (2006) in the case of multivariate lognormal random vectors. In such case the subexponential-like behavior of the multivariate lognormal distribution could have suggested that these algorithms are efficient. However, the following counterexample proves this intuition wrong. We consider a multivariate lognormal random vector with localization vector and dispersion matrix given by

$$\boldsymbol{\mu} = \mathbf{0}, \qquad \boldsymbol{\Sigma} := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

with $\rho > 1/3$, then the estimator will have an efficiency which is weaker than *logarithmic efficiency* (this example is due to Dominik Kortschak).

The reason is that the values of the random vector $\mathbf{X}_{-k}$ affect directly the distribution $G_k(\cdot)$. Moderately large values of the elements of $\mathbf{X}$ and positive correlations will increase the localization parameter of the conditional distribution of X_k in the tail and in consequence the probability of large values will be bigger compared to the unconditional distribution $F(\cdot)$. In the case of the Asmussen-Kroese algorithm it is easily seen that this increases the variance of the estimator.

This analysis suggested to use importance sampling over the vector $\mathbf{X}_{-k}$ before the conditioning which can reduce variance of the estimator. We show this estimator in algorithmically shape.

Asmussen-Kroese Algorithm: Dependent version

- Select an appropriate multivariate distribution function H such that it density function h is known and is possible to simulate from it.
- Simulate K.
- Simulate $X_1, \ldots, X_{K-1}, X_{K+1}, \ldots, X_d$ from H.
- Return
$$\widehat{z}_{AK}(u) = \frac{\overline{G}\big(M_{(-K)} \vee (u - S_{-K}, \mathbf{X}_{-K})\big)}{p_K} \frac{f_{-K}(\mathbf{X}_{-K})}{h_{-K}(\mathbf{X}_{-K})}.$$

As in any importance sampling implementation there are two main issues associated with the election of the change of measure. First, we need to be able to simulate from the proposed importance sampling distribution and also we need to know the likelihood ratio. Secondly, the change of measure should deliver a substantial variance reduction (remember that the implementation of the importance sampling method can in fact increase the variance). As an illustrative example we show the following result which was proved in Rojas-Nandayapa (2006).

Theorem 3.9. *Let (X_1, X_2) be a multivariate lognormal random vector with localization vector and dispersion matrix given by*

$$\boldsymbol{\mu} = \mathbf{0}, \qquad \boldsymbol{\Sigma} := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

$\overline{G}(\cdot, x_2) \sim LN(\rho \log x_2, 1-\rho^2)$ be the conditional distribution of $\big(X_1 | X_2 = x_2\big)$, $Y \sim LN(\rho \log u, 1/(1-\rho))$ an independent random variable and $f(\cdot)$ and $h(\cdot)$ be the densities of X_1 and Y respectively. Then

$$\widehat{z}_{AK} := 2\, G\big(Y \vee (u-Y),\, Y\big) \frac{f(Y)}{h(Y)}$$

is an unbiased estimator of $\mathbb{P}(X_1 + X_2 > u)$ and it has bounded relative error.

This corresponds to a dependent version of the Asmussen-Kroese algorithm using h as importance density. The set up in the hypothesis of the Theorem are too restrictive and its proof is rather long, so I have decided to omit it. However, this Theorem works fine to illustrate some features. We have chosen an importance sampling distribution with an increased localization parameter and smaller dispersion. The idea of this is to concentrate the frequency of samples over the region where the function h is larger than f, so the likelihood ratio is smaller than 1 helping to reduce the variance. The problem of this approach is that the likelihood ratio becomes difficult to handle when we increase the dimensions and the efficiency proofs become more complicated. However, we got the intuition that a good way to control the effect of the lighter conditional distribution tail of $G(\cdot)$ is to use an importance sampling distribution with smaller likelihood ratios $f(\cdot)/h(\cdot)$. Intuitively, this might be obtained by choosing a density $h(\cdot)$ with a heavier tail than $f(\cdot)$.

In the previous section we have suggested to use an importance sampling strategy by scaling the variance which can be combined with the conditional strategy shown in this section. This will be investigated in future work.

3.3 Concluding Remarks

3.3.1 Numerical Examples

In the Examples D.1 to D.3 on pages 103–105 we examined numerically the sharp asymptotic approximation given in Theorem 3.2 on page 43 (*subexponential-type*) and the upper bound obtained suggested in on page 16 (*aggregated tails*). For comparison purposes we included our best Monte Carlo estimate in terms of smallest variance. The examples correspond to low, medium and high correlations respectively

The conclusion here is that the subexponential-type approximation provide better results than the *aggregated tails*. Moreover, it is also observed that its sharpness improved as the correlation parameter is increased. Recall that the random variables

consider in our examples are nonidentically distributed. An intuitive explanation of this is that stronger correlations together with different marginal behavior decreases the probability of observing two or more random variables with large values. A different behavior should be expected in the identically distributed case.

The discussion of the numerical examples of the Monte Carlo estimators can be found in the Subsection 4.3.1 on page 65 where the estimator proposed in the next chapter has been considered for comparison purposes.

Notes and Comments

In recent years, there has been an intense research on the asymptotic tail behavior of sums of dependent heavy-tailed random variables. The following are some results which are related to some extent to the asymptotic approximation given in the Section 3.1.

Several papers have taken the approach of constructing bounds for the tail probability of the sum of general random variables. Denuit et al. (1999) provided upper and lower bounds which do not rely on any assumption on the nature of the dependence. Cossette et al. (2002); Embrechts and Puccetti (2006a); Mesfioui and Quessy (2006) provided extensions on best possible bounds of the distribution of nondecreasing functions of n dependent risks. Embrechts and Puccetti (2006b) studied the tail distribution of functions of dependent random vectors having fixed multivariate marginals. The case $n = 2$ was studied in Albrecher et al. (2006) where they provided bounds for the sum of two heavy-tailed random variables in a general copula framework.

More explicit results had been obtained by considering asymptotic approximations for the tail probability of a sum where it is assumed that the marginals are heavy-tailed distributions. Albrecher et al. (2006) investigated exhaustively the case of two dependent random variables. In the more general case of n random variables, Wüthrich (2003); Alink et al. (2004, 2005) used multivariate extreme value theory to provide asymptotic bounds $\mathbb{P}(S_n > u)$ in the case where the variables are exchangeable with Archimedean copula. This results were generalized in Alink et al. (2007) for a subclass of a symmetric copulas. Barbe et al. (2006) gave asymptotic for the tail probability in the case where the random variables follow a multivariate regularly varying distribution in terms of a measure associated with the extreme value copula. Foss and Richards (2008) consider the case where the involved random variables are conditionally independent.

In the nonexchangeable case several results have been developed recently. Kortschak and Albrecher (2008) generalized some of the results in Alink et al. (2004); Barbe et al. (2006) for nonidentical and not necessarily positive random variables. Goovaerts et al. (2004); Tang and Wang (2006) derived asymptotic expressions for a randomly weighted sum of regularly varying random variables. Wang and Tang (2004); Geluk and Ng (2006); Tang (2006); Ko and Tang (2008) analyzed tail probabilities of sums of negatively associated sums of heavy-tailed random variables. In particular in Ko and Tang (2008); Tang (2006, 2008) dependence structures which do not impact the tail behavior of subexponential marginals are investigated.

In the lognormal case Vanduffel et al. (2008) obtained approximations for the sum of conditionally independent lognormal random variables and investigate some links with the results in Asmussen and Rojas-Nandayapa (2008). Szyszkowicz and Yanikomerloglu (2008) show that the distribtuion of the sum of correlated exchangeable lognormals can be well approximated with a lognormal random variable and provide the parameters.

We leave the discussion of Monte Carlo estimates for the Notes and Comments section in the next chapter.

3.4 Proofs.

Proof of Proposition 3.1. Note that

$$\lim_{u\to\infty}\frac{\mathbb{P}(M_d>u)}{\mathbb{P}(\mu+\sigma\gamma Z>\log u)}\geq\lim_{u\to\infty}\frac{\sum_{i=1}^d\mathbb{P}(X_i>u)-\sum_{j\neq i}\mathbb{P}(X_i>u,X_j>u)}{\mathbb{P}(\mu+\sigma\gamma Z>\log u)},$$

recall $Z\sim N(0,1)$. We claim that the last line in the previous display is asymptotically equivalent to

$$\lim_{u\to\infty}\frac{\sum_{i=1}^d\mathbb{P}(X_i>u)}{\mathbb{P}(\mu+\sigma\gamma Z>\log u)}=m_d.$$

To see this, write

$$\mathbb{P}(X_i>u,X_j>u)=\begin{cases}\mathbb{P}(X_i>u\big|X_j(u)>u)\mathbb{P}(X_j>u)\\ \mathbb{P}(X_j>u\big|X_i(u)>u)\mathbb{P}(X_i>u).\end{cases}$$

If $\mathbb{P}(X_k>u)=o\big(\mathbb{P}(\mu+\sigma\gamma Z>\log u)\big)$ as $u\nearrow\infty$ for $k=i$ or $k=j$, then the claim holds immediately, so, the interesting case is obtained when the X_i and X_j are identically distributed with $\mathbb{P}(X_k>u)=\mathbb{P}(\mu+\sigma\gamma Z>\log u)$. However, it is well known (see, for instance, Joe, 1997; Nelsen, 2006) that if Z_1 and Z_2 are jointly standard Gaussian r.v.'s then $\mathbb{P}(Z_1>u\big|Z_2>u)\to 0$ as $u\to\infty$. A straightforward adaptation of this result to the case of Gaussian random variables with scaled covariance structure allows to conclude the previous claim in this case and, in turn, the result. □

Proof of Theorem 3.2. We proceed by induction. The case $n=1$ is straightforward. For the induction step, we assume that the Theorem holds for any arbitrary lognormal random vector with Gaussian copula of size d. Next, we will prove that the Theorem is true for a random vector of size $d+1$. For the proof, the following assumptions (which are made w.l.o.g.) are convenient:

A1. $X_1(u),\ldots,X_{d+1}(u)$ are ordered in such way that if $\ell<k$ then $X_k(u)$ and $X_\ell(u)$ either have the same marginal distribution, or $X_k(u)$ has lighter tail than $X_\ell(u)$. Thus $\overline{F}_{\mu,\sigma^2}(u)=\mathbb{P}(X_1(u)>u)$.

A2. $\mu=0$. If not, replace $X_k(u)$ and u by $X_k(u)\mathrm{e}^{-\mu}$ and $u\mathrm{e}^{-\mu}$ in $\mathbb{P}(S_{d+1}(u)>u)$.

Choose β as in Lemma 3.10 on the following page and consider the following relations

$$\begin{aligned}&\mathbb{P}(S_{d+1}(u)>u)\\ &\quad=\mathbb{P}(S_{d+1}(u)>u,S_d(u)\leq u^\beta)+\mathbb{P}(S_{d+1}(u)>u,S_d(u)>u^\beta,X_{d+1}(u)\leq u^\beta)\\ &\qquad\quad+\mathbb{P}(S_{d+1}(u)>u,S_d(u)>u^\beta,X_{d+1}(u)>u^\beta)\\ &\quad\leq\mathbb{P}(X_{d+1}(u)>u-u^\beta)+\mathbb{P}(S_d(u)>u-u^\beta)+\mathbb{P}(S_d(u)>u^\beta,X_{d+1}(u)>u^\beta).\end{aligned}$$

From the same Lemma it follows that

$$\limsup_{u\to\infty}\frac{\mathbb{P}(S_{d+1}(u)>u)}{\mathbb{P}(X_1(u)>u)}\leq\limsup_{u\to\infty}\frac{\mathbb{P}(X_{d+1}(u)>u-u^\beta)}{\mathbb{P}(X_1(u)>u)}\;+\;\limsup_{u\to\infty}\frac{\mathbb{P}(S_d(u)>u-u^\beta)}{\mathbb{P}(X_1(u)>u)}.$$

Here assumption A1 and the induction hypothesis guarantee that the two limsup's on the r.h.s. are actually limits, so that the r.h.s. becomes

$$\lim_{u\to\infty}\frac{\mathbb{P}(X_{d+1}(u)>u)}{\mathbb{P}(X_1(u)>u)}+\lim_{u\to\infty}\frac{\mathbb{P}(S_d(u)>u-u^\beta)}{\mathbb{P}(X_1(u)>u-u^\beta)}=\lim_{u\to\infty}\frac{\mathbb{P}(X_{d+1}(u)>u)}{\mathbb{P}(X_1(u)>u)}+m_d^*,$$

where the first step uses assumption A1 and the second step the induction hypothesis with m_d^* denoting the number among $X_1(u), \ldots, X_d$ which have the same tail as $X_1(u)$. From assumption A1 the limit in the r.h.s. is 0 if $X_{d+1}(u)$ has lighter tail than $X_1(u)$, or 1, if $X_{d+1}(u)$ and $X_1(u)$ have the same distribution. Observe that by A1, $X_1(u)$ cannot have lighter tail than $X_{d+1}(u)$. We have proved

$$\limsup_{u\to\infty} \frac{\mathbb{P}(S_{d+1}(u) > u)}{\mathbb{P}(X_1(u) > u)} = \limsup_{u\to\infty} \frac{\mathbb{P}(S_{d+1}(u) > u)}{\mathbb{P}(\mu + \sigma\gamma Z > \log u)} \leq m_{d+1}, \tag{3.6}$$

for the appropriate value of m_{d+1}. Together with Proposition 3.1 on page 43 (which provides a lower bound) this establishes the induction step and completes the proof. □

Lemma 3.10. *Under the hypothesis of Theorem 3.2 on page 43, if A1 and the induction hypothesis hold, then there exists* $0 < \beta < 1$ *such that*

$$\limsup_{u\to\infty} \frac{\mathbb{P}(S_d(u) > u^\beta, X_{d+1}(u) > u^\beta)}{\mathbb{P}(X_1(u) > u)} = 0.$$

Proof. If $\sigma_{d+1} < \sigma$ choose $\sigma_{d+1}/\sigma < \beta < 1$. Then

$$\begin{aligned}\limsup_{u\to\infty} \frac{\mathbb{P}(S_d(u) > u^\beta, X_{d+1}(u) > u^\beta)}{\mathbb{P}(X_1(u) > u)} &\leq \limsup_{u\to\infty} \frac{\mathbb{P}(X_{d+1}(u) > u^\beta)}{\mathbb{P}(X_1(u) > u)} \\ &= \lim_{u\to\infty} \frac{\mathbb{P}(X_{d+1}(u)^{1/\beta} > u)}{\mathbb{P}(X_1(u) > u)}.\end{aligned}$$

But the last limit is 0 since $X_{d+1}(u)^{1/\beta} \sim \mathrm{LN}\big(\mu_{d+1}/\beta, \gamma(u)\sigma_{d+1}^2/\beta^2\big)$ has lighter tail than $X_1(u)$ because of $\sigma_{d+1}^2/\beta^2 < \sigma^2$.

If $\sigma = \sigma_{d+1}$ define $\eta = \max_{k=1,\ldots,d}\{\sigma_{k(d+1)}/\sigma^2\}$. By A1, we have that $\sigma_k^2 = \sigma^2$ for $k = 1, \ldots, d+1$. Then η is the maximum among the correlations between $X_k(u)$ and $X_{d+1}(u)$ and by the hypothesis of the Theorem 3.2 on page 43 it should take values $\eta \in [-1, 1)$. Therefore, we can choose β close enough to 1 to obtain $\max\{1/2, \eta\} < \beta^2 < 1$ and $(\beta - \eta/\beta)^2 + \beta^2 > 1$ (observe that $(1 - \eta/1)^2 + 1^2 > 1$). Consider

$$\begin{aligned}&\limsup_{u\to\infty} \frac{\mathbb{P}(S_d(u) > u^\beta, X_{d+1}(u) > u^\beta)}{\mathbb{P}(X_1(u) > u)} \\ &\quad= \limsup_{u\to\infty} \frac{\mathbb{P}(X_{d+1}(u) > u^{1/\beta})}{\mathbb{P}(X_1(u) > u)} + \limsup_{u\to\infty} \frac{\mathbb{P}(S_d(u) > u^\beta, u^{1/\beta} > X_{d+1}(u) > u^\beta)}{\mathbb{P}(X_1(u) > u)}.\end{aligned}$$

Here the first limit is 0 since $X_{d+1}(u)^\beta \sim \mathrm{LN}\big(\beta\mu_d, \gamma(u)(\beta\sigma)^2\big)$ has lighter tail than $X_1(u)$ because of $\beta < 1$. For the second limit, we define

$$(X_1^c(u,y), \ldots, X_d^c(u,y)) = (X_1(u), \ldots, X_d(u) | X_{d+1}(u) = y), \quad S_d^c(u,y) = \sum_{i=1}^d X_i^c(u,y).$$

So,

$$\begin{aligned}&\limsup_{u\to\infty} \frac{\mathbb{P}(S_d(u) > u^\beta, u^{1/\beta} > X_{d+1}(u) > u^\beta)}{\mathbb{P}(X_1(u) > u)} \\ &\qquad= \limsup_{u\to\infty} \frac{1}{\mathbb{P}(X_1(u) > u)} \int_{u^\beta}^{u^{1/\beta}} \mathbb{P}(S_d^c(u,y) > u^\beta) f_{X_{d+1}(u)}(y) dy.\end{aligned}$$

Standard formulas for the conditional mean vector and conditional covariance matrix in the multivariate normal distribution yield

$$\mathbf{X}^c(u,t) \sim LN(\{\mu_i + \frac{\sigma_{i(d+1)}}{\sigma^2}(\log t - \mu_{d+1})\}_i, \gamma(u)\{\sigma_{ij} - u_{ij}\}_{ij}), \quad u_{ij} = \frac{\sigma_{i(d+1)}\sigma_{j(d+1)}}{\sigma^2}.$$

We can restrict to consider values of $t > 1$ since

$$X_i^c(t) \stackrel{d}{=} X_i^c(1) \exp\big\{\frac{\sigma_{i(d+1)}}{\sigma^2}\log t\big\}.$$

Then it follows that $X_i^c(u,t)$ is smaller than $X_i^c(u,1)\mathrm{e}^{\eta \log t}$ in stochastic order. So, $S_d(u)^c(t) \leq S_d(u)^c(1)t^\eta$ in stochastic order, and the last limit above can be above bounded by

$$\limsup_{u\to\infty} \frac{1}{\mathbb{P}(X_1(u) > u)} \int_{u^\beta}^{u^{1/\beta}} \mathbb{P}(S_d^c(u,1) > u^\beta/y^\eta) f_{X_{d+1}(u)}(y)dy. \tag{3.7}$$

If $\eta \leq 0$, the integral in (3.7) is bounded by $\mathbb{P}(S_d^c(u,1) > u^\beta)\mathbb{P}(X_{d+1}(u) > u^\beta)$ which by the induction hypothesis is asymptotically equivalent to $m_d^c\mathbb{P}(X_k^c(u,1) > u^\beta)\mathbb{P}(X_{d+1}(u) > u^\beta)$ with the appropriate integer value m_d^c and index k. Now, from the form of the distribution of $\mathbf{X}^c(u,y)$ and A1 it follows that $X_k(u)^c(u,1)$ and $X_{d+1}(u)$ have lighter or equivalent tails than $X_1(u)$, so bounding m_d^c by d we have proved that

$$\limsup_{u\to\infty} \frac{\mathbb{P}(S_d > u^\beta, u^{1/\beta} > X_{d+1}(u) > u^\beta)}{\mathbb{P}(X_1(u) > u)} \leq \limsup_{u\to\infty} \frac{d\mathbb{P}^2(X_1(u) > u^\beta)}{\mathbb{P}(X_1(u) > u)}.$$

The last limit is 0 because of the choice $2\beta^2 > 1$ and A1. In the case where $\eta > 0$, the integral in expression (3.7) can be bounded by $\mathbb{P}(S_d^c(u,1) > u^{\beta-\eta/\beta})\mathbb{P}(X_{d+1}(u) > u^\beta)$. Observe that $\beta - \eta/\beta > 0$ since we took $\beta^2 > \eta$; so we can use the same argument as above to conclude that

$$\begin{aligned}\limsup_{u\to\infty} &\frac{\mathbb{P}(S_d(u) > u^\beta, u^{1/\beta} > X_{d+1}(u) > u^\beta)}{\mathbb{P}(X_1(u) > u)} \\ &\leq \limsup_{u\to\infty} \frac{d\,\mathbb{P}(X_1(u) > u^{\beta-\eta/\beta})\mathbb{P}(X_1(u) > u^\beta)}{\mathbb{P}(X_1(u) > u)},\end{aligned}$$

which is again 0 because of the choice $(\beta - \eta/\beta)^2 + \beta^2 > 1$ and the asymptotic relation

$$\begin{aligned}&\mathbb{P}(X_1(u) > u^{\beta-\eta/\beta})\mathbb{P}(X_1(u) > u^\beta) \\ &\qquad = \frac{\sigma^2}{2\pi \log u^{\beta-\eta/\beta} \log u^\beta} \exp\Big\{ - \frac{[(\beta - \eta/\beta)^2 + \beta^2]\log^2 u}{2\sigma^2\gamma^*}\Big\}\big(1 + o(1)\big).\end{aligned}$$

□

Proof of Lemma 3.4 on page 45.

$$\begin{aligned}
\mathbb{E}_\theta \widehat{z}_1(u)^2 &= \int \frac{\mathbb{I}(e^{y_1}+\cdots+e^{y_d}>u)\exp\big\{-2\theta(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2\big\}}{(1-\theta)^d} \\
&\qquad \times \frac{\exp\big\{-(1-\theta)(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2\big\}}{(2\pi)^{d/2}\det\big(\boldsymbol{\Sigma}/(1-\theta)\big)^{1/2}} dy_1\ldots dy_d \\
&= \int \mathbb{I}\big\{e^{y_1}+\cdots+e^{y_d}>u\big\} \\
&\qquad \times \frac{\exp\big\{-(1+\theta)(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2\big\}}{(1-\theta)^{d/2}(1+\theta)^{d/2}(2\pi)^{d/2}\det(\boldsymbol{\Sigma}/(1+\theta))^{1/2}} dy_1\ldots dy_d \\
&= (1-\theta^2)^{-d/2} Q_\theta(S_d>u).
\end{aligned}$$

□

Proof of Theorem 3.5. Theorem 3.2 applied with $\gamma(u) = 1/\big(1+\theta(u)\big)$ together with a straightforward extension of approximation (3.1) in the case of scaled variances yields

$$\begin{aligned}
Q_{\theta(u)}(S_d>u) &= \Theta\big(\mathbb{P}(\mu+\sigma(1+\theta(u))^{-1/2}N(0,1)>\log u)\big) \\
&= \Theta\bigg(\mathbb{P}\bigg(Z>\frac{\log u-\mu}{\sigma(1+\theta(u))^{-1/2}}\bigg)\bigg) \\
&= \Theta\bigg(\frac{1}{\log u-\mu}\exp\bigg\{-\frac{(\log u-\mu)^2(1+\theta(u))}{2\sigma^2}\bigg\}\bigg) \\
&= \Theta\bigg(\frac{1}{\log u}\exp\bigg\{-\frac{(\log u-\mu)^2(2-\psi(u))}{2\sigma^2}\bigg\}\bigg)
\end{aligned}$$

Since we have that

$$z(u)^{2-\epsilon} = \Theta\bigg(\frac{1}{(\log u)^{2-\epsilon}}\exp\bigg\{-\frac{(\log u-\mu)^2(2-\epsilon)}{2\sigma^2}\bigg\}\bigg),$$

the result follows by noting that

$$\big(1-\theta(u)^2\big)^{-d/2} = \Theta\big((\psi(u))^{-d/2}\big)$$

and plugging in this estimate together with that of $Q_{\theta(u)}(S_d>u)$ into representation (3.4). □

Proof of Lemma 3.6. First, we note that existence and uniqueness for sufficiently large u follows easily by virtue of a monotonicity argument. Note that

$$\mathrm{e}^{\mu+\sigma^2/2(1-\theta(u))} \le \mathbb{E}_{\theta(u)}S_d \le d\mathrm{e}^{\mu+\sigma^2/2(1-\theta(u))}.$$

Let $\theta_+(u)$ be the solution to the equation

$$\mathrm{e}^{\sigma^2/2(1-\theta_+(u))} = u\mathrm{e}^{-\mu}/d.$$

We must have that $1-\theta(u) \le 1-\theta_+(u)$. However,

$$\frac{\sigma^2}{2\big(1-\theta_+(u)\big)} = \log u - \mu - \log d.$$

Moreover, we also have that

$$1-\theta(u) \geq 1-\theta_{-}(u) = \frac{\sigma^2}{2(\log u - \mu)}.$$

These observations imply the statement of the lemma. □

Proof of Corollary 3.7. Since $1-\theta(u) = \log(u)^{-1}\sigma^2/2\big(1+o(1)\big)$ by Proposition 3.7 we get

$$\exp\left\{-\frac{(1-\theta(u))\big(\log u-\mu\big)^2}{2\sigma^2}\right\} = u^{1/4}\big(1+o(1)\big).$$

Inserting this in (3.3) the result follows.

□

Proof of Theorem 3.4. Note that

$$Q_{\theta(u)}\Big(S_d > u, \max_{1\leq i\leq d} X_i \leq u\Big) = \sum_{k=1}^{d} Q_{\theta(u)}\Big(S_d > u, X_k = \max_{1\leq i\leq d} X_i \leq u\Big).$$

So, we choose β as in Lemma 3.11 and consider the following decomposition

$$\begin{aligned}
&\sum_{k=1}^{d} Q_{\theta(u)}\Big(S_d > u,\, X_k = \max_{1\leq i\leq d} X_k < u\Big) \\
&\quad = \sum_{k=1}^{d} Q_{\theta(u)}\Big(S_d > u,\, X_k = \max_{1\leq i\leq d} X_k < u,\, S_{d,-k} > u^{\beta}\Big) \\
&\qquad + Q_{\theta(u)}\Big(S_d > u,\, X_k = \max_{1\leq i\leq d} X_k < u,\, S_{d,-k} < u^{\beta}\Big) \\
&\quad \leq \sum_{k=1}^{d} Q_{\theta(u)}\Big(S_{d,-k}(u) > u^{\beta}, X_k(u) > u^{\beta}/d\Big) + Q_{\theta(u)}\Big(u-u^{\beta} < X_k(u) < u\Big). \quad (3.8)
\end{aligned}$$

We take $\gamma = 1-\beta$ in Lemma 3.11 and together with Lemma 3.12 on page 59 yields

$$\frac{Q_{\theta(u)}\big(S_d > u, \max_{1\leq i\leq d} X_i \leq u\big)}{\alpha(u)^2} = O\bigg(\frac{(\log u)^2}{u^{1-\beta}}\bigg),$$

Finally, the result follows by inserting this in representation 3.5 and noting that

$$\big(1-\theta(u)^2\big)^{-d/2} = \Theta\big(\psi(u)^{-d/2}\big).$$

□

Lemma 3.11. *There exists $\beta \in \big(0,1\big)$ such that for any $\gamma \in \mathbb{R}$ it follows that*

$$\frac{Q_{\theta(u)}(S_{d,-k} > u^{\beta}, X_k > u^{\beta}/d)}{u^{\gamma} z(u)^2} = o(1), \qquad k = 1,\ldots,d,$$

where

$$S_{d,-k} := X_1 + \cdots + X_{k-1} + X_{k+1} + \cdots + X_d.$$

Proof. For the proof we will consider two cases. The first when $\sigma_k < \sigma$ and the second when $\sigma_k = \sigma$ (cf. assumption A1).

Case 1. If $\sigma_k \neq \sigma$ take $\beta_k := \sigma_k/\sigma$ and observe that

$$\begin{aligned} Q_{\theta(u)}(S_{d,-k} > u^{\beta_k}, X_k > u^{\beta_k}/d) &\leq Q_\theta(X_k > u^{\beta_k}/d) \\ &= \mathbb{P}\Big(\mu_k + \sigma_k(2-\psi(u))^{-1/2}N(0,1) > \beta_k \log u - \log d\Big) \\ &= \mathbb{P}\bigg(N\bigg(\frac{\mu_k + \log d}{\beta_k}, \frac{\sigma_k^2}{{\beta_k}^2(2-\psi(u))}\bigg) > \log u\bigg). \end{aligned}$$

By Mill's ratio the last term is dominated by the tail of $z^2(u)$ even after premultiplying by a power term $b^\gamma = \mathrm{e}^{\gamma \log b}$. The assertion of the Lemma follows from here.

Case 2. If $\sigma_k = \sigma$ define $\eta = \max\{\sigma_{\ell.k}/\sigma^2 : \ell \neq k\}$. By assumption, we have that $\sigma_k^2 \leq \sigma^2$ for $k = 1, \ldots, d$, and thereover

$$1 > \Big|\max_{\ell \neq k}\{\rho_{k\ell}\}\Big| = \Big|\max_{\ell \neq k}\Big\{\frac{\sigma_{\ell,k}}{\sigma_\ell \sigma_k}\Big\}\Big| \geq \Big|\max_{k=1,\ldots,d}\Big\{\frac{\sigma_{\ell,k}}{\sigma^2}\Big\}\Big| \geq \eta.$$

Therefore $\eta \in [-1, 1)$, so we can choose β_k close enough to 1 such that $\max\{1/2, \eta\} < \beta_k^2 < 1$ and $(\beta_k - \eta/\beta_k)^2 + \beta_k^2 > 1$; note that such β_k can always be chosen by continuity since $(1-\eta/1)^2 + 1^2 > 1$. Consider

$$\begin{aligned} &Q_{\theta(u)}\big(S_{d,-k} > u^{\beta_k}, X_k > u^{\beta_k}/d\big) \\ &\quad \leq Q_{\theta(u)}\big(S_{d,-k}(u) > u^{\beta_k}, u^{\beta_k}/d < X_k < u^{1/\beta_k}\big) + Q_{\theta(u)}\big(S_{d,-k} > u^{\beta_k}, u^{1/\beta_k} < X_k\big) \\ &\quad \leq Q_{\theta(u)}\big(S_{d,-k}(u) > u^{\beta_k}, u^{\beta_k}/d < X_k < u^{1/\beta_k}\big) + Q_{\theta(u)}\big(u^{1/\beta_k} < X_k\big). \end{aligned} \tag{3.9}$$

Define $Q'_{\theta(u),t}(\cdot)$ the probability measure under which

$$\mathbf{Y} \sim N_d\bigg(\boldsymbol{\mu} + \boldsymbol{\Sigma}_{\cdot,k}\frac{t-\mu_k}{\sigma^2}, (1+\theta^2)\bigg(\boldsymbol{\Sigma} - \frac{\boldsymbol{\Sigma}_{\cdot,k}\boldsymbol{\Sigma}_{k,\cdot}}{\sigma^2}\bigg)\bigg),$$

or equivalently the conditional distribution of $\mathbf{Y}|Y_k = t$. Moreover, since $\mathbf{Y}$ has the same distribution under the measure $Q'_{\theta(u),t}$ than $\mathbf{Y} + \boldsymbol{\Sigma}_{\cdot,k}\, t/\sigma_k^2$ under the measure $Q'_{\theta(u),0}$ and η was chosen in such way that $\eta \geq \boldsymbol{\Sigma}_{\cdot,k}/\sigma_k^2$, then if $u > 1$ it holds that

$$\begin{aligned} &Q_{\theta(u)}(S_{d,-k} > u^{\beta_k}, u^{1/\beta_k} > X_k > u^{\beta_k}/d) \\ &\qquad = E^{Q_{\theta(u)}}\big(Q'_{\theta(u),Y_k}(S_{d,-k} > u^{\beta_k})\,;\, \beta_k \log u < Y_k < \log u/\beta_k\big) \\ &\qquad \leq E^{Q_{\theta(u)}}\big(Q'_{\theta(u),0}(S_{d,-k}\, e^{\eta Y_k} > u^{\beta_k})\,;\, \beta_k \log u < Y_k < \log u/\beta_k\big). \end{aligned} \tag{3.10}$$

If $\eta \leq 0$, the previous expectation is bounded by

$$Q'_{\theta(u),0}\big(S_{d,-k} > u^{\beta_k}\big) Q_{\theta(u)}\big(X_k > u^{\beta_k/d}\big).$$

The previous two factors have lognormal tails due to Theorem 3.2. In fact, since the covariances of the Gaussian conditional random variables are never larger than the unconditional ones we obtain the following relation

$$Q'_{\theta(u),0}\big(S_{d,-k} > u^{\beta_k}\big) Q_{\theta(u)}\big(X_k > u^{\beta_k/d}\big) = o\big(\mathbb{P}_{\theta(u)}(X_k > u^{\beta_k})\mathbb{P}_{\theta(u)}(X_k > u^{\beta_k/d})\big),$$

and in turn we have

$$\mathbb{P}_{\theta(u)}(X_k > u^{\beta_k})\mathbb{P}_{\theta(u)}(X_k > u^{\beta_k/d}) = o\big(\mathbb{P}^2_{\theta(u)}(X_k > u^{\beta_k/d})\big).$$

By Mill's ratio we obtain that the last expression is equivalent to

$$\Theta\Big(\frac{1}{\log(u)^2}\exp\Big\{-\frac{2\beta_k^2(1+\theta(u))(\log(u)-(\mu_k+\log(d))/d)^2}{2\sigma^2}\Big\}\Big).$$

Since we choose $2\beta_k^2 > 1$ the last expression is dominated by the tail of $z^2(u)$ and this result result holds after multiplying by a power term $u^\gamma = \exp\big(\gamma \log u\big)$.

In the case where $\eta > 0$, the expression (3.10) can be bounded by

$$Q'_{\theta(u),0}\big(S_{d,-k} > u^{\beta_k-\eta/\beta_k}\big)Q_{\theta(u)}\big(X_k > u^{\beta_k}/d\big)$$

Observe that $\beta_k - \eta/\beta_k > 0$ since we took ${\beta_k}^2 > \eta$ – otherwise $u^{\beta_k-\eta/\beta_k} \to 0$ and the first term will go to 1, so we can use a similar argument as above to conclude that

$$\begin{aligned}&Q'_{\theta(u),0}\big(S_{d,-k} > u^{\beta_k-\eta/\beta_k}\big)Q_{\theta(u)}\big(X_k > u^{\beta_k}/d\big)\\ &\quad= o\Big(\frac{1}{\log(u)^2}\exp\Big\{-\frac{2\big((\beta_k-\eta/\beta_k)^2+{\beta_k}^2\big)(1+\theta(u))(\log(u)-\mu_k)^2}{2\sigma^2}\Big\}\Big)\end{aligned}$$

which again is dominated by $z^2(u)$ because of the choice $(\beta_k-\eta/\beta_k)^2+{\beta_k}^2 > 1$. Again, multiplying by a power function will not alter the result of the Theorem. We conclude the proof by selecting β such that

$$\max\{\beta_1,\dots,\beta_d\} < \beta < 1.$$

□

Lemma 3.12.

$$\frac{Q_{\theta(u)}(u-u^\beta < X_k < u)}{\alpha(u)^2} = O\big(\frac{(\log u)^2}{u^{1-\beta}}\big), \qquad k = 1,\dots,d,$$

for any $0 < \beta < 1$.

Proof. Take

$$\begin{aligned}&Q_{\theta(u)}(u-u^\beta < X_k < u)\\ &\quad= Q_{\theta(u)}\big(X_k > u(1-u^{\beta-1})\big) - Q_{\theta(u)}(X_k > u)\\ &\quad\sim \Big(\frac{1+\theta(u)}{\sqrt{2\pi}\sigma_k\log(u-u^{\beta_k})}\exp\Big\{-\frac{(\log(u-u^{\beta_k})-\mu_k)^2}{2\sigma_k^2/(1+\theta(u))}\Big\} - Q_{\theta(u)}(X_k > u)\Big)\\ &\quad= \Big(Q_{\theta(u)}(X_k > u)\Big[\exp\Big\{-\frac{2(\log u-\mu_k)\log(1-u^{\beta-1})+\log^2(1-u^{\beta-1})}{2\sigma_k^2/(1+\theta(u)^2)}\Big\}-1\Big]\Big)\end{aligned}$$

Using basic calculus we can verify that the expression in the brackets is

$$\Theta\big(\frac{\log u}{u^{1-\beta}\sigma^2}\big).$$

Inserting this expansion in the limit we prove the lemma. □

Chapter 4

Tail Probabilities for Log-elliptical Distributions

In this chapter we will consider the class of multivariate distributions obtained via an exponential transformation of the so called *elliptical distributions*. The motivation is the analysis of a more general class of multivariate distributions with heavy-tailed marginal distributions.

The class of *elliptical* distributions has gained some attention during the last few years since it allows to analyze under the same framework a large class of multivariate distributions which share similar properties. Several well known multivariate distributions belong to this class: the multivariate normal, the multivariate t, the normal mixtures and generalized hyperbolic distributions.

A further attractive characteristic of this class of distributions is its relative flexibility for modeling the marginal behavior. It will be explained later how the particular choice of a radial random variable will lead to different marginal behaviors. However, this freedom is quite limited compared to that of a copula. For instance, not every marginal behavior can be obtained and the marginal distributions of the elements of an elliptical distribution belong to the same parametric class of distributions. One may raise some concerns as those in Mikosch (2006) for this class of multivariate distributions. However, our main interest in this chapter will be to suggest an efficient Monte Carlo algorithm for the sum of the elements of correlated lognormal random variables, and analyze it from the more general perspective of *elliptical distributions*.

The rest of the chapter is organized as follows. In the Section 4.1 we will briefly study the class of distributions obtained by an exponential transformation of an *elliptical* distribution. We will refer to this class of multivariate distributions as *log-elliptical* distributions.

In the Section 4.2 we will propose and analyze a Monte Carlo algorithm for estimating the tail probability

$$\mathbb{P}(X_1 + \cdots + X_d > u), \qquad u \to \infty,$$

where the X_i's have a log-elliptical distribution. Besides a general description algorithm we provide some general guidelines to prove that the algorithm proposed is efficient. Moreover, we prove that in the particular case of a multivariate lognormal distribution the algorithm can achieve *zero relative error*.

Finally, in the Section 4.3.1 we provide some discussion and references of recent research related to the material presented in the chapters 3 and 4.

4.1 The Log-elliptical Distribution

A more detailed description of the class of *elliptical* distributions and its properties can be found in A.2.3 on page 93. Here we will just limit ourselves to provide the definition of a *log-elliptical* distributions together with some useful details for the implementation of the Monte Carlos estimator in the next section.

For the rest of this chapter we will define the set in $\mathbb{R}^d$

$$\mathcal{S}_\epsilon(\mathbf{y}) := \{\mathbf{x} \in \mathbb{R}^d : \|x - y\| = \epsilon\},$$

and refer to it as the *spheroid* of radius ϵ centered in $\mathbf{y}$. When $\epsilon = 1$ we will simply say *unit spheroid*. In a similar fashion, we define

$$\mathcal{B}_\epsilon(\mathbf{y}) := \{\mathbf{x} \in \mathbb{R}^d : \|x - y\| \leq \epsilon\}$$

and refer to it as the ball of radius ϵ centered in $\mathbf{y}$.

Definition 4.1 (Log-elliptical distribution)**.** We say that a random vector $\mathbf{X}$ follows a log-elliptical distribution if it has the stochastic representation

$$\log \mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + R\,\mathbf{A}\,\mathbf{C}, \tag{4.1}$$

where $\mathbf{C}$ is a random vector uniformly distributed over the unit spheroid $\mathcal{S}_1(\mathbf{0})$, R is a random variable which takes values in $[0, \infty)$, $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\mathbf{A} \in \mathbb{R}^{k \times d}$.

We denote $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ where $\boldsymbol{\mu}$ is known as the *localization parameter*, $\boldsymbol{\Sigma} = \mathbf{A}^T\mathbf{A}$ as the *dispersion matrix* and ψ is the *characteristic generator* of the distribution (see the Section A.2.2 in the appendix part) . It should be noted that the stochastic representation (4.1) is not unique. A random vector $\mathbf{X}$ with $\log \mathbf{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ will also have the alternative stochastic representation

$$\log \mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + R\,\mathbf{B}\,\mathbf{C},$$

where $\mathbf{B}$ is any other square matrix such that $\mathbf{B}^T\mathbf{B} = \boldsymbol{\Sigma}$. Another important feature of the stochastic representation is that if R has no point mass at the origin then

$$R \stackrel{d}{=} \sqrt{(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})}, \qquad \mathbf{C} \stackrel{d}{=} \frac{\boldsymbol{\Sigma}^{-1/2}\big(X - \boldsymbol{\mu}\big)}{\sqrt{(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})}}.$$

From this expression it can be easily seen that in the multivariate normal case the square of the radial random variable R^2 will follow a χ^2_d chi-square distribution with d degrees of freedom. We show the method to simulate from an elliptical $E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ distribution. The simulation of log-elliptical random vectors follows by a simple exponential transformation of an elliptical random vector (see page 96).

4.2 Monte Carlo Estimation

We want to obtain a Monte Carlo estimate of the tail probability

$$\mathbb{P}(X_1, \ldots, X_d > u), \qquad u \to \infty.$$

where $\log \mathbf{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$. Here the main idea will be to use the stochastic representation (4.1) and condition over the random vector $\mathbf{C}$ which is uniformly distributed over the unit sphere. This will come out as

$$\mathbb{P}(S_d > u) = \mathbb{P}(h(R, \mathbf{C}) > u) = \mathbb{E}\left[\mathbb{P}(h(R, \mathbf{C}) > u | \, \mathbf{C})\right],$$

where we have simplified our notation by taking

$$h(r, \mathbf{c}) := \mathrm{e}^{r(A_{1,1}c_1 + \cdots + A_{1,d}c_d) + \mu_1} + \cdots + \mathrm{e}^{r(A_{d,1}c_1 + \cdots + A_{d,d}c_d) + \mu_d}, \tag{4.2}$$

where $A_{i,j}$ correspond to entries of the matrix $\mathbf{A}$. The algorithm should determine the region $\{r \in \mathbb{R} : h(r, \mathbf{c}) > u)$ for every possible outcome $\mathbf{c}$, calculate its probability and return it as an unbiased estimator of $\mathbb{P}(S_d > u)$.

By taking the second derivative it is easily verified that $h(r, \mathbf{c})$ is strictly convex, so the problem $h(r, \mathbf{c}) = u$ should have at most two solutions, say $\Psi_L(u, \mathbf{c})$ and $\Psi_U(u, \mathbf{c})$. Therefore, the aforementioned region should be the union of at most two intervals of the form $(-\infty, \Psi_L(u, \mathbf{c}))$ and $(\Psi_U(u, \mathbf{c}), \infty)$ (where the subindexes stand for *lower* and *upper*). First we should note that for any given outcome $\mathbf{c}$ there exist three possible scenarios listed below.

S1. The function $h(r, \mathbf{c})$ is strictly decreasing. In such case there exist at most one solution $\Psi_L(u, \mathbf{c})$ which tends to minus infinity as $u \to \infty$ (if no solution exists we define $\Psi_L(u, \mathbf{c}) := -\infty$). We set $\Psi_U(u, \mathbf{c}) := \infty$.

S2. The function $h(r, \mathbf{c})$ is both decreasing and increasing. In such case there exist at most two solutions $\Psi_L(u, \mathbf{c})$ and $\Psi_U(u, \mathbf{c})$ tending to $\pm\infty$ respectively as $u \to \infty$. In the cases where only one or none solutions exists we set $\Psi_L(u, \mathbf{c}) = \Psi_U(u, \mathbf{c}) = \infty$.

S3. The function $h(r, \mathbf{c})$ is strictly increasing. In such case there exist at most one solution $\Psi_U(u, \mathbf{c})$ which tends to infinity as $u \to \infty$ (if no solution exists we define $\Psi_U(u, \mathbf{c}) = \infty$). We set $\Psi_L(u, \mathbf{c}) = -\infty$.

Then, we obtain the final shape of our estimator

$$\widehat{z}_B(u) := \mathbb{P}\big(R < \Psi_L(u, \mathbf{c})\big)\, \mathbb{I}\big(\Psi_L(u, \mathbf{c}) > 0\big) + \mathbb{P}\big(R > \Psi_U(u, \mathbf{c})\big),$$

where the indicator function has been added since the random variable R should take values only on $(0, \infty)$. Algorithmically:

Conditional Monte Carlo Algorithm.

1. Simulate $\mathbf{C}$ uniformly over the unit spheroid in $\mathbb{R}^d$.
2. Find (numerically) the values $\psi_L = \Psi_L(u, \mathbf{C})$ and $\psi_U = \Psi(u, \mathbf{C})$ (see the discussion on the current page).
3. Return
$$\widehat{z}_1(u) := \mathbb{P}\big(R < \psi_L\big)\, \mathbb{I}\big(\psi_L > 0\big) + \mathbb{P}\big(R > \psi_U\big).$$

The efficiency of this estimator is established in the following Theorem whose proof can be found on page 66.

Theorem 4.2. *Let* $\log \mathbf{X}$ *have a elliptical distribution* $E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ *with* $\boldsymbol{\Sigma}$ *a nonsingular positive definite matrix,* $\mathbf{C}$ *a random vector uniformly distributed over the unit spheroid in* $\mathbb{R}^d$ *and* $\Psi_L(\cdot)$ *and* $\Psi_U(\cdot)$ *as defined in S1–S3. If*

$$\limsup_{u\to\infty} \frac{\mathbb{P}^2\left(R > \frac{\log u - \mu_1 + \sqrt{(\log u-\mu)^2 - 4\sigma_1}}{2\sigma_1}\right)}{\mathbb{P}^{2-\epsilon}(S_d > u)} < \infty, \qquad \epsilon > 0.$$

Then the estimator

$$\widehat{z}_1(u) := \mathbb{P}\big(R < \Psi_L(u, \mathbf{C})\big)\, \mathbb{I}\big(\Psi_L(u, \mathbf{C}) > 0\big) + \mathbb{P}\big(R > \Psi_U(u, \mathbf{C})\big)$$

is an unbiased and logarithmic efficient estimator of $\mathbb{P}(S_d > u)$*. Moreover, if the limit holds for* $\epsilon = 0$ *then the estimator has bounded relative error.*

The next corollary can be easily proved from the results of the last Theorem and it proof is given in on page 68.

Corollary 4.3. *Let* $\log \mathbf{X}$ *have a multivariate normal distribution* $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *with* $\boldsymbol{\Sigma}$ *a nonsingular positive definite matrix,* $\mathbf{C}$ *a random vector uniformly distributed over the unit spheroid in* $\mathbb{R}^d$ *and* $\Psi_L(\cdot)$ *and* $\Psi_U(\cdot)$ *as defined in S1–S3. Let*

$$\widehat{z}_1(u) := \mathbb{P}\big(\chi^2_d < \psi^2_L(u, \mathbf{C})\big)\, \mathbb{I}\big(\Psi_L(u, \mathbf{C}) > 0\big) + \mathbb{P}\big(\chi^2_d > \psi^2_U(u, \mathbf{C})\big)$$

where χ^2_d *is a chi-square distribution with* d *degrees of freedom. Then* $\widehat{z}_1(u)$ *is an unbiased and logarithmic efficient estimator of* $\mathbb{P}(S_d > u)$*.*

Usually, the problem of finding the solution of a fixed point problem of the type $h(r, \mathbf{c}) = u$ is an easy task. However, we need to implement an algorithm which systematically solves a random problem at the time for each replication. A careless implementation of this problem may yield to a slow algorithm or even wrong solutions, so we need to do some analysis of it. Recall that Newton-Raphson is a method for finding roots of a differentiable function which provide reliable results if the derivative of the function is not very close to zero in large regions (similar problems arise for other algorithms for finding roots). In fact, in our numerical implementations it has been observed cases where one of the solutions lie in a region where the derivative is close to zero and the methods implemented with arbitrary initial values become slow and deliver poor solutions of the problem. The solution is provide an initial value which is close to the real solution; generally speaking, a root-finding algorithm becomes faster and more accurate as the initial value is chosen closer to the real solution of the problem.

Remember that in the previous chapter we found that for the case of correlated lognormals it holds that $\mathbb{P}(S_d > u) = \mathbb{P}(M_d > u)(1 + o(1))$. This relation will not necessarily hold for the general case of log-elliptical random vectors, however it provides an excellent approximation for the problems we are interested here. For the general case it should be investigated whether or not this is an appropriated choice. So, our suggestion is to use the solutions involving the maximum as initial values. Let $\mathbf{c} \in \mathcal{S}_1(\mathbf{0})$ and define

$$h_i(r, \mathbf{c}) := \mathrm{e}^{r\alpha_i(\mathbf{c}) + \mu_i} := \mathrm{e}^{r(A_{i1}c_1 + \cdots + A_{in}c_d) + \mu_i}. \tag{4.3}$$

Hence, $h_i(r, \mathbf{c})$ is convex and monotone as function of r and therefore there exist at most one solution to the problem $h_i(r, \mathbf{c})$. We consider two possible scenarios:

S1'. The function $h_i(r, \mathbf{c})$ is decreasing. We define $\Upsilon_i^-(u, \mathbf{c})$ to be the solution of the problem $h_i(r, \mathbf{c}) = u$ (if no solution exists we let $\Upsilon_i^-(u, \mathbf{c}) = -\infty$). We define $\Upsilon_i^+(u, \mathbf{c}) = \infty$.

S2'. The function $h_i(r, \mathbf{c})$ is increasing. We define $\Upsilon_i^+(u, \mathbf{c})$ to be the solution of the problem $h_i(r, \mathbf{c}) = u$ (if no solution exists we let $\Upsilon_i^+(u, \mathbf{c}) = \infty$). We define $\Upsilon_i^-(r, \mathbf{c}) = \infty$.

By defining

$$\Upsilon_-(u, \mathbf{c}) := \max_{i=1,\ldots,d} \{\Upsilon_i^-(u, \mathbf{c})\}, \qquad \Upsilon_+(u, \mathbf{c}) := \min_{i=1,\ldots,d} \{\Upsilon_i^+(u, \mathbf{c})\} \tag{4.4}$$

The root-finding algorithm is given next.

Root-finding algorithm.

1. Given $h(r, \mathbf{c})$ determine the regions where the function is either increasing or decreasing (see scenarios S1-S3). An easy way to do this is to check the signs of the functions $\alpha_i(\mathbf{c})$ defined in (4.3).

2. If $h(r, \mathbf{c})$ is decreasing in some region use Newton-Raphson to find $\Psi_L(r, \mathbf{c})$ taking as initial value $\Upsilon_-(u, \mathbf{c})$. Otherwise set $\Psi_L(u, \mathbf{c}) = -\infty$.

3. If $h(r, \mathbf{c})$ is increasing in some region use Newton-Raphson to find $\Psi_U(r, \mathbf{c})$ taking as initial value $\Upsilon_+(u, \mathbf{c})$. Otherwise set $\Psi_U(u, \mathbf{c}) = \infty$.

4.2.1 Estimation of the Residual Probability via Conditional Monte Carlo

As in the previous chapter we can improve substantially our algorithm by using a stratification strategy as the one suggested in Juneja (2007). That is we can rewrite

$$\mathbb{P}(S_d > u) = \mathbb{P}(S_d > u, M_d < u) + \mathbb{P}(M_d \geq u).$$

Here we deal with the estimation of the residual probability $\mathbb{P}(S_d > u, M_d < u)$ via the conditional Monte Carlo method suggested before. From the discussion for the root-finding algorithm and the stochastic representation 4.1 it follows easily that

$$\widehat{z}_R(u) := \mathbb{P}\big(\Upsilon_-(u, \widetilde{\mathbf{Z}}) < R < \Psi_L(u, \widetilde{\mathbf{Z}})\big) + \mathbb{P}\big(\Psi_U(u, \widetilde{\mathbf{Z}}) < R < \Upsilon_+(u, \mathbf{C})\big)$$

is an unbiased estimator of the residual probability $\mathbb{P}(S_d > u, M_d < u)$.

In the following we will restrict ourselves to the case of correlated lognormals. In fact, we prove that the resulting estimator will have vanishing relative error. The more general case will depend on the construction of an efficient algorithm for $\mathbb{P}(M_d > u)$ and the generalization of the proof presented here. This is still work in progress.

The following Theorem, whose proof can be found on page 68, shows that the contribution to the variation coefficient of the estimator for the residual part will vanish as $u \to \infty$.

Theorem 4.4. *Let $(X_1, \ldots, X_d)$ have a $LN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution such that $\boldsymbol{\Sigma}$ is nonsingular positive definite matrix, $\mathbf{C}$ a random vector uniformly distributed over the unit spheroid $\mathcal{S}_1(\mathbf{0})$, $\Psi_L(\cdot)$, $\Psi_U(\cdot)$ defined as in S1–S3 and $\Upsilon_-(\cdot)$, $\Upsilon_+(\cdot)$ defined as in* (4.4). *Then*

$$\widehat{z}_R(u) := \mathbb{P}\Big(\Upsilon_-(u, \mathbf{C}) < \sqrt{\chi^2_d} < \Psi_L(u, \mathbf{C})\Big) + \mathbb{P}\big(\Psi_U(u, \mathbf{C}) < \sqrt{\chi^2_d} < \Upsilon_+(u, \mathbf{C})\big)$$

is an unbiased of the residual probability $\mathbb{P}(S_d > u, M_d < u)$ and is such that

$$\limsup_{u \to \infty} \frac{\mathbb{E}\ \widehat{z}^2_R(u)}{\mathbb{P}^2(S_d > u)} \ = \ 0\,.$$

Hence, by using an estimator as the one suggested in Adler et al. (2008) for estimating the second term $\mathbb{P}(M_d > u)$ we obtain an algorithm with *vanishing relative error*.

4.3 Concluding Remarks

4.3.1 Numerical Examples

We implemented the estimators described in Chapters 3 and 4 in three examples corresponding to low (Example D.4 on page 107), medium (Example D.5 on page 108) and high correlations (Example D.6 on page 109). A more detailed description of the setting used in the examples can be found on page 106. The estimators analyzed were: the Importance Sampling via the Scaled Variance (**SV**) and its Cross-Entropy version (**SVCE**); the dependent version of the Asmussen-Kroese algorithm (**AK**) and the conditional algorithm based on the radial random variable (**RE**). The conclusion is that the four algorithms analyzed show excellent numerical results in our experiments. However:

The estimator **AK** appears to be favored over the other proposed algorithms. However, Juneja (2007) indicated that this algorithm is inefficient for estimating the probability $\mathbb{P}(S_n > u, M_n < u)$ and in consequence the estimator of the variance shown in the numerical examples might not be reliable. Note for example in the left graph of figure D.6 on page 109 that the non stratified **AK** algorithm shows the largest Variation Coefficient. However, after an entirely similar improvement via the stratification in all algorithms, the **AK** shows the smallest Variation Coefficient. A possibility could be to analyze this problem using the efficiency concepts suggested in L'Ecuyer et al. (2008).

Leaving aside the estimator **AK**, the **RE** estimator outperform significatively the other two algorithms in terms of the Variation Coefficient. However, since it requires a find-rooting procedure for each iteration it is not that efficient in terms of Time-Relative Error – a more fair measure of the performance of an algorithm (cf. Asmussen and Glynn, 2007). Moreover, the Cross-entropy based estimator **SVCE** provides the best results for estimating the probability $\mathbb{P}(S_n > u, M_n < u)$.

It also should be noted that our initial guess of the value θ in the scaling variance algorithm, chosen as $\mathbb{E}[S_n] = u$, gives excellent numerical results. This a fundamental feature for the Cross-Entropy estimator since it provides a close initial point to the optimal parameter. Although we have used a very demanding criterion to stop the iteration in the Cross-Entropy, it was observed that few iterations are enough. This is a relevant observation since it translates in an improvement in the Time-Relative Error criterion.

The conclusion here is that the Stratified Importance Sampling via a Scaled Variance plus a few iterations of the Cross-Entropy method provides excellent results in every regime. In fact, the modifications – stratification and Cross-Entropy – are cheap in terms of cpu time and implementation effort. However, the value of the conditional algorithm based on the radial random variable lies on its stronger theoretical properties and the flexibility to be generalized to a wider class of dependent distributions.

Notes and Comments

Glasserman et al. (2000) developed an importance sampling estimator based on a quadratic approximation which can be used to estimate the tail probability of a sum of correlated lognormal random variables. In Rojas-Nandayapa (2006), some numerical comparisons were performed against some of the algorithms contained in this dissertation. Also Glasserman et al. (2002) develop importance sampling estimators based on a quadratic approximation for calculating the Value-At-Risk when the underlying risk factors follow a t distribution. More details of the quadratic approximation can be found in Glasserman (2000).

A somehow related model was studied in Glasserman and Li (2005) where an efficient simulation method based on importance sampling was proposed to estimate the probability of large losses resulting from a number of defaults in a portfolio. The model analyzed there was the so called *normal copula model*. However, it has been argued that this model fails to capture some of the behavior observed in financial variables. Also recent work shows that the Student t-copula provides a better fit (see e.g. Bassamboo et al., 2008, and references therein). In recent work, Bassamboo et al. (2008); Chan and Kroese (2008) suggested algorithms for the t based model which are claimed to be efficient. The first set of estimators is based on importance sampling strategies while the second on conditional Monte Carlo. Moreover, Chan and Kroese (2008) also explore the case of a skew t-copula.

4.4 Proofs

Proof of Theorem 4.2. Observe that there exists $u_1 > 0$ such that for all $u > u_1$ we get

$$\mathbb{I}\big(\Psi_L(u, \mathbf{c}) > 0\big) = 0.$$

This follows from the observation $\Psi_L(u, \mathbf{c}) \to -\infty$ made in S1 and S2. Hence, for $u > u_0$ we get

$$\mathbb{E}\left[\widehat{z}_1^2(u)\right] = \mathbb{E}\left[\mathbb{P}^2(R > \Psi_U(u, \mathbf{C}))\right].$$

Moreover, it is observed that the solution $\Psi_U(u, \mathbf{c}) < \infty$ exists only for those $\mathbf{c} \in \mathcal{S}_1(\mathbf{0})$ such that $h(r, \mathbf{c})$ is ultimately increasing in r (otherwise the probability is 0 since we have defined $\Psi_U(u, \mathbf{c}) = \infty$). From this observation and Lemma 4.5 on the facing page we deduce that there exists a value u_0 (which depends on r_0) such that for all $u > u_0$ the following inequality is fulfilled

$$\mathbb{E}\left[\mathbb{P}^2(h(R, \mathbf{C}) > u)|\mathbf{C}\right] \leq \mathbb{P}^2(\mathrm{e}^{R\sigma_1 + \mu_1 + 1/R} > u).$$

Hence, we can easily solve the equation $\mathrm{e}^{R\sigma_1+\mu_1+1/R}=u$ for R and use the larger root for making to obtain

$$\limsup_{u\to\infty}\frac{\mathbb{E}\left[\widehat{z}_1^2(u)\right]}{\mathbb{P}^{2-\epsilon}(S_d>u)}=\limsup_{u\to\infty}\frac{\mathbb{P}^2\left(R>\frac{\log u-\mu_1+\sqrt{(\log u-\mu)^2-4\sigma_1}}{2\sigma_1}\right)}{\mathbb{P}^{2-\epsilon}(S_d>u)}.$$

By hypothesis the last lim sup is bounded, so the result of the Theorem follows. □

Lemma 4.5. *Let $h(r,\mathbf{z})$ be defined as in (4.2). Then there exists a value $r_0>0$ such that for all $r>r_0$ the following inequality is satisfied*

$$h(r,\mathbf{c})\leq \mathrm{e}^{r\sigma_1+\mu_1+1/r},\qquad \forall\,\mathbf{c}\in\mathcal{S}_1(\mathbf{0})$$

Proof. Our aim is to prove that there exist $0<\delta<1$ such that

$$\max_{\|\mathbf{y}\|=1}h(r,\mathbf{c})<\left(\mathrm{e}^{r\sigma_1+\mu_1}+e^{r\sigma_1\delta+c_1}\right),\qquad r>0,\tag{4.5}$$

with c_1 a fixed constant. To simplify some of the notation used in this proof we define

$$h_i(r,\mathbf{c}):=\mathrm{e}^{r\alpha_i(\mathbf{c})+\mu_i}:=\mathrm{e}^{r(A_{i1}c_1+\cdots+A_{in}c_d)+\mu_i},\qquad \forall\,\mathbf{c}\in\mathcal{S}_1(\mathbf{0}).$$

Using Lagrangian multipliers we can easily verify that

$$\max_{\mathbf{c}\in\mathcal{S}_1(\mathbf{0})}\{h_i(r,\mathbf{c})\}=\mathrm{e}^{r\sqrt{A_{i1}^2+\cdots+A_{in}^2}+\mu_i}\tag{4.6}$$

and the solution of every individual maximization problem is given by the vector

$$\mathbf{c}_i^*:=\operatorname*{argmax}_{\mathbf{c}\in\mathcal{S}_1(\mathbf{0})}\{h_i(r,\mathbf{c})\}=\frac{\left(A_{i1},\ldots,A_{in}\right)}{\sigma_i}.$$

That is the i-th row vector of the matrix $\mathbf{A}$ divided by σ_i. Moreover, the solution is unique and do not depend on r. If we let $K:=\#\{i:\sigma_1=\sigma_i\}$ it follows that

$$\max_{\mathbf{c}\in\mathcal{S}_1(\mathbf{0})}\{h(r,\mathbf{c})\}\leq\max_{\mathbf{c}\in\mathcal{S}_1(\mathbf{0})}\{h_1(r,\mathbf{c})+\cdots+h_K(r,\mathbf{c})\}+\mathrm{e}^{r\sigma_{K+1}+\mu_{K+1}}+\cdots+\mathrm{e}^{r\sigma_d+\mu_d}.$$

In fact, a single vector $\mathbf{c}\in\mathcal{S}_1(\mathbf{0})$ cannot be the solution of more than one of the first K maximization problems given in (4.6). The reason lies in the nonsingularity of the matrix $\mathbf{\Sigma}$ which implies that we cannot have $(A_{i\ell})_{\ell=1,\ldots,d}=(A_{j\ell})_{\ell=1,\ldots,d}$ with $i\neq j$ for $i,j<K$. Hence, we can choose $\epsilon>0$ small enough in such way that the following sets are disjoint

$$\mathcal{B}_\epsilon(\mathbf{c}_i^*):=\{\mathbf{c}:\|\mathbf{c}-\mathbf{c}_i^*\|\leq\epsilon\},\qquad i=1,\ldots,K,$$

(balls of radius ϵ centered in $\mathbf{c}_i^*$). Furthermore, since the solutions are unique and the functions $h_i(r,\mathbf{c})$ are continuous we can take a value α^* close enough to σ_1 such that

$$\max_{i=1,\ldots,K}\ \sup\ \{h_i(r,\mathbf{c}):\mathbf{c}\in\mathcal{B}_\epsilon(\mathbf{c}_i^*)^c\cap\mathcal{S}_1(\mathbf{0})\}<\mathrm{e}^{r\alpha^*+\mu_1},\qquad\forall r>0.$$

Since an arbitrary vector $\mathbf{c}$ can be at most in a single set $\mathcal{B}_\epsilon(\mathbf{c}_i^*)$ it follows that

$$\begin{aligned}\sup_{\mathbf{c}\in\mathcal{S}_1(\mathbf{0})}\{h(r,\mathbf{c})\}&\leq\mathrm{e}^{r\sigma_1+\mu_1}+(K-1)\mathrm{e}^{r\alpha^*+\mu_1}+\mathrm{e}^{r\sigma_{K+1}+\mu_{K+1}}+\cdots+\mathrm{e}^{r\sigma_d+\mu_d}\\&\leq\mathrm{e}^{r\sigma_1+\mu_1}+(K-1)\mathrm{e}^{r\delta+\mu_1}+\mathrm{e}^{r\delta+\mu_{K+1}}+\cdots+\mathrm{e}^{r\delta+\mu_d}\\&=\mathrm{e}^{r\sigma_1+\mu_1}+\mathrm{e}^{r\sigma_1\delta+k_1},\end{aligned}$$

with $\max\{\alpha^*, \sigma_{K+1}, \ldots, \sigma_d\}/\sigma_1 < \delta < 1$ and k_1 the appropriate constant. We rewrite this bound as

$$\exp\Big\{r\sigma_1 + \mu_1 + \log\Big(1 + \mathrm{e}^{-r\sigma_1(1-\delta)+c_1-\mu_1}\Big)\Big\}.$$

The proof is completed by noting that

$$\lim_{r\to\infty} \frac{\log\big(1 + \mathrm{e}^{-r\sigma_1(1-\delta)+k_1-\mu_1}\big)}{1/r} = \lim_{r\to\infty} \frac{r^2\sigma_1(1-\delta)\ \mathrm{e}^{-r\sigma_1(1-\delta)+k_i-\mu_1}}{1 + \mathrm{e}^{-r\sigma_1(1-\delta)+k_1-\mu_1}} = 0.$$

We have used L'Hopital Theorem to solve the limit. □

Proof of Corollary 4.3. The result will follow immediately from Theorem 4.2 on page 63 if we prove that the following limit remains bounded for all $\epsilon > 0$

$$\begin{aligned}
\limsup_{u\to\infty} \frac{\mathbb{E}\left[\widehat{z}_1^2(u)\right]}{\mathbb{P}^{2-\epsilon}(S_d > u)} &= \limsup_{u\to\infty} \frac{\mathbb{P}^2\Big(R > \frac{\log u - \mu_1 + \sqrt{(\log u-\mu)^2 - 4\sigma_1}}{2\sigma_1}\Big)}{\big[m_d \overline{F}_{\mu,\sigma^2}(u)\big]^{2-\epsilon}} \\
&= \limsup_{u\to\infty} \frac{\mathbb{P}^2\Big(R^2 > \frac{(\log u-\mu)^2}{\sigma_1^2}\Big)}{\big[m_d \overline{F}_{\mu,\sigma^2}(u)\big]^{2-\epsilon}} \\
&= \limsup_{u\to\infty} k\, \frac{\big(\log u - \mu\big)^{d-2} \exp\Big\{-\frac{\left(\log u-\mu\right)^2}{\sigma_1^2}\Big\}}{\exp\Big\{-\frac{\left(\log u-\mu\right)^2}{\sigma_1^2}\big(1-\epsilon/2\big)\Big\}/u^{2-\epsilon}}.
\end{aligned}$$

All constants have been grouped in k. Hence, it follows that the last limit goes to 0 for all $\epsilon > 0$ showing that the estimator is logarithmically efficient. □

Proof of Theorem 4.4. The notation will become rather complicated in parts. So to simplify things we take $\mathbf{c} \in \mathcal{S}_1(\mathbf{0})$ and define

$$h_i(r, \mathbf{c}) := \mathrm{e}^{r\alpha_i(\mathbf{c})+\mu_i} = \mathrm{e}^{r(A_{i1}c_1+\cdots+A_{in}c_d)+\mu_i} \qquad H(r, \mathbf{C}) := \max_{i=1,\ldots,d} \big\{h_i(r, \mathbf{C})\big\}$$

So, we will start by expanding the second moment of the estimator $\widehat{z}_{RB}^2(u)$ as follows

$$\begin{aligned}
\mathbb{E}\left[\widehat{z}_1^2(u)\right] &= \mathbb{E}\ \mathbb{P}^2\Big(h(R, \mathbf{C}) > u,\ H(R, \mathbf{C}) < u \,\Big|\, \mathbf{C}\Big) \\
&= \mathbb{E}\left(\sum_{k=1}^{d} \mathbb{P}\Big(h(R, \mathbf{C}) > u,\, h_k(R, \mathbf{C}) = H(R, \widetilde{\mathbf{C}}) < u \,\Big|\, \mathbf{C}\Big)\right)^2.
\end{aligned}$$

We choose β, γ and k as in Lemma 4.6 and use the same argument as in (3.8) to obtain the following upper bound for the previous expression

$$\begin{aligned}
\mathbb{E}\Bigg[\Bigg(\sum_{k=1}^{d} &\mathbb{P}\Big(h(R, \mathbf{C}) - h_k(R, \mathbf{C}) > u^\beta,\, h_k(R, \mathbf{C}) > u^\beta/d \,\Big|\, \mathbf{C}\Big) \\
&+ \mathbb{P}\Big(u - u^\beta < h_k(R, \mathbf{C}) < u | \mathbf{C}\Big)\Bigg)^2\Bigg].
\end{aligned}$$

From Lemmas 4.6 and 4.7 we have that for all $u > (u_0 \vee u_0')$ the last expectation is bounded above by

$$d^2\left[\mathbb{P}\left(\chi_k^2 > \frac{(\log u - k)^2}{\gamma}\right) + \mathbb{P}\left(\frac{(\log(u-u^\beta)-\mu_k)^2}{\sigma_k^2} < \chi_k^2 < \frac{(\log u - \mu_k)^2}{\sigma_k^2}\right)\right]^2.$$

Notice that we have removed the expectation since these results apply for all $\mathbf{z} \in \mathcal{S}_1(\mathbf{0})$. Hence, it just remains to prove that

$$\limsup_{u\to\infty} \frac{\mathbb{P}\left(\chi_k^2 > \frac{(\log u-k)^2}{\gamma}\right) + \mathbb{P}\left(\frac{(\log(u-u^\beta)-\mu_k)^2}{\sigma_k^2} < \chi_k^2 < \frac{(\log u-\mu_k)^2}{\sigma_k^2}\right)}{\overline{F}_{\mu,\sigma^2}(u)} = 0$$

(notice that we have taken the square root of the relative error). For the first term we use L'Hopital to get

$$\limsup_{u\to\infty} \frac{\mathbb{P}\left(\chi_k^2 > \frac{(\log u-k)^2}{\gamma}\right)}{\overline{F}_{\mu,\sigma^2}(u)} = \limsup_{u\to\infty} \frac{k_0(\log u)^{d-1}\exp\left\{-\frac{(\log u-k)^2}{2\gamma^2}\right\}}{\exp\left\{-\frac{(\log u-\mu)^2}{2\sigma^2}\right\}} = 0$$

where we have grouped all constants in k_0 and we have used the fact that $\gamma < \sigma$. Similarly, we obtain for the second term that

$$\begin{aligned}\limsup_{u\to\infty} &\frac{\mathbb{P}\left(\frac{(\log(u-u^\beta)-\mu_k)^2}{\sigma_k^2} < \chi_d^2 < \frac{(\log u-\mu_k)^2}{\sigma_k^2}\right)}{\overline{F}_{\mu,\sigma^2}(u)} \\ &= \limsup_{u\to\infty}\ k_1(\log u)^{d-1}\ \frac{\exp\left\{-\frac{(\log u-\mu_k+\log(1-u^{\beta-1}))^2}{2\sigma_k^2}\right\} - \exp\left\{-\frac{(\log u-\mu_k)^2}{2\sigma_k^2}\right\}}{\exp\left\{-\frac{(\log u-\mu)^2}{2\sigma^2}\right\}}.\end{aligned}$$

Since we have chosen $0 < \beta < 1$ we can reduce the last expression to

$$\limsup_{u\to\infty}\ k_1(\log u)^{d-1}\ \frac{\exp\left\{-\frac{(\log u-\mu_k)^2}{2\sigma_k^2}\right\}}{\exp\left\{-\frac{(\log u-\mu)^2}{2\sigma^2}\right\}}\left[\exp\left\{-\frac{\log u\log(1-u^{\beta-1})}{\sigma_k^2}\right\} - 1\right]. \quad (4.7)$$

Using Taylor expansions of $\log(1-x)$ and e^x around $x = 0$ we deduce that

$$\lim_{u\to\infty} \frac{\exp\left\{-\frac{\log u\log(1-u^{\beta-1})}{\sigma_k^2}\right\} - 1}{\frac{\log u}{u^{1-\beta}}} = \lim_{u\to\infty} \frac{\exp\left\{\frac{\log u}{u^{1-\beta}\sigma_k^2}\right\} - 1}{\frac{\log u}{u^{1-\beta}}} = 1$$

From this result it is easily seen that the limit (4.7) goes to 0 and the result follows. □

Lemma 4.6. *Under the hypothesis of Proposition 4.4 there exists values $\beta \in (0,1)$, $\gamma \in (0, \sigma_1)$, $u_0 > 0$ and $k \in \mathbb{R}$ such that the following set of inequalities is satisfied*

$$\mathbb{P}\big(h(R,\mathbf{c}) - h_k(R,\mathbf{c}) > u^\beta, h_k(R,\mathbf{c}) > u^\beta/d\big) < \mathbb{P}\left(\chi_d^2 > \frac{(\log u - k)^2}{\gamma^2}\right), \quad k = 1,\dots,d$$

for all $\mathbf{c} \in \mathcal{S}_1(\mathbf{0})$ and $u > u_0$.

Proof. Observe that if $\mathbf{c} \in \mathcal{S}_1(\mathbf{0})$ is such that either:

i) $\alpha_k(\mathbf{c}) \leq 0$. Therefore $h_k(r, \mathbf{c})$ is a decreasing function in r.

ii) $\alpha_\ell(\mathbf{c}) < 0$ for all $\ell \neq k$. Therefore $h(r, \mathbf{z}) - h_k(r, \mathbf{z})$ is a decreasing function in r.

Hence, we conclude that in any of the cases above there exists $u_k > 0$ which does not depend on $\mathbf{c}$ and such that for all $u > u_k$ it holds that

$$\mathbb{P}\big(h(R, \mathbf{c}) - h_k(R, \mathbf{c}) > u^\beta, h_k(R, \mathbf{c}) > u^\beta/d\big) = 0$$

and the inequality is trivially satisfied. Thus, we consider only those vectors $\mathbf{c} \in \mathcal{S}_1(\mathbf{0})$ such that $\alpha_k(\mathbf{c}) > 0$ and $\alpha_\ell(\mathbf{c}) > 0$ for at least one $\ell \neq k$. Let $K := \#\{i : \sigma_i = \sigma_1\}$. We verify

$$\max_{\mathbf{c}\in\mathcal{S}_1(\mathbf{0})} \{h_i(r, \mathbf{c})\} = \mathrm{e}^{r\sigma_i + \mu_i}, \qquad \mathbf{c}_i^* := \operatorname*{argmax}_{\mathbf{c}\in\mathcal{S}_1(\mathbf{0})} \{h_i(r, \mathbf{c})\} = \frac{\big(A_{i1}, \ldots, A_{in}\big)}{\sigma_i},$$

where the $\mathbf{c}_i^*$'s are unique and do not depend on r. Moreover, we can choose $\epsilon > 0$ and $\alpha^* \in (0, \sigma_1)$ such that the sets

$$\mathcal{B}_\epsilon(\mathbf{c}_i^*) := \{\mathbf{c} : \|\mathbf{c} - \mathbf{c}_i^*\| \leq \epsilon\}, \qquad i = 1, \ldots, K,$$

are disjoint and

$$\max_{i=1,\ldots,K} \sup \{h_i(r, \mathbf{c}) : \mathbf{c} \in \mathcal{B}_\epsilon(\mathbf{c}_i^*)^c \cap \mathcal{S}_1(\mathbf{0})\} < \mathrm{e}^{r\alpha^* + \mu_1} \tag{4.8}$$

for all $r > 0$ (for details see the proof of Lemma 4.5). We consider the three following cases.

Case 1. Suppose that $k > K$. Hence it follows that

$$\begin{aligned}
&\mathbb{P}\big(h(R, \mathbf{z}) - h_k(R, \mathbf{c}) > u^{\beta_k}, h_k(R, \mathbf{c}) > u^{\beta_k}/d\big) \\
&\quad \leq \mathbb{P}\big(h_k(R, \mathbf{c}) > u^{\beta_k}/d\big) = \mathbb{P}\big(\mathrm{e}^{R\alpha_k(\mathbf{c})/\beta_k + (\mu_k + \log d)/\beta_k} > u\big) \\
&\quad \leq \mathbb{P}\big(\mathrm{e}^{R\sigma_k/\beta_k + (\mu_k + \log d)/\beta_k} > u\big) = \mathbb{P}\big(\mathrm{e}^{R\gamma_k + c_k} > u\big),
\end{aligned}$$

where $\gamma_k := \sigma_k/\beta$ and $c_k := (\mu_k + \log d)/\beta$. So, we choose β_k such that $\sigma_k/\sigma_1 < \beta < 1$ to ensure $\gamma_k < \sigma_1$.

Case 2. Suppose that $k \leq K$ and $\mathbf{c} \in \mathcal{B}_\epsilon^c(\mathbf{c}_k^*) \cap \mathcal{S}_1(\mathbf{0})$. Hence, it follows by (4.8) that

$$\begin{aligned}
&\mathbb{P}\big(h(R, \mathbf{c}) - h_k(R, \mathbf{c}) > u^{\beta_k}, h_k(R, \mathbf{zy}) > u^{\beta_k}/d\big) \\
&\quad \leq \mathbb{P}\big(h_k(R, \mathbf{c}) > u^{\beta_k}/d\big) = \mathbb{P}\big(\mathrm{e}^{R\alpha_k(\mathbf{c})/\beta_k + (\mu_k + \log d)/\beta} > u\big) \\
&\quad \leq \mathbb{P}\big(\mathrm{e}^{R\alpha^*/\beta_k + (\mu_k + \log d)/\beta} > u\big) = \mathbb{P}\big(\mathrm{e}^{R\gamma_k + c_k} > u\big),
\end{aligned}$$

where $\gamma_k := \alpha^*/\beta$ and $c_k := (\mu_k + \log d)/\beta_k$. So, we choose β_k such that $\alpha^*/\sigma_1 < \beta < 1$ to ensure $\gamma_k < \sigma_1$.

Case 3. Suppose that $k \leq K$ and $\mathbf{c} \in \mathcal{B}_\epsilon(\mathbf{c}_k^*) \cap \mathcal{S}_1(\mathbf{0})$. We define

$$\delta_\sigma := \max\{\alpha^*, \sigma_{K+1}, \ldots, \sigma_d\} \qquad \text{and} \qquad \delta_\mu := \max\{\mu_1, \ldots, \mu_d\}.$$

Hence it follows by (4.8) that

$$\begin{aligned}
&\mathbb{P}\big(h(R,\mathbf{c}) - h_k(R,\mathbf{c}) > u^{\beta'_k}, h_k(R,\mathbf{c}) > u^{\beta'_k}/d\big) \\
&\quad \leq \mathbb{P}\big(h(R,\mathbf{c}) - h_k(R,\mathbf{c}) > u^{\beta'_k}\big) \leq \mathbb{P}\big(d^{(1/\beta'_k)} \mathrm{e}^{R\delta_\sigma/\beta'_k + \delta_\mu/\beta'_k} > u\big) \\
&\quad = \mathbb{P}\big(\mathrm{e}^{R\gamma'_k + c'_k} > u\big),
\end{aligned}$$

where $\gamma'_k = \delta_\sigma/\beta'_k$ and $c_k = (\delta_\mu + \log d)/\beta'_k$. So, we choose β'_k such that $\delta_\sigma/\sigma_1 < \beta'_k < 1$ to ensure $\gamma'_k < \sigma_1$. The result follows easily by choosing

$$\begin{aligned}
u_0 &:= \max\{u_1, \ldots, u_d\} \\
\beta &:= \max\{\beta_1, \ldots, \beta_d, \beta'_1, \ldots, \beta'_K\} \\
\gamma &:= \max\{\gamma_1, \ldots, \gamma_d, \gamma'_1, \ldots, \gamma'_K\} \\
c &:= \max\{c_1, \ldots, c_d, c'_1, \ldots, c'_K\}.
\end{aligned}$$

□

Lemma 4.7. *Under the hypothesis of Proposition 4.4 there exists $u'_0 > 0$ and $c' \in \mathbb{R}$ such that for all $u > u'_0$ it holds true that*

$$\mathbb{P}\Big(u - u^\beta < h_k(R,\mathbf{c}) < u\Big) < \mathbb{P}\bigg(\frac{(\log(u - u^\beta) - c')^2}{\sigma_1^2} < \chi_d^2 < \frac{(\log u - c')^2}{\sigma_1^2}\bigg)$$

for any $\beta \in (0,1)$ and all $\mathbf{c} \in \mathcal{S}_1(\mathbf{0})$.

Proof. Observe that for all $\mathbf{c}$ such that $\alpha_k(\mathbf{c}) \leq 0$ there exists a fixed value u'_k which does not depend on $\mathbf{c}$ and such that for all $u > u'_k$ the probability on the l.h.s. of the expression in Lemma (4.6) is 0 and the inequality is trivially satisfied. Hence, we consider only those vectors $\mathbf{c} \in \mathcal{S}_1(\mathbf{0})$ such that $\alpha_k(\mathbf{c}) > 0$.

Next observe that

$$\mathbb{P}\Big(u - u^\beta < h_k(R,\mathbf{c}) < u\Big) < \mathbb{P}\bigg(\Big(\frac{\log(u - u^\beta) - \mu_k}{\alpha_k(\mathbf{c})}\Big)^2 < \chi_d^2 < \Big(\frac{\log u - \mu_k}{\alpha_k(\mathbf{c})}\Big)^2\bigg).$$

To simplify notation we let $g_1 = (\log(u - u^\beta) - \mu_k)^2$, $g_2 = (\log u - \mu_k)^2$ and $\alpha := \alpha_k^2(\mathbf{c})$ to obtain

$$\mathbb{P}\bigg(\frac{g_1}{\alpha} < \chi_d^2 < \frac{g_2}{\alpha}\bigg).$$

So, we just need to prove that for large u the last expression is increasing in α. So, the first derivative with respect to α will be positive if and only if

$$\frac{g_1}{\alpha^2} f_{\chi_d^2}\Big(\frac{g_1}{\alpha}\Big) > \frac{g_2}{\alpha^2} f_{\chi_d^2}\Big(\frac{g_2}{\alpha}\Big).$$

Note that we can complete the density of a χ_{d+2}^2 random variable on both sides

$$f_{\chi_{d+2}^2}\Big(\frac{g_1}{\alpha}\Big) > f_{\chi_{d+2}^2}\Big(\frac{g_2}{\alpha}\Big),$$

and it is clear that this inequality holds for $0 \ll g_1 < g_2$ and all $\alpha > 0$ or equivalently, for all $u > u''_k$ for some $u''_k > 0$. Since $\sigma_k \geq \alpha_k(\mathbf{y})$ for all $\mathbf{y} \in \mathcal{S}_1(\mathbf{0})$ the result of the Lemma follows by taking

$$u'_0 := \max\{u'_1, \ldots, u'_d, u''_1, \ldots, u''_d\}.$$

□

Chapter 5

Small Tail Probabilities of Sums of Lognormals

In this chapter we study the tail probability $\mathbb{P}(S_n < nx)$ as $x \to 0$ in the case of independent lognormal random variables.

The results in this chapter will be centered around the existence of the *Laplace-Stieltjes transform* of a lognormal distribution function for negative values of the argument. In fact, the implementation of a importance sampling algorithm using the optimal change of measure will be possible due to this feature (cf. Siegmund, 1976; Bucklew et al., 1990; Sadowsky and Bucklew, 1990; Sadowsky, 1996). Moreover, some results form *large deviations* and *saddlepoint approximations* which make use of the Laplace transform–moment generating function will be available for this tail probability. However, the main difficulty lies in the calculation of the moment generating function itself. An analytic solution is not available, so it will be necessary to construct some approximations.

The rest of the chapter is divided as follows. In the Section 5.1 we study the Laplace transform of a lognormal random variable. We will introduce some concepts and give a large deviations-kind result as a first approximation to the tail probability $\mathbb{P}(S_n < nx)$. Then we illustrate how this result can be used to design efficient Monte Carlo algorithms based on importance sampling ideas. Later, based on a saddlepoint approximation we will construct an approximation for the Laplace transform of a lognormal random variable as well as for the moments of the distribution generated by the exponential change of measure.

In the Section 5.2 we propose two importance sampling algorithms for estimating the probability $\mathbb{P}(S_n < nx)$ as $x \to 0$. The first of them will be based in the optimal exponential change of measure where we use the approximations obtained in the previous section. From a theoretical point of view this algorithm should be efficient, however it will be shown that the final estimator depends directly on approximations so the resulting algorithm will turn to be biased. For the second algorithm we derive the parameters of a lognormal distribution in order to approximate the distribution generated by the optimal exponential change of measure and use it to implement an importance sampling algorithm. Finally, the two algorithms are combined with a stratification strategy which can produce theoretical efficient algorithms. Finally, we conduct an empirical test of the two algorithms.

5.1 The Laplace Transform of a Lognormal Random Variable

Let F be the distribution function of a random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the *Laplace-Stieltjes transform* of F is defined as

$$\mathcal{L}^* F(\theta) = \mathbb{E}\left[\mathrm{e}^{\theta X}\right] = \int \mathrm{e}^{\theta x} F(dx).$$

Note that the previous definition correspond in fact to the moment generating function and not to the *Laplace-Stieltjes* transform itself. However, with some slightly abuse of notation we will suppress the negative sign for computational convenience, adopt the notation $\varphi(\theta)$ and for the rest of this chapter we will refer to it as the *Laplace-Stieltjes* transform of the distribution F.

The *domain of the transform* is defined as $\Theta = \{\theta \in \mathbb{R} : \varphi(\theta) < \infty\}$ and the *cumulant transform* as $\kappa(\theta) = \log \varphi(\theta)$. Then for each $\theta \in \Theta$ we can define a new distribution function which is as follows

$$F_\theta(dx) := \mathrm{e}^{\theta x - \kappa(\theta)} F(dx), \qquad \theta \in \Theta.$$

This set of distributions is known as the *exponential family generated by* F and its members will be denoted with F_θ. The following, which is a *large deviations*-kind result, will be useful for the construction of the algorithms (see for instance Dembo and Zeitouni, 1998).

Proposition 5.1. *Let X_1, X_2 be a sequence of i.i.d. random variable with second moment finite, $x \in \mathbb{R}$ such that $x < \mathbb{E}(X)$ and $\theta \in \mathbb{R}$ the solution of $x = \mathbb{E}_\theta[X]$ (assuming that it exists). Then*

$$\mathbb{P}(S_n < nx) \sim \frac{1}{|\theta|\sqrt{2\pi\sigma_\theta^2 n}} \mathrm{e}^{-nx\theta + n\kappa(\theta)}, \qquad n \to \infty.$$

Usually, this type of asymptotic result are stated for the probabilities $\mathbb{P}(S_n > nx)$. The proof of this Proposition follows entirely in a similar way (see for instance Asmussen, 2003). The following heuristic argument is useful for illustrating that an *exponential distribution generated by* F is appropriate distribution for rare event simulation. Recall that the distribution associated to the *zero-variance estimator* is obtained by taking the original measure and conditioning with respect to the rare event $\{S_n < nx\}$. So, for a sequence $X_1, \ldots, X_n$ of nonnegative i.i.d. random variables we will have

$$\begin{aligned}\mathbb{P}(X_n \in dx | S_n < nx) &= \frac{\mathbb{P}(S_n \leq nx | X_n \in dx)\mathbb{P}(X_n \in dx)}{\mathbb{P}(S_n < nx)} \\ &= \frac{\mathbb{P}(S_{n-1} \leq (n-1)x)\mathbb{P}(X_n \in dx)}{\mathbb{P}(S_n < nx)}.\end{aligned}$$

So if $x \leq \mathbb{E}[X_i]$, the X_i's have second moment finite and θ is the solution of $\mathbb{E}_\theta[X] = x$, then the hypothesis of the Proposition (5.1) are satisfied and we will have

$$\mathbb{P}(X_n \in dx | S_n < nx) \approx \mathrm{e}^{x\theta - \kappa(\theta)} \mathbb{P}(X_n \in dx), \qquad n \to \infty.$$

In fact, this says that the *zero-variance* importance sampling converge to F_θ which motivates its use. Asmussen and Rubinstein (1995) proved that an importance sampling

distribution taken from the *exponential family generated by F* (also referred as *exponential change of measure* or *exponential twisted distribution*) will be logarithmically efficient if and only if θ is the solution of $\mathbb{E}_\theta X = x$, that is, the expectation take with respect to the probability measure induced by F_θ.

However, the implementation of this importance sampling algorithm depends upon an explicit expression of

$$\int_0^\infty \mathrm{e}^{-\theta x} F(dx),$$

which is clearly not available in closed analytical form. Even in the case that we could approximate precisely the solution θ of $\mathbb{E}_\theta[X] = x$ and the value of $\kappa(\theta)$ we would obtain a biased estimator due to the approximations. We provide an alternative method which consists in finding a lognormal distribution which approximates precisely the optimal exponential change of measure described above. However, the result in Asmussen and Rubinstein (1995) exclude it from being efficient.

The approximation of the Laplace transform of a lognormal distribution we propose here makes use a saddlepoint approximation and exploits the properties of the function LambertW – also called Omega function (cf. Corless et al., 1996). This function is defined as the multivalued inverse of the function $f(W) = W\mathrm{e}^W$, so it satisfies the relation

$$\mathrm{LambertW}(x)\ \mathrm{e}^{\mathrm{LambertW}(x)} = x, \qquad x \in \mathbb{C}. \tag{5.1}$$

In particular, if x is real the function $\mathrm{LambertW}(x)$ is defined for $x \in [-\mathrm{e}^{-1}, \infty)$ with two possible solutions in $[-\mathrm{e}^{-1}, 0)$ and one solution in $[0, \infty)$ defining two real branches. The *principal branch* is defined in $[-\mathrm{e}^{=1}, \infty)$ and satisfies $\mathrm{LambertW}(x) \geq -1$. The second branch will be denoted $\mathrm{LambertW}_{-1}(x)$ is defined in the interval $[-\mathrm{e}^{-1}, 0)$ and satisfies $\mathrm{LambertW}_{-1}(x) \leq -1$. For the remaining of the chapter we will consider the principal branch of the $\mathrm{LambertW}(x)$. More details can be found in Corless et al. (1996).

The property 5.1 turns out to be of great value for our purposes. It will allow us to get simplifications of long and complicated expressions. This function can be easily evaluated in the software Maple. In fact, this software can manipulate easily this function saving us a lot of work. In the following lines some of the stated equivalences are far from being obvious. To prove them formally it would take many pages and a lot of effort that will contribute little to our knowledge. Since our main interest is to get the final approximation we rely on Maple to do the tedious work for us.

Let $X \sim LN(\mu, \sigma^2)$ with distribution function F, so the k-th moment of the exponential distribution generated by F X after the change of variable $y = \log t$ is given by

$$\mathbb{E}\left[X^k \mathrm{e}^{\theta X}\right] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left\{\theta \mathrm{e}^y + ky - \frac{(y-\mu)^2}{2\sigma^2}\right\} dy. \tag{5.2}$$

The first important observation is that we have no means to solve this integral analytically. So our objective in the next paragraphs will be to provide a good approximation of it.

Note that the *Laplace-Stieltjes* transform of a lognormal random variable is finite only if the argument θ is not positive – recall that we have dropped the sign form the argument in the definition. Then, if $\theta < 0$ the function inside the brackets is convex. Even more, we can easily verify that this function goes to minus infinity if we let the $y \to \pm\infty$. This observation is important because of the exponential transformation the

main contribution to the total value of the integral will come from the region around the value ρ that maximizes the function inside the brackets, that is

$$\theta \mathrm{e}^{y} + ky - \frac{(y-\mu)^2}{2\sigma^2}.$$

By making the first derivative equal to zero and using the definition of the function LambertW given in (5.1) we can easily obtain that this is given by

$$\rho := -\mathrm{LambertW}(-\theta\sigma^2 \mathrm{e}^{k\sigma^2+\mu}) + \mu,$$

(Maple is not really necessary here). Recall that the main contribution for the integral comes around this value. So, the idea is to take the second order Taylor expansion around ρ of the function e^{y} an insert it in the integral (5.2) to obtain the following approximation

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left\{\theta \mathrm{e}^{\rho}\Big[1 + (y-\rho) + \frac{(y-\rho)^2}{2}\Big] + ky - \frac{(y-\mu)^2}{2\sigma^2}\right\} dy. \tag{5.3}$$

There were several reasons for choosing this approximation. The first one, is that by substituting the term e^{y} for a polynomial of second order we will be left with the kernel of a normal distribution, so we just need to complete its distribution to get rid of the integral. In fact this will yield to an approximation via a lognormal distribution which will be explained later. Secondly, the functions inside the brackets in expressions 5.2 and 5.3 are both convex functions which tend to minus infinity as $\to \pm\infty$. Since these functions differ significatively only in regions away from the value ρ, the exponential transformation will yield a small error of the approximation. These kind of methods are known as *saddlepoint approximation*.

We proceed to complete the density function of a normal random variable and leave all constants outside the integral. This is a tedious work and the expressions become more complicated. In order to simplify the notation we define $\nu = (1 + |\theta|\mathrm{e}^{\rho}\sigma^2)^{-1}$. Making the necessary algebra and substitutions we get

$$\frac{\exp\left\{\theta\, \mathrm{e}^{\rho}(1-\rho+\frac{\rho^2}{2}) + \frac{[k+\mu/\sigma^2-|\theta|\mathrm{e}^{\rho}(1-\rho)]^2\sigma^2}{2(1+|\theta|\mathrm{e}^{\rho}\sigma^2)} - \frac{\mu^2}{2\sigma^2}\right\}}{\sqrt{1+|\theta|\mathrm{e}^{\rho}\sigma^2}} \tag{5.4}$$

$$\times \int_{-\infty}^{\infty} \frac{\exp\left\{-\frac{\left(y-[k+\mu/\sigma^2-|\theta|\mathrm{e}^{\rho}(1-\rho)]\nu\sigma^2\right)^2}{2\nu\sigma^2}\right\}}{\sqrt{2\pi\nu\sigma^2}} dy. \tag{5.5}$$

The integral vanishes form the previous expression, so we are left only with the expression (5.4) which still depends on ρ. However, this is where we can take advantage of the function LambertW and a software like Maple to see that after a large simplification we will obtain

$$\mathbb{E}\left[X^k \mathrm{e}^{\theta X}\right] \approx \frac{\exp\left\{-\frac{\mathrm{LambertW}^2(|\theta|\sigma^2\mathrm{e}^{k\sigma^2+\mu})+2\,\mathrm{LambertW}(-\theta\sigma^2\mathrm{e}^{k\sigma^2+\mu})}{2\sigma^2}\right\}}{\sqrt{\mathrm{LambertW}(-\theta\sigma^2\mathrm{e}^{\mu})+1}}. \tag{5.6}$$

By taking $k=0$ in the last expression we obtain and approximation for $\varphi(\theta)$ which will be denoted $\widetilde{\varphi}(\theta)$.. Hence, we may use this to approximate the distributions in the

exponential family generated by a lognormal random variable. As discussed previously, this is given by

$$d\mathbb{P}_\theta(\omega) = \frac{e^{\theta X}}{\varphi(\theta)} d\mathbb{P}(\omega) \approx \frac{e^{\theta X}}{\widetilde{\varphi}(\theta)}\, d\mathbb{P}(\omega).$$

This clearly will not integrate to 1 due to the approximation. However, if we consider the probability $\mathbb{P}_\theta(X < x)$ we could follow exactly the same procedure as before to show that

$$\begin{aligned}\mathbb{P}_\theta(X < x) &= \int_0^x \frac{\mathrm{e}^{\theta x}}{\varphi(\theta)} F(dx) \approx \frac{1}{\widetilde{\varphi}(\theta)} \int_0^x \mathrm{e}^{\theta x} F(dx) \\ &\approx \frac{\widetilde{\varphi}(\theta)}{\widetilde{\varphi}(\theta)} \int_{-\infty}^{\log x} \frac{\exp\Big\{ -\frac{\left(y - [\mu/\sigma^2 - |\theta| \mathrm{e}^\rho (1-\rho)]\nu\sigma^2\right)^2}{2\nu\sigma^2} \Big\}}{\sqrt{2\pi\nu\sigma^2}} dy.\end{aligned}$$

Recall 5.5. So, the most relevant feature of the previous approximation was that the second order Taylor expansion allowed us to identify the kernel of a lognormal distribution, so we could have safely ignored the real value of $\varphi(\theta)$ since it was going to cancel itself in the next approximation. Moreover, the function LambertW will allow us to obtain further simplifications of the localization and dispersion parameters. That is

$$\frac{\exp\Big\{ -\frac{\left(y - [\mu/\sigma^2 - |\theta| \mathrm{e}^\rho (1-\rho)]\nu\sigma^2\right)^2}{2\nu\sigma^2} \Big\}}{\sqrt{2\pi\nu\sigma^2}} = \frac{\exp\Big\{ -\frac{\left(y - \mu + \mathrm{LambertW}(-\theta\sigma^2\mathrm{e}^\mu)\right)^2}{2\sigma^2(1+\mathrm{LambertW}(-\theta\sigma^2\mathrm{e}^\mu))^{-1}} \Big\}}{\sqrt{2\pi\sigma^2(1+\mathrm{LambertW}(-\theta\sigma^2\mathrm{e}^\mu))^{-1}}}.$$

Hence, we have approximated the elements in the *exponential family* generated by a random variable $X \sim LN(\mu, \sigma^2)$ with a lognormal distribution $LN(\mu_\theta, \sigma_\theta^2)$ where

$$\mu_\theta := \mu - \mathrm{LambertW}(-\theta\sigma^2\mathrm{e}^\mu), \qquad \sigma_\theta^2 = \frac{\sigma^2}{1 + \mathrm{LambertW}(-\theta\sigma^2\mathrm{e}^\mu)}.$$

To emulate the optimal exponential change of measure we will need to find the value θ such that $\mathbb{E}_\theta X = x$. Recall that the expectation of a lognormal random variable with localization parameter μ_θ and dispersion parameters σ_θ^2 is given by $\mathbb{E}_\theta[X] = \exp\{\mu_\theta + \sigma_\theta^2/2\}$. So, for this approximation we just need to insert the expressions we found previously, set $\mathbb{E}_\theta[X] = x$ and solve for θ. In fact, if $x < 1$ then there exist two solutions which are implicit given by

$$\mathrm{LambertW}(-\theta\sigma^2\mathrm{e}^\mu) = \frac{-1 + \mu - \log x \pm \sqrt{(1 - \mu - \log x)^2 + 2\sigma^2}}{2}. \tag{5.7}$$

Inserting this in the expression for the parameters μ_θ and σ_θ^2 we arrive to

$$\begin{aligned}\mu_\theta &:= \frac{1 + \mu + \log x \mp \sqrt{(1 - \mu - \log x)^2 + 2\sigma^2}}{2}, \\ \sigma_\theta^2 &:= \frac{2\sigma^2}{1 + \mu - \log x \pm \sqrt{(1 - \mu - \log x)^2 + 2\sigma^2}}.\end{aligned}$$

The value of θ can be obtained using the properties of the function LambertW and it is given by

$$\theta = -\frac{\gamma \mathrm{e}^\gamma}{\sigma^2 \mathrm{e}^\mu}, \qquad \gamma := \frac{-1 + \mu - \log x \pm \sqrt{(1 - \mu - \log x)^2 + 2\sigma^2}}{2}.$$

Note that if we let $x \to 0$ then $|\log x| \to \infty$. If we choose the negative root in (5.7), then as we let $x \to 0$ we will obtain

$$\mu_\theta \to \mu, \qquad \sigma_\theta^2 \to \sigma^2, \qquad \theta \to 0,$$

while if we choose the positive root, the limits will be instead

$$\mu_\theta \to -\infty, \qquad \sigma_\theta^2 \to 0, \qquad \theta \to -\infty,$$

which is in accordance with our intuition. That is, the value of θ should become smaller as we let $x \to 0$.

5.2 Importance Sampling Algorithms

Next we will implement importance sampling algorithms for estimating the rare event probability $\mathbb{P}(S_n < nx)$ as $x \to 0$ in the case of i.i.d. lognormal random variables using the results from the discussion above.

First, with the approximation of the *Laplace transform*. Recall that the only member of the *exponential family generated by* F which yields to a logarithmically efficient algorithm is the distribution F_θ such that θ is the solution of $\mathbb{E}_\theta = x$. Therefore, since the implementation depends on several approximation we have no warranty that this estimator will inherit the logarithmic efficiency. The second idea is to use the approximated lognormal distribution of the optimal F_θ.

However, the following simple algorithm can be easily implemented together with any of the strategies implemented above and it is easily proved to have logarithmic efficiency for a fixed n. We discuss it next.

Recall that the optimal importance sampling is obtained as the distribution of the random vector $(X_1, \ldots, X_n)$ conditioned to the rare event $\{S_n < nx\}$. In particular if the X_i's are nonnegative random variables it follows in a straightforward way that the random variable $(X_i | S_n < nx)$ should be supported over the interval $(0, nx)$. Then, if we apply importance sampling using a truncated lognormal random variable we will obtain a significative variance reduction. For instance, we can see this strategy either as an importance sampling or a stratification: just divide the probability of interest in two components, namely

$$\mathbb{P}(S_n < nx) = \mathbb{P}(S_n < nx, M_n < nx) + \mathbb{P}(S_n < nx, M_n > nx) = \mathbb{P}(S_n < nx, M_n < nx).$$

Denote Q the probability measure under which the X_i's are truncated lognormal random variables restricting its support over the interval $(0, nx)$ and let F_Q be its distribution function. Then

$$\begin{aligned}\mathbb{P}(S_n < nx, M_n \leq nx) &= \mathbb{P}(S_n < nx | M_n < nx)\mathbb{P}(M_n < nx) \\ &= Q(S_n < nx)\mathbb{P}(M_n < nx).\end{aligned}$$

So, if define $\widetilde{X}_1, \ldots, \widetilde{X}_n \sim F_Q$ in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we will have that

$$\widehat{z}(x) = \mathbb{P}(M_n < nx)\mathbb{I}(\widetilde{X}_1 + \cdots + \widetilde{X}_n < nx)$$

is an unbiased estimator of $\mathbb{P}(S_n < nx)$ with second moment given by

$$\mathbb{E}\left[\widehat{z}^2(x)\right] = \mathbb{P}^2(M_n < nx)Q(S_n < nx) = \mathbb{P}(M_n < nx)\mathbb{P}(S_n < nx).$$

From this expression, we can prove the efficiency of this estimator.

Theorem 5.2. *Let $X_1, \dots, X_n$ be a sequence of i.i.d. lognormal random variables and $\widetilde{X}_i \stackrel{d}{=} (X_i | X_i < ny)$. Then*

$$\widehat{z}(x) = \mathbb{P}(M_n < nx)\mathbb{I}(\widetilde{X}_1 + \cdots + \widetilde{X}_n < ny)$$

is an unbiased estimator of $\mathbb{P}(S_n < ny)$ which is logarithmically efficient.

Proof. For $\epsilon > 0$ we have

$$\begin{aligned}\limsup_{x\to 0} \frac{\mathbb{E}\,\widehat{z}^2(x)}{\mathbb{P}(S_n < nx)^{2-\epsilon}} &= \limsup_{x\to 0} \frac{\mathbb{P}(M_n < nx)\mathbb{P}(S_n < nx)}{\mathbb{P}^{2-\epsilon}(S_n < nx)} \\ &= \limsup_{x\to 0} \frac{\mathbb{P}(M_n < nx)}{\mathbb{P}^{1-\epsilon}(S_n < nx)} \\ &\leq \limsup_{x\to 0} \big(\frac{\mathbb{P}(X_1 < nx)}{\mathbb{P}^{1-\epsilon}(X_1 < x)}\big)^n.\end{aligned}$$

By *Mill's ratio* we have that

$$\mathbb{P}(X_1 < x) \sim \frac{\sigma}{\sqrt{2\pi}|\log x|} \exp\big\{-\frac{\log^2 x}{2\sigma^2}\big\}, \qquad x \to 0.$$

So it can be easily verified that

$$\frac{\mathbb{P}(X_1 < nx)}{\mathbb{P}^{1-\epsilon}(X_1 < x)} \sim \frac{|\log x|}{|\log nx|} \frac{\exp\big\{-\frac{\log^2 nx}{2\sigma^2}\big\}}{\exp\big\{-\frac{\log^2 x}{2\sigma^2}(1-\epsilon)\big\}} \to 0, \qquad x \to 0$$

with n fixed. The last proves that the estimator $\widehat{z}(x)$ is logarithmically efficient. □

Trivially, this algorithm has zero variance for $n = 1$ and its efficiency will decrease as $n \to \infty$. Numerically, it provides a modest variance reduction, but its significance is that it attains logarithmical efficiency. Moreover, the exponential change of measure was inspired by a large deviation result which means that the importance sampling estimator based on it should improve its efficiency as $n \to \infty$. So by implementing an algorithm which incorporates these two features we expect to obtain a solid algorithm for both large and small values of n.

For the rest of the discussion we will assume that $\theta(x)$ solves $\mathbb{E}_\theta[X] = x$ where X follows a $LN(0, \sigma^2)$ distribution. The assumption that $\mu = 0$ is made without lost of generality since we can substitute X_i and x by $X_i \mathrm{e}^{-\mu}$ and $x\mathrm{e}^{\mu}$ respectively in the expression $X_1 + \cdots + X_n < nx$. We denote f the common density function of $X_1, \dots, X_n$.

Importance Sampling via approximated exponential change of measure. Since we are considering i.i.d. random variables the discussion will be centered in the marginal distributions. The first strategy is to use the following importance sampling.

$$\widetilde{f}(t) := \frac{\mathrm{e}^{\theta t} f(t)\mathbb{I}(t \in (0, nx))}{\mathbb{E}\,[\mathrm{e}^{\theta X}; (0, nx)]}.$$

Hence, if $V_1, \dots, V_n$ are i.i.d. random variables with density function as above, the importance sampling estimator will be given by

$$\mathbb{I}(V_1 + \cdots + V_n < nx) \frac{\mathbb{E}^{\,n}[\mathrm{e}^{\theta X}; (0, nx)]}{\mathrm{e}^{\theta(V_1 + \cdots + V_n)}},$$

We give the estimator in algorithmic form. A discussion of steps 1 and 2 is given later.

1. Simulate $V_1, \ldots, V_n$ from $\widetilde{f}$.
2. Approximate $\mathbb{E}\left[\mathrm{e}^{\theta X}; (0, nx)\right]$.
3. Return
$$\widehat{z}_1(x) = \mathbb{I}(V_1 + \cdots + V_n < nx) \frac{\mathbb{E}^{\,n}[\mathrm{e}^{\theta X}; (0, nx)]}{\mathrm{e}^{\theta(V_1 + \cdots + V_n)}}.$$

The expression $\mathbb{E}\left[\mathrm{e}^{\theta X}; (0, nx)\right]$ in the step 2 is unknown. We can approximate it using the expression in (5.6). However, this will introduce some bias which cannot be measured so far. Hence, we choose instead to approximate it via Monte Carlo estimates using truncated lognormal random variables to estimate it. To generate the V_i's from $\widetilde{f}$ in the step 1 we suggest to use an Acceptance-Rejection strategy. An algorithm with good results is given next.

1. Simulate $V \sim LN(0, \sigma^2)$ restricted to $(0, nx)$.
2. Simulate $U \sim U(0, 1)$.
3. If $U > e^{\theta X}$ reject X and go back to Step 1. Else accept X.

This algorithm can be improved significatively and this will be part of future work. For now it has shown good performance in our numerical examples.

Theoretically this estimator should be logarithmically efficient. However, recall that we have used an approximation for θ, so it is uncertain if it remains logarithmically efficient even if we have used truncated lognormal random variables. This will be part of future work.

Moreover, for implementing this algorithm we should also be able to simulate from $\widetilde{F}$ which can be easily accomplished by implementing a *acceptance-rejection* algorithm. Note also that we do not have an analytic expression for $\mathbb{E}\left[\mathrm{e}^{\theta X}; (0, nx)\right]$. One way to sort this problem is to follows the procedure described in Section 5.1 to provide an approximation. Another option is to estimate it via Monte Carlo. But this procedure will introduce another source of variability in the estimator. The next alternative estimator avoids these two problems.

Importance Sampling via an approximated distribution of the optimal ECM. The second proposal is to use the approximated distribution of the optimal *exponential distribution generated by* F that we found in the Section 5.1. That is a $LN(\mu_\theta, \sigma_\theta^2)$ distribution where

$$\mu_\theta = \frac{1 - |\log x| - \sqrt{(1 - |\log x|)^2 + 2\sigma^2}}{2}$$
$$\sigma_\theta^2 = \frac{2\sigma^2}{1 + |\log x| + \sqrt{(1 - |\log x|)^2 + 2\sigma^2}}.$$

We define f_θ as the common density function of the $\widetilde{Z}_i$'s. Algorithmically

1. Simulate $\widetilde{Z}_1, \ldots, \widetilde{Z}_n$ from $LN(\mu_\theta, \sigma_\theta^2)$. and restricted $(0, nx)$.
2. Return
$$\widehat{z}_1(x) = \mathbb{I}(\widetilde{Z}_1 + \cdots + \widetilde{Z}_n < nx) \frac{f(\widetilde{\mathbf{Z}})}{f_\theta(\widetilde{\mathbf{Z}})} \mathbb{P}^n(\widetilde{Z}_1 < nx).$$

Clearly this algorithm is unbiased and much simpler to implement than the previous one. The results in Asmussen and Rubinstein (1995) do not exclude it from been logarithmically efficient but so far we have not found either a proof of it or a counterexample. In the subsection of numerical examples we make an empirical study of these two algorithms.

5.3 Concluding Remarks

5.3.1 Numerical Examples

We have implemented the suggested Monte Carlo estimates to approximate the probability $\mathbb{P}(S_n < nu)$ in the case of i.i.d. lognormal random variables with common distribution $LN(0, 1)$. The results are summarized in the Example D.11 on page 116.

The conclusion is that both algorithms provide good numerical results. It is, however, somewhat surprising that the Importance sampling estimator based on a Lognormal distribution shows a substantial better performance than the one suggested by means of the approximated optimal exponential change of measure. This will be investigated in future work.

Notes and Comments

Recently, Rossberg (2008); Rossberg et al. (2008) proposed a method for approximating the moment-generating function of a lognormal random variable. Part of future research might include some numerical comparison of that method against the one proposed here.

Chapter 6

The Economical Value of Realized Covariance

The variances and covariances of assets returns are of major importance in the study of option pricing, portfolio and risk management. Recently a new model which exploits high-frequency information has been proposed in Andersen et al. (2001); Barndorff-Nielsen and Shephard (2002). The main idea of the model is that if the log prices are generated by a multivariate continuous stochastic process and a sample path of the process is observed, then the theory of quadratic variation implies that the covariance becomes in fact observable. Hence, increasing the sampling frequency will yield to arbitrarily precise estimates which are usually called *realized volatility* or *realized covariance*.

In this chapter we analyze empirically the economical value of *realized covariance* in investment decisions. Our study is conducted within the framework developed by Fleming et al. (2001, 2003) where it is consider an investor who rebalances his portfolio daily following a volatility-timing strategy. This means that the weights in his portfolio vary only with with changes in his estimates of the conditional covariance matrix of daily returns. Our contribution is to provide a similar evaluation using alternative estimation of the covariance matrix than the one employed in Fleming et al. (2003). The reason being that for most high-frequency the prices are observed at discrete times in a non-synchronous manner (cf. Epps, 1979) and available only for part of a day, so the information unfolded during the inactive period is lost.

The first issue was addressed in Hayashi and Yoshida (2005) where they proposed an alternative estimator to the realized covariance named *cumulative covariance*. In there, they consider the problem of estimating the covariance of two diffusion processes observed at discrete non-synchronous times which is free of bias in the absence of noise. For the second issue, Hansen and Lunde (2005) proposed a *realized variance for the whole day* by means of an optimal combination of the realized variance for the active period and the squared overnight return. A further issue which affects the quality of the estimator is that the data is contaminated by noise. This problem has been largely studied in the literature (see for instance Hansen and Lunde, 2006, and references therein). In particular two methods have been proposed to address this problem: *subsampling* and *realized kernels* proposed in Zhang et al. (2005); Barndorff-Nielsen et al. (2008a) respectively. A mixed approach is given in Barndorff-Nielsen et al. (2008b). Furthermore, Voev and Lunde (2007) prove that the *cumulative covariance* estimator is biased and inconsistent in the presence of noise and propose a noise bias-

correction based on *realized kernels* and *subsampling*.

The objective here is unify the ideas of Hayashi and Yoshida (2005) and Hansen and Lunde (2005). That is, to construct a realized covariance matrix for the whole day based on high-frequency non-synchronous data and analyze its economical value in investment decisions using stock data from the NYSE. The rest of the chapter is divided as follows. In the Section 6.1 we described the proposals by Hayashi and Yoshida (2005); Hansen and Lunde (2005) in order to provide the guidelines on how to construct a realized covariance matrix for the whole day. In Section 6.2 we described the empirical framework following the work by Fleming et al. (2001, 2003) to measure the economical value of realized covariance in investment decisions. The material in this chapter is work in progress; further details, empirical results and conclusions will be reported elsewhere. Noise corrections will also be investigated.

6.1 Cumulative Covariance for the Whole Day

6.1.1 Cumulative Covariance

We describe the methodology given in Hayashi and Yoshida (2005) for estimating the covariance of two processes recorded at random times not necessarily regularly spaced. Here we are interested in the covariation of two diffusion process $X(s)$ and $Y(s)$ (continuous-time Itô semimartingales) which will be denoted

$$V(X,Y) := \langle X(s), Y(s) \rangle.$$

In the case where the two process are observed at discrete times, say, $s_0\, s_1, \ldots, s_m$, the following estimator

$$V_{\pi(m)}(X,Y) := \sum_{i=1}^{m} \big(X(s_i) - X(s_{i-1})\big)\big(Y(s_i) - Y(s_{i-1})\big), \qquad \pi(m) := \max_{1 \le i \le m},$$

is known as the *realized covariance* and is such that if $\pi(m) \to 0$ then $V_{\pi(m)} \to V$ in probability. This statistic has been largely studied and it has been proved to have significative advantages over other volatility models. However, the statistic $V_{\pi(m)}$ has limited practical use since financial data is usually registered at random times and that implies that the processes $X(s)$ and $Y(s)$ are not observed at the same time. Hayashi and Yoshida (2005) document on the methods available for synchronizing the data by means of an interpolation or an imputation of the prices. Hence, the resulting estimator will have a serious bias.

They proposed a new estimator completely different which makes uses of the original data exclusively, do not depend on interpolations or imputations so it is free of bias typically caused by non-synchronous observation.

For a continuous process $X(s)$ and a semiopen interval I=(a,b] we let $\Delta X(I) := X(b) - X(a)$. For a given time interval $(0, S]$ we consider two partitions $I := \{(i_k, i_{k+1}] : k = 1, \ldots, n\}$ and $J := \{(j_k, j_{k+1}] : k = 1, \ldots, m\}$. That is, I and J are sets of disjoint intervals of the form $(a, b]$ whose respective union is equal to the set $(0, S]$. Then we define.

Definition 6.1 (Cumulative Covariance)**.** The *Cumulative Covariance Estimator* among two non-synchronous processes X and Y observed at the times determined by the par-

titions Π_X and Π_Y is given by

$$V_{I,J}(X,Y) = \sum_{k,\ell} \Delta X\big((i_k, i_{k+1}]\big)\Delta Y\big((j_\ell, j_{\ell+1}]\big)\mathbb{I}\big((i_k, i_{k+1}] \cap (j_\ell, j_{\ell+1}] \neq \emptyset\big).$$

That is, the product of any pair of increments $\Delta X(I_i)$ and $\Delta Y(J_j)$ will make a contribution to the summation only when the respective observation intervals are overlapping. Moreover, Hayashi and Yoshida (2005) provide an estimator for the correlation given by

$$C_{I,J}(X,Y) = \frac{\sum\limits_{k,\ell} \Delta X\big((i_k, i_{k+1}]\big)\Delta Y\big((j_\ell, j_{\ell+1})\big)\mathbb{I}\big((i_k, i_{k+1}] \cap (j_\ell, j_{\ell+1}] \neq \emptyset\big)}{\big[\sum\limits_{k} \Delta X\big((i_k, i_{k+1}]\big)^2\big]^{1/2}\big[\sum\limits_{\ell} \Delta Y\big((j_\ell, j_{\ell+1}]\big)^2\big]^{1/2}}.$$

These estimators are unbiased and consistent in the absence of noise. Voev and Lunde (2007) propose a bias-corrected estimator and ultimately a subsampling version of the bias-corrected version which improves efficiency. However, it is not clear that this corrected estimator yields to a positive-definite covariance matrix.

6.1.2 Variance for the Whole Day

A second concern is to obtain a covariance estimator for the whole day. I am not aware of any paper dealing with this problem. However, Hansen and Lunde (2005) propose an estimator for the realized variance for the whole day using and optimal combination of the realized variance and squared overnight return. A naive solution to this problem could be combine this estimator with the correlation matrix given in Hayashi and Yoshida (2005).

Recall that the main concern here is that high-frequency data is only available during the trading period which just represents a fraction of the day, and therefore a substantial part of the daily volatility which unfolds during the inactive period is lost. Hansen and Lunde (2005) propose to approximate this lost information by using the square of the return registered during the whole inactive period of the t-th day, say $r_t^2(X)$, and determine the weights ω_1 and ω_2 which yield to the optimal linear combination

$$V_t(X) := \omega_1 r_t^2 + \omega_2 V_{I_t}(X),$$

where $V_t(X)$ stands for the realized variance for the whole t-th day, $V_{I_t}(X)$ the realized variance for the active period of the t-th day and I_t the corresponding partition. Here we have defined $V_{I_t}(X)$ with some abuse of notation instead of $V_{\pi(m)}(X,X)$ as could be denoted from the discussion above. In future work we will come with a more appropriated notation for this. The realized covariance for t whole days is implicit in the following definition for the weights.

Definition 6.2. Let $\mu_1 := \mathbb{E}\,(r_t^2)$ $\eta_1^2 := \mathbb{V}\mathrm{ar}\,(r_t^2)$, $\mu_2 := \mathbb{E}\,(V_{I_t}(X))$ $\eta_2^2 := \mathbb{V}\mathrm{ar}\,(V_{I_t}(X))$ and $\eta_{12} := \mathbb{C}\mathrm{ov}\,(r_t^2, V_{I_t}(X))$ and define

$$\omega_1 := (1-\varphi)\frac{\mu_0}{\mu_1}, \qquad \omega_2 := \varphi\frac{\mu_0}{\mu_2},$$

where $\mu_0 := \mathbb{E}\,[V_{I_t}(X)]$ and φ is a factor defined by

$$\varphi := \frac{\mu_2^2\eta_1^2 - \mu_1\mu_2\eta_{12}}{\mu_2^2\eta_1^2 + \mu_1^2\eta_2^2 - 2\mu_1\mu_2\eta_{12}}.$$

This solution is straightforward to implement since the mean and variances can be estimated by averages.

Hence our proposal is to estimate independently pairwise realized correlations and scale them by the realized variance for the whole day obtained previously to obtain the realized covariance for $n = 30$ stocks of the Dow Jones Industrial Average. With these estimates we will construct a realized covariance matrix for the t-th day which we will denote $\mathbf{V}_t \in \mathbb{R}^{n\times n}$. Further improvements may include subsampling for removing noise from the suggested estimate.

6.2 Empirical Framework

In this section we describe the empirical framework proposed in Fleming et al. (2001, 2003) where it is consider two cases of an investor allocating funds across the set of 30 stocks of the Dow Jones Industrial Average plus cash using conditional mean-variance analysis. We explain the details in the rest of the section.

Our investor requires one-step-ahead estimates of the conditional means and the conditional covariance matrix. So, we denote $\mathbf{X}_t(s)$ as the multivariate process of prices comprising exclusively the day t. Let $\mathcal{F}_t$ the information available until day $t-1$. It is assumed that conditional mean vector $\boldsymbol{\mu} := \mathbb{E}\left[\mathbf{X}_t|\mathcal{F}_t\right]$ is constant since there is little variability during one day. We construct the conditional covariance matrix $\boldsymbol{\Sigma}_t := \mathbb{V}\mathrm{ar}\left[\mathbf{X}_t|\mathcal{F}_t\right]$ of daily returns using rolling estimators. Fleming et al. (2003) point out that this approach is advantageous since it avoids parametric assumptions, it nests a variety of GARCH and stochastic volatility models as special cases and it is computational efficient. The optimal weighting scheme is of the form

$$\widehat{\boldsymbol{\Sigma}}_t = \exp(-\alpha)\widehat{\boldsymbol{\Sigma}}_{t-1} + \alpha\exp(-\alpha)\mathbf{V}_{t-1},$$

where $\mathbf{V}_{t-1}$ is the covariance matrix for the whole t-th day proposed previously. This estimator is optimal in the sense that it produces the smallest asymptotic MSE. It guarantees that $\widehat{\boldsymbol{\Sigma}}_t$ is positive-definite, and a single parameter α controls the rate at which the weights decay with the *lag* length. This approach should be more efficient since realized variances and covariances provide better estimates of the daily return innovations.

With these estimates, our risk-averse investor wishes to allocate his funds across the 30 stocks of the Dow Jones Industrial Average and cash rebalancing his portfolio daily. In order to avoid restrictions on short selling and transaction costs he trade future contracts over the risky assets. Before we continue, a few notes of futures are included (for more details see Björk, 2004).

Definition 6.3 (Forwards)**.** Consider a contingent claim X. A forward contract on X, contracted at time s with time of delivery S and forward price $f(s,S,X)$ is defined by the payment scheme:

1. The holder of the forward receives X at time S and pays $f(s,S,X)$.

2. The forward price $f(s,S,X)$ is determined at time s and in such way that the price at time t (when the contract is made) is 0.

The main difference between a futures and a forward contract lies in the way the payments are made:

Definition 6.4 (Futures)**.** Consider a contingent claim X. A futures contract on X, with time of delivery S is a financial asset with the following properties:

1. At any point s in $(0, S]$ there exists in the market an object $F(s, S, X)$ known as the futures price for X at time s, for delivery at T.

2. During an arbitrary time interval $(s_1, s_2]$, the holder of the contract receives the amount $F(s_2, S, X) - F(s_1, S, X)$.

3. At time S of delivery the holder receives X and pays $F(S, S, X)$.

4. The spot price, at any time s prior the deliver, of obtaining the futures contract is by definition 0.

One of the reasons to trade futures over markets is the difficulty of trading or hedging directly in the underlying object particularly with commodities where it is not allowed to go short. We must have $F(S, S, X) = X$, so there is no reason to deliver X at the time S. The only contractual obligations is the stream of payments described above. There is no cost or gain of entering or closing a futures contract at any time. Most futures contracts are closed before the delivery date. Forward and futures prices are the same when interest rates are deterministic. There is a difference when interest rates are stochastic. The forward price is a martingale under the forward measure, whereas the futures price is a martingale under the risk neutral measure.

Going back to our investor, he decides to use *conditional mean-variance* analysis to make his allocation decisions. Two strategies are possible:

- He targets a specific expected portfolio return μ_* while minimizing the variance. The following portfolio weights delivers the solution to this optimization problem

$$\mathbf{w}_\mu(t, \mu_*) = \frac{(\mu_* - R_f)\mathbf{\Sigma}_t^{-1}\boldsymbol{\mu}}{\boldsymbol{\mu}'\mathbf{\Sigma}_t^{-1}\boldsymbol{\mu}}, \tag{6.1}$$

  where R_f the return of the risk free asset.

- He targets a specific portfolio variance σ_* while maximizing the expected return. The solution is given by

$$\mathbf{w}_\sigma(t, \sigma_*) = \frac{\sigma_*\mathbf{\Sigma}_t^{-1}\boldsymbol{\mu}}{\sqrt{\boldsymbol{\mu}'\mathbf{\Sigma}_t^{-1}\boldsymbol{\mu}}}. \tag{6.2}$$

The optimization is carried out using Lagrangian multipliers. The solutions of both problems differ just by a scalar term. The weight in cash is one minus the sum of the weights of the assets.

Finally we want to measure the performance associated with the use of the proposed realized covariance for the whole day and compare against other approaches as the ones discussed in Fleming et al. (2001, 2003). For doing so, they propose the use of the *quadratic utility*. The quadratic utility function is defined as

Expected return - Risk Aversion × Expected Variance / 2.

Such function can be seen as a second-order approximation of the investor's true utility. The investor's quadratic *realized* utility in period $t+1$ is given by

$$U(W_{t+1}) = W_{t+1} - \frac{aW_{t+1}^2}{2} = W_t \mathbf{R}_{t+1} - \frac{aW_t^2}{2}\mathbf{R}_{t+1}^2, \tag{6.3}$$

where W_t is the investor's wealth at time t, a is his absolute risk aversion (coefficient) and

$$\mathbf{R}_{t+1} = R_f + \mathbf{w}'\mathbf{X}_{t+1}.$$

Consider $\mathbf{R}_1$ and $\mathbf{R}_2$ the vector or returns for two different strategies, then the measure of comparison of the performance of the two strategies will be given by the value Δ which solves

$$\overline{U}(\mathbf{R}_1 - \Delta\mathbf{1}) = \overline{U}(\mathbf{R}_2).$$

Such value is reported as annualized fee in basis points (.01%).

Appendix A

Probability Distributions

A.1 Univariate Distributions

A.1.1 The Lognormal Distribution

Definition A.1 (Lognormal Random Variable)**.** A random variable X is said to have a *three parameter lognormal distribution* if there exist $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}^+$ and $\gamma \in \mathbb{R}$ such that

$$\frac{\log(X-\gamma)-\mu}{\sigma^2} \sim \mathrm{N}(0,1).$$

In particular, if we take $\gamma = 0$ the resulting random variable X is said to have *two parameter lognormal distribution* or simply *lognormal distribution* (Johnson et al., 1994).

It is easily deduced that X only takes values over (γ, ∞). Moreover, since it has been defined as a transformed distribution it is possible to use many of the properties of the normal standard distribution. For instance its probability density function is given by

$$f(t;\mu,\sigma^2,\gamma) := \frac{1}{\sqrt{2\pi\sigma^2}(t-\gamma)} \exp\Big\{ -\frac{\big(\log(t-\gamma)-\mu\big)^2}{2\sigma^2}\Big\}, \qquad t \geq \gamma. \tag{A.1}$$

and its cumulative distribution function can be evaluated as

$$F(t;\mu,\sigma,\gamma) := \Phi\bigg(\frac{\log(t-\gamma)-\mu}{\sigma}\bigg),$$

where Φ is the normal standard cumulative distribution function. From this expression it is possible to observe that γ is a localization parameter. That is, changes on its value will only affect the localization of the distribution but not its shape. Hence if we fix the value of γ most of the properties of this distribution can be easily transferred to the general case.

A relevant feature of the lognormal distribution is that

$$\mathbb{E}\left[\mathrm{e}^{\theta X}\right] = \infty, \qquad \theta > 0,$$

meaning that the moment generating function does not exist. However, all the moments of the *lognormal distribution* exist and can be calculated explicitly. In fact, the moment of order k is given by

$$\mathbb{E}\left[X^k\right] = \mathrm{e}^{k\mu + k^2\sigma^2/2}, \qquad k \in \mathbb{N}.$$

from where its mean and variance are easily obtained and given by

$$\mathbb{E}[X] = \mathrm{e}^{\mu+\sigma^2/2}, \tag{A.2}$$

$$\mathbb{V}\mathrm{ar}[X] = \mathrm{e}^{2\mu+\sigma^2}\big(\mathrm{e}^{\sigma^2}-1\big). \tag{A.3}$$

Define $\overline{F}(t) = 1-F(t)$ to be the *tail probability* of a lognormal random variable. The following is an asymptotically equivalent expression as $t \to \infty$ of the tail probability of a lognormal random variable

$$\overline{F}(t) \sim \frac{\sigma}{\sqrt{2\pi}(\log t-\mu)} \exp\Big\{-\frac{(\log t-\mu)^2}{2\sigma^2}\Big\}. \tag{A.4}$$

Although this is based in a well-known asymptotic expression for the normal distribution, this fact can be proved in a straightforward way using L'Hopital rule:

$$\lim_{t\to\infty} \frac{\frac{\sigma}{\sqrt{2\pi}(\log t-\mu)} \exp\Big\{-\frac{(\log t-\mu)^2}{2\sigma^2}\Big\}}{\overline{F}(t)} = \lim_{t\to\infty} \frac{f(t)\frac{\log^2 t-2\log t\mu+\mu^2+\sigma^2}{(\log t-\mu)^2}}{f(t)} = 1.$$

The *failure rate function* $\lambda(t)$ is defined to be the ratio among the pdf and the tail probability of a random variable. For the lognormal random variable we employ the expression (A.4) to obtain an asymptotic expression for $\lambda(t)$ as $t \to \infty$:

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)} = \frac{\log t-\mu}{\sigma^2 t}\big(1+o(1)\big) = \frac{\log t}{\sigma^2 t}\big(1+o(1)\big),$$

which clearly goes to 0 as $t \to \infty$. A useful fact about the lognormal failure rate function is that it remains bounded in the set $[0,\infty)$. To prove this it is just necessary to observe that $\lambda(t) \to 0$ when $t \to 0$ as well. The assertion follows from the continuity of the function and the two limits.

A.1.2 Regularly Varying Distributions

Definition A.2 (Slow Variation)**.** A function L with values in $(0,\infty)$ is said to be *slowly varying at infinity* if

$$\lim_{x\to x_0} \frac{L(tx)}{L(x)} = 1, \qquad t > 0.$$

The family of slowly varying functions is commonly denote $\mathcal{R}_0$.

Functions with limits at x_0 are typical examples of slowly varying functions. Note however, that while convergence is a sufficient condition it is not necessary. In fact a slowly varying function can be oscillatory at x_0. The next function, taken from Embrechts et al. (1997), is an example of a slowly varying function at ∞ with an oscillatory behavior

$$L(x) = \exp\Big\{\sqrt{\log(1+x)}\,\cos\big(\sqrt{\log(1+x)}\big)\Big\}.$$

Definition A.3 (Regular Variation)**.** A function h with values in $(0,\infty)$ is said to be *regularly varying at infinity* with index α if

$$\lim_{x\to x_0} \frac{h(tx)}{h(x)} = t^\alpha, \qquad t > 0.$$

The family of regularly varying functions with index α is commonly denoted $\mathcal{R}_0$.

Remark A.4. It is often useful to see the family $\mathcal{R}_\alpha$ through its close relation with the power function x^α and the family $\mathcal{R}_0$. More precisely, a function $h \in \mathcal{R}_\alpha$ can be written as $x^\alpha L(x)$ with $L \in \mathcal{R}_0$.

A.2 Multivariate Distributions

A.2.1 Multivariate Normal Distribution

Definition A.5. A random vector $\mathbf{Y} := (Y_1, \ldots, Y_d)^T$ is said to have a *Multivariate Normal* (or *Gaussian*) distribution if

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\,\mathbf{Z},$$

where $\mathbf{Z} := (Z_1, \ldots, Z_k)^T$ is a vector of i.i.d. normal standard random variables, $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\mathbf{A} \in \mathbb{R}^{k\times d}$.

Since $\mathbf{Y}$ is a linear transformation of the random vector $\mathbf{Z}$ we can easily calculate its mean vector and covariance matrix. That is

$$\mathbb{E}\,[\mathbf{Y}] = \boldsymbol{\mu}, \qquad\qquad \mathbb{C}\mathrm{ov}\,[\mathbf{Y}] = \boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma} := \mathbf{A}\mathbf{A}'$ is a positive (semi)definite matrix. Since the multivariate normal distribution is fully characterized by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ we will use the notation $\mathbf{Y} \sim N_d(\mu, \boldsymbol{\Sigma})$ and refer to it as *multivariate normal distribution with parameters μ and $\boldsymbol{\Sigma}$*.

An alternative (and equivalent) definition is stated in terms of the characteristic function. A random vector $\mathbf{Y}$ is said to be a multivariate normal with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ if its characteristic function can be written as

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \mathrm{e}^{i\,\boldsymbol{\mu}\,\mathbf{t}' - \frac{1}{2}\mathbf{t}\,\boldsymbol{\Sigma}\,\mathbf{t}'},$$

where $\boldsymbol{\mu}, \mathbf{t} \in \mathbb{R}^d$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ positive (semi)definite matrix.

Under the assumption that $\mathrm{rank}(A) = d$, the variance-covariance $\boldsymbol{\Sigma}$ has full rank and therefore it is *non-singular* and *positive semi-definite*. In such case, its joint density function is given by

$$f_{\mu,\boldsymbol{\Sigma}}(\mathbf{t}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\Big\{ -\frac{(\mathbf{t}-\mu)'\mathbf{K}(\mathbf{t}-\mu)}{2}\Big\}, \tag{A.5}$$

where $\mathbf{K} = \boldsymbol{\Sigma}^{-1}$ is known as the concentration matrix of the distribution. From the form of the joint density it is observed that the level sets do form *ellipsoids*.

Let $\mathbf{X}_1 \sim N_d(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathbf{X}_2 \sim N_d(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ be independent multivariate normal vectors. Several important properties of the multivariate distribution are listed below.

1. **Convolution.** The class of multivariate normal distributions is closed under convolutions. In particular

   $$\mathbf{X}_1 + \mathbf{X}_2 \sim N_d(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2).$$

2. **Linear transformations.** If $\mathbf{A} \in \mathbb{R}^{k\times d}$ and $\mathbf{u} \in \mathbb{R}^d$, then

   $$\mathbf{A}\mathbf{X}_1 + \mathbf{u} \sim N_k(\mathbf{A}\mu_1 + \mathbf{u}, \mathbf{A}\boldsymbol{\Sigma}_1\mathbf{A}').$$

   In particular

   $$\mathbf{u}'\mathbf{X}_1 \sim N(\mathbf{u}'\boldsymbol{\mu}_1, \mathbf{u}'\boldsymbol{\Sigma}\mathbf{u}).$$

3. **Transformation to the Normal Standard.** Let $\mathbf{U}^{-1}\Lambda\mathbf{U}$ be the eigenvalue decomposition of $\mathbf{\Sigma}$. Then

$$\Lambda^{-1/2}\mathbf{U}\mathbf{X}_1 \sim N_d(0, \mathbf{I}).$$

Partition $\mathbf{X}_1$ into $\mathbf{X}_A \in \mathbb{R}^r$ and $\mathbf{X}_B \in \mathbb{R}^s$ with $r + s = d$. Consequently

$$\boldsymbol{\mu}_1 = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \qquad \mathbf{\Sigma}_1 = \begin{pmatrix} \mathbf{\Sigma}_{AA} & \mathbf{\Sigma}_{AB} \\ \mathbf{\Sigma}_{BA} & \mathbf{\Sigma}_{BB} \end{pmatrix}.$$

4. **Marginal distributions.** $\mathbf{X}_A \sim N_r(\boldsymbol{\mu}_A, \mathbf{\Sigma}_{AA})$ and $\mathbf{X}_B \sim N_s(\boldsymbol{\mu}_B, \mathbf{\Sigma}_{BB})$.

5. **Conditional Distributions.** $\mathbf{X}_A|\mathbf{X}_B \sim N_r(\boldsymbol{\mu}_{A|B}, \mathbf{\Sigma}_{A|B})$ where

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A + \mathbf{\Sigma}_{AB}\mathbf{\Sigma}_{BB}^{-1}(\mathbf{X}_B - \boldsymbol{\mu}_B) \qquad \text{and} \qquad \mathbf{\Sigma}_{A|B} = \mathbf{\Sigma}_{AA} - \mathbf{\Sigma}_{AB}\mathbf{\Sigma}_{BB}^{-1}\mathbf{\Sigma}_{BA}.$$

A.2.2 Elliptical Distributions

Elliptical distributions constitute a large class of multivariate distributions which provide a natural generalization of the multivariate normal distribution. The material presented here follows the exposition of McNeil et al. (2005) and Fang et al. (1987). We start by defining the subclass of *spherical* distributions which will be useful for giving a full characterization of the *elliptical* distributions.

Definition A.6 (Spherical Distribution)**.** We say that a random vector $\mathbf{X}$ has an elliptical distribution if it has the stochastic representation

$$\mathbf{X} \stackrel{d}{=} R\,\mathbf{S},$$

where $\mathbf{S}$ is a random vector uniformly distributed over the unit sphere $\mathcal{S}$ and R is an independent random variable supported over $[0, \infty)$ (radial random variable).

The following definition will be useful to characterize the class of elliptical distributions.

Definition A.7 (Characteristic generator)**.** Let $\mathbf{X}$ be a random vector with an elliptical distribution. The *characteristic generator* $\psi(t)$ of $\mathbf{X}$ is defined as the characteristic function of $\|\mathbf{X}\|$.

Then we denote $\mathbf{X} \sim S_d(\psi)$. In particular, the *characteristic function* $\phi(\mathbf{t})$ of a spherical distribution is equivalent to the *characteristic generator* of $\mathbf{X}$ evaluated at $\|\mathbf{t}\|$. That is,

$$\phi(\mathbf{t}) = \psi(\|\mathbf{t}\|).$$

Spherical distributions have uncorrelated components (not necessarily independent) and their distributions are invariant under rotations. However, it is relevant to stress that the only spherical distribution with independent components is the multivariate normal distribution.

Definition A.8 (Elliptical distribution)**.** We say that a random vector $\mathbf{X}$ follows an elliptical distribution if it has the stochastic representation

$$\mathbf{X} = \boldsymbol{\mu} + R\,\mathbf{A}\,\mathbf{S},$$

where $\mathbf{S}$ is a random vector uniformly distributed over the unit sphere $\mathcal{S}$, R is a random variable which takes values in $[0, \infty)$, $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\mathbf{A} \in \mathbb{R}^{k \times d}$.

We denote $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ where $\boldsymbol{\mu}$ is known as the *localization parameter*, $\boldsymbol{\Sigma} = \mathbf{A}'\mathbf{A}$ as the *dispersion matrix* and ψ is the *characteristic generator* of the distribution.

A practical feature of the stochastic representation $\mathbf{X} \stackrel{d}{=} R\,\mathbf{S}$ is the fact that if R has no point mass at the origin then

$$R \stackrel{d}{=} \sqrt{(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})}, \qquad S \stackrel{d}{=} \frac{\boldsymbol{\Sigma}^{-1/2}\big(X - \boldsymbol{\mu}\big)}{\sqrt{(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})}},$$

and for details about the joint density function see McNeil et al. (2005); Fang et al. (1987). Below are listed some of the properties of the class of spherical distributions.

Let $\mathbf{X}_1 \sim S_d(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}, \psi_1)$ and $\mathbf{X}_2 \sim S_d(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}, \psi_2)$ have independent spherical distributions with common dispersion matrix $\boldsymbol{\Sigma}$, then

1. **Convolutions.**
$$\mathbf{X}_1 + \mathbf{X}_2 \sim S_d(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{\Sigma}, \psi_1\,\psi_2).$$
However, the convolution of two independent spherical distributions with non-identical dispersion matrices will not necessarily be elliptical.

2. **Linear transformations.** If $\mathbf{A} \in \mathbb{R}^{k \times d}$ and $\mathbf{u} \in \mathbb{R}^d$, then
$$\mathbf{A}\mathbf{X}_1 + \mathbf{u} \sim E_k(\mathbf{A}\mu_1 + \mathbf{u}, \mathbf{A}\boldsymbol{\Sigma}_1\mathbf{A}', \psi).$$
In particular
$$\mathbf{u}'\mathbf{X}_1 \sim E_1(\mathbf{u}'\boldsymbol{\mu}_1, \mathbf{u}'\boldsymbol{\Sigma}\mathbf{u}, \psi).$$

3. **Relation with the Spherical distributions**
$$\mathbf{X}_1 \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi) \quad \Longleftrightarrow \quad \boldsymbol{\Sigma}^{-1/2}\big(\mathbf{X}_1 - \boldsymbol{\mu}\big) \sim S_d(\psi).$$
Partition $\mathbf{X}_1$ into $\mathbf{X}_A \in \mathbb{R}^r$ and $\mathbf{X}_B \in \mathbb{R}^s$ with $r + s = d$. Consequently
$$\mu_1 = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \qquad \boldsymbol{\Sigma}_1 = \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}.$$

4. **Marginal distributions.** $\mathbf{X}_A \sim S_r(\mu_A, \boldsymbol{\Sigma}_{AA}, \psi)$ and $\mathbf{X}_B \sim S_s(\mu_B, \boldsymbol{\Sigma}_{BB}, \psi)$.

5. **Conditional Distributions** Conditional Elliptical Distributions are also elliptical, but not necessarily with the same characteristic generation function ψ. For more details see Fang et al. (1987).

A.2.2.1 Normal Variance Mixture Distributions

We will introduce the class of Normal Mixture Distributions which constitutes an important subclass of Elliptical Distributions.

Definition A.9. A random vector $\mathbf{Y}$ is said to have a *normal variance mixture distribution* if it has the stochastic representation

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\,\mathbf{Z},$$

where $\mathbf{Z} := (Z_1, \ldots, Z_k)$ is a vector of i.i.d. normal standard r.v.'s, $\mathbf{A} \in \mathbb{R}^{k\times d}$, $\mu \in \mathbb{R}^d$ and W is a random variable supported over $(0, \infty)$ and independent of $\mathbf{Z}$.

This type of multivariate distributions are named *variance mixtures* since the conditional distribution of $\mathbf{X}$ given $W = w$ has a $N_d(\boldsymbol{\mu}, w\boldsymbol{\Sigma})$ distribution where $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$. Hence, we may think of such distribution as a multivariate normal distribution with a *random covariance matrix*. Notice however that this is just an heuristic intuition since the resulting distribution will not be multivariate normal.

Its mean and covariance are given by

$$\mathbb{E}\left[\mathbf{X}\right] = \boldsymbol{\mu}, \qquad\qquad \mathbb{C}\mathrm{ov}\left(\mathbf{X}\right) = \mathbb{E}\left[W\right]\boldsymbol{\Sigma}$$

and its characteristic function

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathrm{e}^{i\,\mathbf{t}\,\boldsymbol{\mu}}\,\widehat{H}\Big(\frac{1}{2}\,\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\Big),$$

where $\widehat{H}$ is the Laplace transform of the distribution of W and the notation $\mathbf{X} \sim M_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \widehat{H})$ is adopted. Furthermore, if $\mathbf{X} \sim M_d(\mathbf{0}, \mathbf{I}_d, \widehat{H})$, it follows that its characteristic function $\phi(\mathbf{t}) = \widehat{H}(\mathbf{t}'\mathbf{t}/2)$ is indeed a *characteristic generator* of a spherical distribution. Moreover, since an elliptical distribution is obtained by a linear transform, it follows that the family of distributions $M_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \widehat{H})$ is a subclass of the elliptical distributions.

For instance, if $\mathbf{X} \sim N_d(\mathbf{0}, \mathbf{I}_d)$ it immediately follows that $R^2 \sim \chi^2_d$. Furthermore, we can use this result to obtain the distribution of the radial random variable R^2 for an uncorrelated normal mixture vector $\mathbf{X} \sim M_d(\mathbf{0}, \mathbf{I}_d, H)$. Given the stochastic representation $\mathbf{X} = \sqrt{W}\,\mathbf{Y}$ with $\mathbf{Y} \sim N_d(\mathbf{0}, \mathbf{I}_d)$ we obtain

$$R^2 \stackrel{d}{=} \mathbf{X}'\mathbf{X} \stackrel{d}{=} W\,\mathbf{Y}'\mathbf{Y} \stackrel{d}{=} W\,\chi^2_d.$$

If $\boldsymbol{\Sigma}$ is non-singular and provided that W has no point mass in the origin., the joint density function is given by the following expression

$$f(\mathbf{t}) = \int \frac{\omega^{-d/2}}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\Big\{ -\frac{(\mathbf{t}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{t}-\boldsymbol{\mu})}{2\omega}\Big\} dH(\omega),$$

where H is the distribution function of W.

Example A.10 (Multivariate k-point mixture distributions)**.** *If the support of the random variable W is finite, say with cardinality k then we say that the random vector $\mathbf{X}$ has a k-point mixture model. Obviously, when $k = 1$ this is simply the multivariate normal distribution.*

A random variable W has an inverse gamma distribution, denoted as $\mathrm{IG}(\alpha, \beta)$ if

$$1/W \sim \mathrm{Gamma}(\alpha, \beta), \qquad \alpha, \beta > 0.$$

Its mean and variance are given by $\mathbb{E}[W] = \beta/(\alpha - 1)$ $(\alpha > 1)$ and $\mathbb{V}\mathrm{ar}(W) = \beta^2/((\alpha - 1)^2(\alpha - 2))$ $(\alpha > 2)$.

Example A.11 (Multivariate t Distribution)**.** *If we take $W \sim IG(\nu/2, \nu/2)$, or equivalently $\nu/W \sim \chi^2_\nu$ (*chi square distribution with ν degrees of freedom*) then we say that the random vector $\mathbf{X}$ has a multivariate t distribution with ν degrees of freedom and denote it as $\mathbf{X} \sim t_d(\nu, \mu, \mathbf{\Sigma})$. Then it follows that $R^2/d \sim F(d, \nu)$, an F distribution with d and ν degrees of freedom, its covariance is given by*

$$\mathbb{C}\mathrm{ov}[\mathbf{X}] = \frac{\nu}{\nu - 2}\mathbf{\Sigma}, \qquad \nu > 2$$

and its joint density function by

$$f(\mathbf{t}) = \frac{\Gamma\big((\nu + d)/2\big)}{\Gamma(\nu/2)\,(\pi\nu)^{d/2}\,|\mathbf{\Sigma}|^{1/2}}\left(1 + \frac{(\mathbf{t} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{t} - \boldsymbol{\mu})}{\nu}\right)^{-(\nu+d)/2}.$$

A.2.3 Log-elliptical Distributions

Definition A.12. A random vector $\mathbf{X} := (X_1, \ldots, X_d)$ is said to have a *log-elliptical distribution* if

$$\log \mathbf{X} := (\log X_1, \ldots, \log X_d) \sim E_d(\boldsymbol{\mu}, \mathbf{\Sigma}, \phi),$$

with $\boldsymbol{\mu} \in \mathbb{R}^d$, $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$ a positive (semi) definite matrix and ϕ a characteristic generator. We denote it $\mathbf{X} \sim LE_d(\boldsymbol{\mu}, \mathbf{\Sigma}, \phi)$ where $\boldsymbol{\mu}$ is known as the localization parameter, $\mathbf{\Sigma}$ as the dispersion matrix and ϕ as the characteristic generator.

In particular a log-elliptical distribution will not be elliptical by itself. However, it inherits many properties of the elliptical distribution which make attractive this class of distributions. If $\log \mathbf{X} \sim E_d(\boldsymbol{\mu}, \mathbf{\Sigma}, \phi)$ has a joint density function, say $f(\mathbf{t})$, then it follows that the joint density of $\mathbf{X}$ will be given by

$$f(\log \mathbf{t})\left(\prod_{i=1}^{d} t_i\right)^{-1}.$$

For more details details on properties of log-elliptical distributions see Fang et al. (1987).

Example A.13 (Multivariate Lognormal Distribution)**.** We say that a vector has a *multivariate lognormal* distribution with localization vector μ and dispersion positive (semi)-definite matrix $\mathbf{\Sigma}$ if

$$\log \mathbf{X} \sim N_d(\mu, \mathbf{\Sigma}).$$

We will denote $\mathbf{X} \sim LN_d(\boldsymbol{\mu}, \mathbf{\Sigma})$. If the matrix $\mathbf{\Sigma}$ is nonsingular, then the joint density function of $\mathbf{X}$ is explicitly given by

$$f_{\mu,\mathbf{\Sigma}}(\mathbf{t}) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}\prod_{i=1}^{d} t_i}\exp\Big\{-\frac{(\log \mathbf{t} - \mu)'\mathbf{\Sigma}^{-1}(\log \mathbf{t} - \mu)}{2}\Big\}.$$

Moreover, its mean and covariance matrix are calculated as

$$\mathbb{E}[\mathbf{X}] = \mathrm{e}^{\boldsymbol{\mu} + \mathrm{diag}(\mathbf{\Sigma})/2}, \qquad \mathbb{C}\mathrm{ov}[\mathbf{X}]_{ij} = \mathrm{e}^{\mu_1 + \mu_j + (\sigma_i^2 + \sigma_j^2)/2}\big(\mathrm{e}^{\sigma_{ij}} - 1\big). \tag{A.6}$$

To make emphasis in the marginal behavior, we will refer to the elements of $\mathbf{X} \sim LN_d(\boldsymbol{\mu}, \mathbf{\Sigma})$ as *lognormal random variables with Gaussian copula.*

A.2.4 Copulas

The multivariate normal random distribution is considered the natural extension to a non-independent multidimensional setting of the normal distribution in the sense that the marginal distributions remain normal. However, in the general case there is not such a natural way to define a dependent version for a random variable. Copulas provide a method to create multivariate distributions with a rather general structure dependence and specific marginal distributions.

The idea is to separate the dependence structure from the univariate marginal behavior and represent it with a non-independent vector of uniform standard random variables. More precisely, let F be a multivariate distribution with univariate marginal distributions $F_1, \ldots, F_n$. The copula associated with F is defined as a distribution function $C : [0,1]^n \to [0,1]$ that satisfies

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_n(x_n)). \tag{A.7}$$

This result is usually known as *Sklar's Theorem*. Moreover, if $F_1^{-1}, \ldots, F_n^{-1}$ are the *right-continuous inverse* functions of $F_1, \ldots, F_n$, then

$$C(\mathbf{u}) = \left(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\right)$$

satisfies (A.7) and if F is continuous the solution is unique. This follows from the following Proposition:

Proposition A.14. *Let X be a random variable with cdf F and right-continuous inverse function F^{-1} and U a uniform standard random variable, then*

a) $F^{-1}(U)$ has c.d.f. F.

b) If F is continuous, then $F(X)$ is a uniform standard r.v.

A couple of basic examples of copulas are the *completely positively dependent copula* where $U_1 = \cdots = U_n$ and the *completely negatively dependent copula with $n = 2$* where $U_1 = 1 - U_2$. Two slightly more complicated (but still basic) examples of copulas are the *multivariate normal or Gaussian copula* and the *multivariate t copula* which are defined as follows. Let $\mathbf{Z} = (Z_1, \ldots, Z_n)$ be a multivariate normal (or multivariate t) random vector and $\Phi_i(\cdot)$ the marginal cdf of Z_i. The copula is defined as the distribution of the transformed vector

$$\Phi(\mathbf{Z}) := (\Phi_1(Z_1), \ldots, \Phi_n(Z_n)),$$

where $\Phi_i(Z_i)$ are non-independent uniform standard variables. These two copulas are widely used since they posses many useful properties which make them very flexible to work with. For more examples of copulas look at Joe (1997); Nelsen (2006).

The *tail dependence coefficient* describes the *amount* of dependence in the upper-quadrant tail or lower-quadrant tail of a bivariate distribution. The definition is given in terms of copulas because of the property of invariance to increasing transformations. Consider a bivariate copula C. Define the *Upper tail dependence coefficient* as

$$\lambda_U := \lim_{t \to 1} \frac{\overline{C}(t,t)}{1-t} = \lim_{t \to 1} \mathbb{P}(U_1 > t | U_2 > t).$$

A similar definition is given for the *Lower tail dependence coefficient* λ_L]. Note that λ_U, $\lambda_L \in [0,1]$. Then it is said that C has upper (lower) tail dependence if $\lambda_U \in (0,1]$ ($\lambda_L \in (0,1]$), and no upper (lower) dependence or tail independent if $\lambda_U = 0$ ($\lambda_L = 0$).

Appendix B

Simulation

B.1 Random Number Simulation

A *pseudo-random number generator* is a deterministic algorithm whose output is a stream of numbers that mimic the properties of a sequence of i.i.d. uniform standard r.v.'s. The efficient simulation of random variables from diverse distributions involve a set of methods and techniques based on transformations of these streams.

Many software packages such as Matlab, S-plus or Maple already contain large libraries for generating a wide variety of random numbers with great efficiency. The aim of this section is to describe some general algorithms for generating random numbers which are particularly useful when it is necessary to generate random numbers from less common distributions (usually not contained in these software packages).

B.1.1 Univariate Distributions

The most common methods used to generate independent random variables are *Inversion* and *Acceptance-Rejection*. For the Inversion method it is assumed that the right-continuous inverse F^{-1} of the c.d.f. is explicitly known, meanwhile, for the acceptance-rejection it is required the knowledge of the density function f plus a *similar* density g from which can be easily simulated. Both methods are sketched below.

Inversion. The way of simulate a random variable whose right-continuous inverse is known is to generate a uniform random variable and apply the right-continuous inverse. So, by part (a) of Proposition A.14, $F^{-1}(U)$ has the desired distribution. Note that almost every known discrete random variable can be generated with this method.

Acceptance-Rejection. Roughly speaking, by *similar* it is understood a random variable with density $g(\cdot)$, such that the minimum value of c such that $f(t) \leq cg(t)$ for all t in the support of X is close to 1. The inverse of the value that maximizes the ratio $f(t)/g(t)$ is the optimal value of c. The method is better understood from the algorithm itself.

Step 1. Generate Y.

Step 2. Generate a uniform standard U.

Step 3. If $U \leq f(Y)cg(Y)$, let $X := Y$. Otherwise return to Step 1.

The following argument shows that X has the right density f or equivalently that the algorithm produces random variables with the desired distribution. Let $\{Y_n\}$ a sequence of i.i.d. random variables with density $g(\cdot)$ and $p_A = \mathbb{P}(\text{Accept})$, then

$$\begin{aligned}
\mathbb{P}(X \leq x) &= \sum_{n=0}^{\infty} \mathbb{P}(Y_n \leq x|\text{Accept}) p_A (1-p_A)^n \\
&= \mathbb{P}(Y_1 \leq x|\text{Accept}) \sum_{n=0}^{\infty} p_A (1-p_A)^n = \mathbb{P}(Y_1 \leq x|\text{Accept}) \\
&= \frac{\mathbb{P}(Y_1 \leq x; \text{Accept})}{\mathbb{P}(\text{Accept})} = \frac{\int_{-\infty}^{x} \mathbb{P}(\text{Accept}|Y_1 = t) g(t) dt}{\int_{\infty}^{\infty} \mathbb{P}(\text{Accept}|Y_1 = t) g(t) dt} \\
&= \frac{\int_{-\infty}^{x} f(t) c g(t) g(t) dt}{\int_{-\infty}^{\infty} f(t) c g(t) g(t) dt} = \frac{\frac{1}{c}\int_{-\infty}^{x} f(t) dt}{\frac{1}{c}\int_{-\infty}^{\infty} f(t) dt} = \int_{-\infty}^{x} f(t) dt.
\end{aligned}$$

From this argument it is observed that $p_A = 1/c$. The speed of this algorithm is based in the rate of acceptance. Then, it is desirable to find a density g whose optimal value of c is as close as 1 as possible or intuitively that it looks as much alike to f as possible.

B.1.2 Multivariate Distributions

For realistic simulation it is more useful to generate non-independent random variables. However, many difficulties arise from the dependence structure and this problem becomes more difficult: The inversion method cannot be generalized to simulate dependent random variables. If the joint density is available then it is possible to use acceptance-rejection. However, it becomes very inefficient in large dimension settings due to the high probability of rejection. Also the choice of a *similar* density becomes less obvious.

In a few cases the properties of the distribution allow to make only a few transformations and operations with independent random variables to simulate from the desired distribution.

Spherical and Elliptical Distributions The algorithm for simulating an elliptical random vector is given next.

1. Simulate $\mathbf{Z} := (Z_1, \ldots, Z_d)$ where the components are independent normal standard random variables. Simulate the radial random variable R with representation ψ.
2. Find a square matrix $\mathbf{A}$ such that $\mathbf{A}^T\mathbf{A} = \boldsymbol{\Sigma}$.
3. Return
$$\boldsymbol{\mu} + R\,\mathbf{A}\,\frac{\mathbf{Z}}{\|\mathbf{Z}\|}.$$

The last random vector has the correct distribution since $\mathbf{C} \stackrel{d}{=} \mathbf{Z}/\|\mathbf{Z}\|$ is uniformly distributed over the unit spheroid $\mathcal{S}_1(\mathbf{0})$ – recall that the level curves of a multivariate normal random vector with marginal distributions $N(0,1)$ form spheroids in $\mathbb{R}^d$.

In the case of the multivariate normal and multivariate t distributions we can avoid the normalization of the normal vector. For the multivariate normal the new radial

random variable is taken to be a constant while in the case of the multivariate t, it an inverse gamma distribution is used.

Copulas A wide variety of dependent random variables can be generated using copulas. Since the copulas separate the dependence structure from the marginal behavior it is necessary to be able to simulate from the copula (acceptance-rejection could be used) and also from the marginal distribution.

A simple example is to export the dependence structure from the multivariate normal or multivariate t to other marginal distributions. The following method generates random variables with copula C and marginal distributions $G_i(\cdot)$. It is assumed that a method for simulating from C is readily available. Simulate $(X_1, \ldots, X_n)^T$ from a C and apply the transformation

$$\left(G_1^{-1}(X_1), \ldots, G_n^{-1}(X_n)\right)$$

where $G_i^{-1}(\cdot)$ is the right-continuous inverse of $G_i(\cdot)$.

B.2 Monte Carlo Estimators and Variance Reduction

B.2.1 Crude Monte Carlo

Let X be a random variable with cdf F. Suppose that it is of interest to calculate

$$\theta = \mathbb{E}\left[h(X)\right] = \int h(t)F(dt)$$

for some specific function h with domain in the support of X. It is often the case, that this integral has not a closed analytic form, so, we turn to simulation to approximate the value θ. The idea is to generate a sequence $X_1, \ldots, X_n$ of i.i.d. random variables and estimate θ with

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} h(X_1).$$

This estimator is called *crude Monte Carlo (CMC)*. It is unbiased by independence and by the strong law of large numbers it is strongly consistent. Let $\theta(x)$ be a function defined in $\mathbb{R}$. An *algorithm* is defined to be a family of estimators $\{Z(x)\}$ such that $\mathbb{E}\, Z(x) = \theta(x)$ for all x. Monte Carlo estimators are useful to approximate the probability of an event A. In particular, if $h(x) = \mathbb{I}_A(x)$, then

$$\mathbb{P}(A) \approx \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_A(X_i).$$

The variance of the CMC estimator is given by

$$\frac{\mathbb{V}\mathrm{ar}\, h(X_i)}{n}.$$

It is desirable to obtain an alternative estimator with smaller variance than the one produced by the CMC method. Variance reduction methods have this purpose, but usually require extra theoretical work and programming effort, so it is only worthwhile if it provides a substantial variance reduction. Among the most common methods are Importance Sampling, Stratification and Conditional Monte Carlo. These are briefly described below.

B.2.2 Importance Sampling

The following is a well-known consequence of the Radon-Nykodym Theorem: Let ν be a finite and absolutely continuous measure with respect to σ-finite measure μ, and h is any integrable function with respect to ν, then

$$\int h d\nu = \int h \frac{d\nu}{d\mu} d\mu,$$

where the function $d\nu/d\mu$ (written in Leibnizian notation) is the Radon-Nykodym derivative of ν with respect to μ. When ν and μ are probability measures and the densities are available, say $f(\cdot)$ and $g(\cdot)$, the last equality becomes

$$\int h(t)F(dt) = \int h(t) \frac{f(t)}{g(t)} G(dt),$$

with the condition that the support of g is contained in the support of f. The idea is to simulate from a density g instead of f which concentrates more mass around the real value of the expected value than f and, in consequence, obtain a variance reduction. The estimator is given by

$$h(X) \frac{f(X)}{g(X)},$$

where X is simulated from the density g. By the arguments shown above it is seen that such estimator is unbiased but it is not evident if it provides a variance reduction and the main problem becomes to make an efficient choice of g.

B.2.3 Stratification

Let Z be a random variable defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The stratification method for estimating $z := \mathbb{E}\, Z$ consists in defining a partition of the sample space Ω, say $\Omega_1, \ldots, \Omega_s$ (*strata*), such that the probabilities $p_i := \mathbb{P}(\Omega_i)$ are known and we can build estimators $\widehat{z}_i$ for every $z_i := \mathbb{E}\,[Z|\Omega_i]$. The stratification estimator comes out as

$$\widehat{z}_S := \sum_{i=1}^{s} p_i \widehat{z}_i.$$

The idea of the stratification method consists in eliminating the variance among the strata. In principle, it always produces a variance reduction if we use *proportional allocation*. That simply means that the number of replications R_i used to generate every $\widehat{z}_i$ should be taken such that $R_i / \sum R_i$ is proportional to p_i. In such case

$$\mathbb{V}\mathrm{ar}\, \widehat{z}_S = \sum_{i=1}^{s} p_i \mathbb{V}\mathrm{ar}\,(Z|\Omega_i).$$

A further characteristic of the stratification method is that, since

$$\mathbb{E}\, \widehat{z}_S = \sum_{i=1}^{s} \mathbb{E}\,(Z|\Omega_i),$$

the variance coming from a set Ω_i can be completely removed if $\mathbb{E}\,(Z|\Omega_1)$ is known in closed form.

B.2.4 Conditional Monte Carlo

Suppose that we want to estimate a quantity z and for doing so we already count with a random variable Z defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $z = \mathbb{E}(Z)$. *Conditional Monte Carlo* is a tool which improves the efficiency of a given estimator by conditioning it to a particular σ-algebra $\mathcal{G} \subset \mathcal{F}$. That is,

$$\widehat{z} = \mathbb{E}(Z|\mathcal{G}),$$

This estimator is clearly unbiased and by the Rao-Blackwell Theorem it follows that

$$\mathbb{V}\text{ar}(\widehat{z}) = \mathbb{V}\text{ar}(\mathbb{E}(Z|\mathcal{G})) \leq \mathbb{V}\text{ar}(Z).$$

Therefore, it always provides a variance reduction. As discussed before the problem is to choose a a σ-algebra $\mathcal{F}$ such that it is possible to simulate the events in $\mathcal{F}$ and the conditional expectation $\mathbb{E}(S_n|\mathcal{F})$ is known in closed form. Obviously, it is also desired that the resulting estimator provides a substantial variance reduction.

Appendix C

Lévy Processes

In this appendix we give the basic definition of a Lévy process plus some fundamental characterizations and properties which are considered to be enough for the needs of this dissertation. It also may serve as illustrative of the richness of the class of Lévy process. Standard references for Lévy processes are Bertoin (1996); Sato (1999); Kyprianou (2006).

Definition C.1 (Lévy Process)**.** A stochastic process $X(t)$ defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a Lévy process if

1. The process $X(t)$ is *càdlàg*. That means that it has sample paths which are right continuous with left limits a.s.

2. The process is started at 0 with probability 1.

3. The process is stationary. That is, $X(t-s) \stackrel{d}{=} X(t) - X(s)$ for all $0 < s < t$.

4. The process has independent increments. This means $X(t) - X(s)$ is independent of $\{X(u) : u \leq s\}$.

The two basic examples of a Lévy process are the Poisson process and the Brownian motion. Some other well known processes which are *Lévy processes* are the compound Poisson processes, Gamma processes, Normal Inverse Gaussian processes and Stable processes. The following definition will be useful to characterize *Lévy processes* and relate them to the class of *infinitely divisible* distributions which are defined later.

Definition C.2 (Characteristic exponent of a Lévy process)**.** For a Lévy process $X(t)$ we define

$$\Psi(\theta) := -\log \mathbb{E}\left[\mathrm{e}^{-i\theta X(1)}\right],$$

and say that $\Psi(\theta)$ is the *characteristic exponent* of the process $X(t)$.

The following Theorem provides a characterization of Lévy process in terms of its *characteristic exponent*.

Theorem C.3 (Lévy-Khinchine Theorem for Lévy processes)**.** *Let $a \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$ and Π a measure with no point mass in the origin and such that*

$$\int_{\mathbb{R}} \left(\epsilon \wedge t^2\right) \Pi(dt) < \infty,$$

for some arbitrary value $\epsilon > 0$ *and define*

$$\Psi(\theta) := ia\theta + \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}} (1 - \mathrm{e}^{i\theta t} + i\theta t \mathbb{I}(|t| < \epsilon))\Pi(dx).$$

Then there exists a Lévy process $X(t)$ *whose characteristic exponent is given by* $\Psi(\theta)$.

The choice of ϵ can be rather arbitrary but in the literature is commonly taken to be 1. The class of Lévy processes it is closely related with the class of *infinitely divisible* distributions which are defined as follows.

Definition C.4 (Infinitely Divisibility)**.** We say that a random variable Y has an infinitely divisible distribution if for all $n \in \mathbb{N}$ there exist a set of i.i.d. random variables, say $Y_{1,n}, \ldots, Y_{n,n}$, such that Y admits the stochastic representation

$$Y \stackrel{d}{=} Y_{1,n} + \cdots + Y_{n,n}.$$

In terms of the characteristic function $\Phi(\theta)$, this just means that for all n there exist a characteristic exponents $\Phi_n(\theta)$ such that $\Phi(\theta) = n\Phi_n(\theta)$. Moreover, a similar *Lévy-Khinchine* Theorem can be proved for infinitely divisible distributions. In fact, this shows that the the increments of a Lévy process follow an infinitely divisible distribution. Conversely, for any given a Lévy process can be constructed from any infinitely divisible distribution.

We close this appendix with the *Lévy-Ito* decomposition which identifies the three key ingredients of any Lévy process. We can use the *Lévy-Khinchine* formula to rewrite the characteristic exponent of any *Lévy process* as the sum of the following three terms

$$\begin{aligned}
\Psi_1(\theta) &:= ia\theta + \frac{1}{2}\sigma^2\theta^2, \\
\Psi_2(\theta) &:= \int_{|t|>\epsilon} (1 - \mathrm{e}^{i\theta t})\Pi(dt), \\
\Psi_3(\theta) &:= \int_{0<|t|<\epsilon} (1 - \mathrm{e}^{i\theta t} + i\theta t)\Pi(dt),
\end{aligned}$$

The *Lévy-Ito* decomposition shows that each of these functions can be identified as the characteristic exponent of three different types of Lévy processes. The function Ψ_1 corresponds to the characteristic exponent of a Brownian motion with drift. After some rewriting, Ψ_2 is identified as the characteristic function of a compound Poisson process with jumps bounded below by ϵ. The remaining element Ψ_3 is a very interesting one. In fact, here we have chosen to take an arbitrary value ϵ (rather than the common practice of taking it equal to one) to give an heuristic argument of the nature of this process. Note that similar decompositions are obtained by choosing taking different values of ϵ in the *Lévy-Khinchine* formula. The idea is to take a decreasing sequence of values $\epsilon_1, \epsilon_2, \ldots$ such that $\epsilon_i \to 0$. Therefore, the function Ψ_3 can be interpreted as the characteristic exponent of the superposition of a countable number of independent compound Poisson processes with jumps sizes in $[\epsilon_{i+1}, \epsilon_i)$. Formally speaking, this process is a square integrable martingale with an almost surely countable number of jumps on a finite time interval which are bounded above by ϵ.

Appendix D

Numerical Examples

The numerical results of the methods discussed in this dissertation had been collected in this appendix. It has been divided in sections where a brief description is included. The main discussion of the results was kept in the corresponding chapters.

D.1 Tail Probabilities of Sums of Lognormal Random Variables

The Subsection D.1.1 contains the aymptotic approximations while the Subsection D.1.2 on page 106 contains the Monte Carlo estimates developed throughout the dissertation.

D.1.1 Asymptotic Approximations

The following set of examples show the numerical results of the asymptotic approximations discussed in Section 3 on page 41. These are:

1. The subexponential-type (**ST**) approximation.
2. The approximation via the aggregated tails (**AT**).

For comparison purposed we have taken the best Monte Carlo estimate in terms of the lowest variance. As a comparison measure I have included the *absolute relative difference* which is defined as the absolute value of the difference between the approximation and the Monte Carlo estimate divided by the Monte Carlo estimate. The idea is to measure the speed of convergence of the approximations.

Example D.1. For this example we have used $n = 10$ independent lognormal random variables such that the i-th random variable has a $LN(-i, i)$ distribution.

Method	Approximation	Relative Difference
Monte Carlo estimate	1.021341e-03	
Subexponential type	9.448482e-04	7.489502e-02
Aggregated tails	1.107537e-03	8.439438e-02

Table D.1: $\mathbb{P}(S_{10} > 20000)$.

Method	Approximation	Relative Difference
Monte Carlo estimate	6.457905e-04	
Subexponential type	6.025234e-04	6.699864e-02
Aggregated Tails	6.967287e-04	7.887727e-02

Table D.2: $\mathbb{P}(S_{10} > 30000)$.

Method	Approximation	Relative Difference
Monte Carlo estimate	4.624322e-04	
Subexponential type	4.339349e-04	6.162468e-02
Aggregated Tails	4.972447e-04	7.528137e-02

Table D.3: $\mathbb{P}(S_{10} > 40000)$.

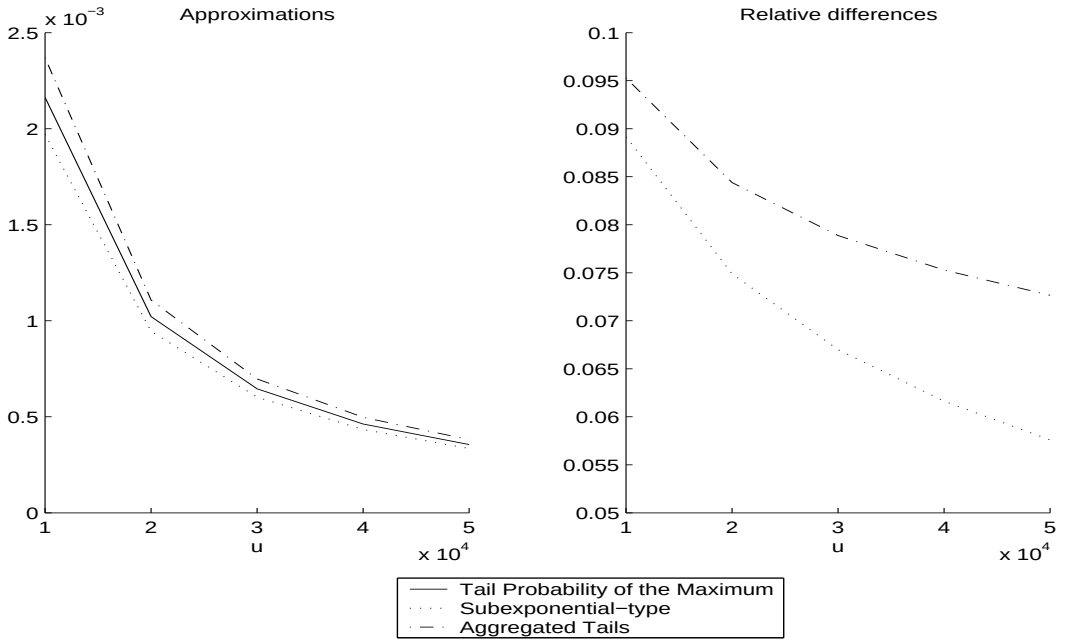


Figure D.1: Asymptotic approximations for the tail probability of a sum of independent lognormal random variables.

Example D.2. For this example we have used $n = 10$ correlated lognormal random variables such that the i-th random variable has a $LN(-i, i)$ distribution and common correlation parameter $\rho = 0.4$.

Method	Approximation	Relative Difference
Monte Carlo estimate	2.140006e-03	
Subexponential type	1.970454e-03	7.922999e-02
Aggregated Tails	2.369424e-03	1.072044e-01

Table D.4: $\mathbb{P}(S_{10} > 20000)$.

Method	Approximation	Relative Difference
Monte Carlo estimate	1.014621e-03	
Subexponential type	9.448482e-04	6.876803e-02
Aggregated Tails	1.107537e-03	9.157635e-02

Table D.5: $\mathbb{P}(S_{10} > 30000)$.

Method	Approximation	Relative Difference
Monte Carlo estimate	6.421017e-04	
Subexponential type	6.025234e-04	6.163859e-02
Aggregated tails	6.967287e-04	8.507537e-02

Table D.6: $\mathbb{P}(S_{10} > 40000)$.

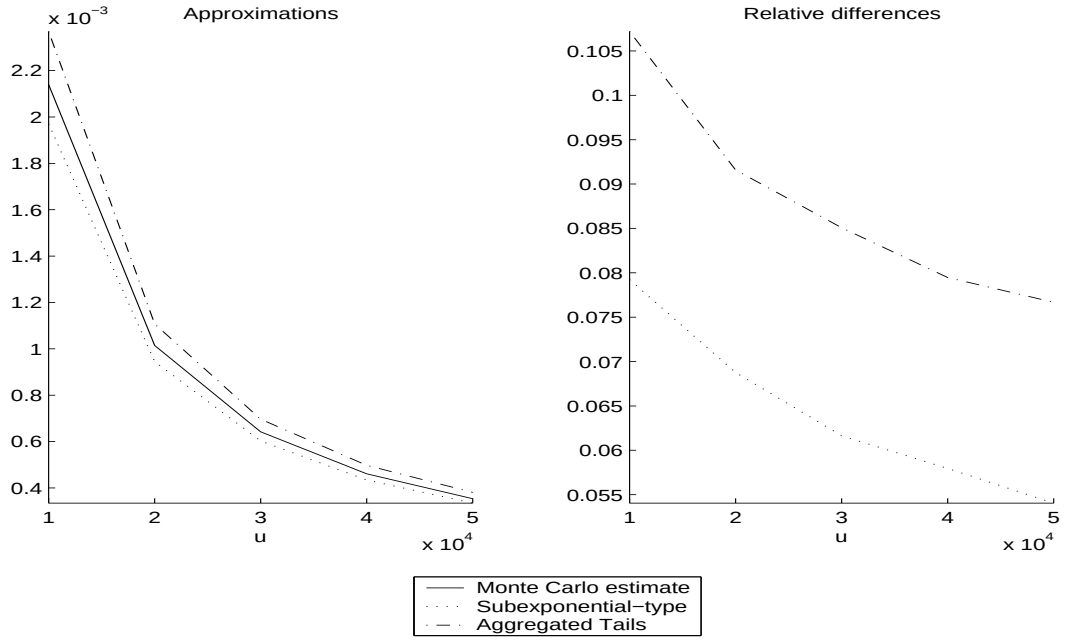


Figure D.2: Asymptotic approximations for the tail probability of a sum of correlated lognormal random variables (medium correlation).

Example D.3. For this example we have used $n = 10$ correlated lognormal random variables such that the i-th random variable has a $LN(-i, i)$ distribution and common correlation parameter $\rho = 0.9$.

Method	Approximation	Relative Difference
Monte Carlo estimate	1.021341e-03	
Subexponential type	9.448482e-04	7.489502e-02
Aggregated tails	1.107537e-03	8.439438e-02

Table D.7: $\mathbb{P}(S_{10} > 20000)$.

Method	Approximation	Relative Difference
Monte Carlo estimate	6.457905e-04	
Subexponential type	6.025234e-04	6.699864e-02
Aggregated Tails	6.967287e-04	7.887727e-02

Table D.8: $\mathbb{P}(S_{10} > 30000)$.

Method	Approximation	Relative Difference
Monte Carlo estimate	4.624322e-04	
Subexponential type	4.339349e-04	6.162468e-02
Aggregated Tails	4.972447e-04	7.528137e-02

Table D.9: $\mathbb{P}(S_{10} > 40000)$.

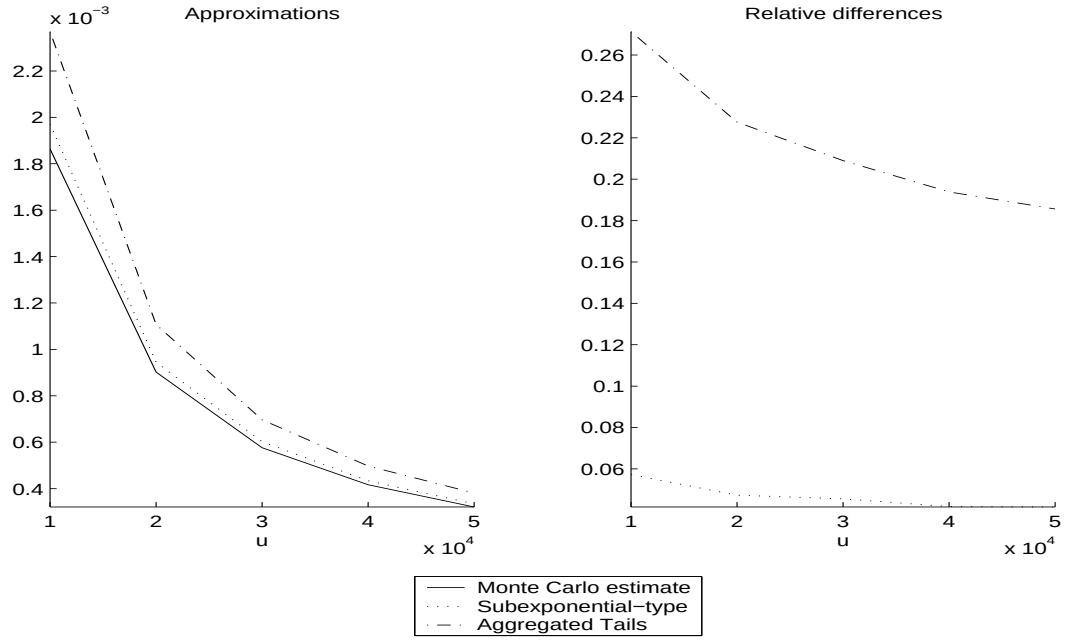


Figure D.3: Asymptotic approximations for the tail probability of a sum of correlated lognormal random variables (high correlations).

D.1.2 Monte Carlo Estimates

The following set of examples show numerical comparisons of the Monte Carlo estimators discussed across the dissertation. These are:

1. The modified *Asmussen-Kroese* algorithm (**AK**) which have been described in subsections 1.3.3 on page 12 and 3.2.2 on page 48.

2. The Importance sampling algorithm based in the Scaling of the Variance (**SV**) and its Cross-Entropy version (**SVCE**) discussed in subsection 3.2.1 on page 44.

3. The conditional algorithm designed for Log-elliptical distributions based on the radial random variable (**RE**) which was discussed in Section 4.2 on page 61.

4. Finally every estimator was improved by using the stratification strategy described in Subsection 3.2.1.2 on page 47 which are distinguished by adding **S-** before the abbreviation of the names of the algorithms. In the case of the graphs and $*$ had been added to the style of the line.

The results are displayed in tables as well as in graphs. The tables show the estimates for every estimator, the standard deviation, the variation coefficient and the cpu time needed to generate the total number of replications. We have used $R = 50,000$ with the purpose of obtaining better estimates of the variance of the estimates, however, it should be noted that fairly precise estimates might be obtained with less replications. The figures show the variation coefficient of every algorithm plus the Time-relative error which is consider to be a more fair measure of the performance of an algorithm since it takes into account the time needed for generating a single replication. It is simple obtained by multiplying the variation-coefficient by the cpu time.

Example D.4. For this example we have used $n = 10$ independent lognormal random variables such that the i-th random variable has a $LN(-i, i)$ distribution.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
SV	1.084545e-03	1.752382e-02	1.615775e 01	0.2380
SVCE	1.030550e-03	9.121518e-03	8.851113e 00	3.3669
AK	1.023973e-03	1.611999e-05	1.574258e-02	0.1460
RE	9.922026e-04	9.050250e-03	9.121372e 00	0.2219
S-SV	1.024992e-03	4.046023e-04	3.947370e-01	0.3249
S-SVCE	1.022912e-03	4.982704e-05	4.871095e-02	0.2850
S-AK	1.021341e-03	0.000000e-00	0.000000e 00	0.1460
S-RE	1.023783e-03	3.561371e-05	3.478635e-02	3.3730

Table D.10: $\mathbb{P}(S_{10} > 20000)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
SV	5.205030e-04	9.924635e-03	1.906739e 01	0.2379
SVCE	4.110338e-04	4.557306e-03	1.108742e 01	3.3549
AK	4.632485e-04	9.821872e-06	2.120216e-02	0.1400
RE	4.845668e-04	7.215657e-03	1.489094e 01	0.2190
S-SV	4.625195e-04	1.402947e-05	3.033271e-02	0.3469
S-SVCE	4.629841e-04	3.125797e-05	6.751413e-02	0.2749
S-AK	4.624322e-04	0.000000e-00	0.000000e 00	0.1409
S-RE	4.630605e-04	1.066126e-05	2.302347e-02	3.3740

Table D.11: $\mathbb{P}(S_{10} > 40000)$.

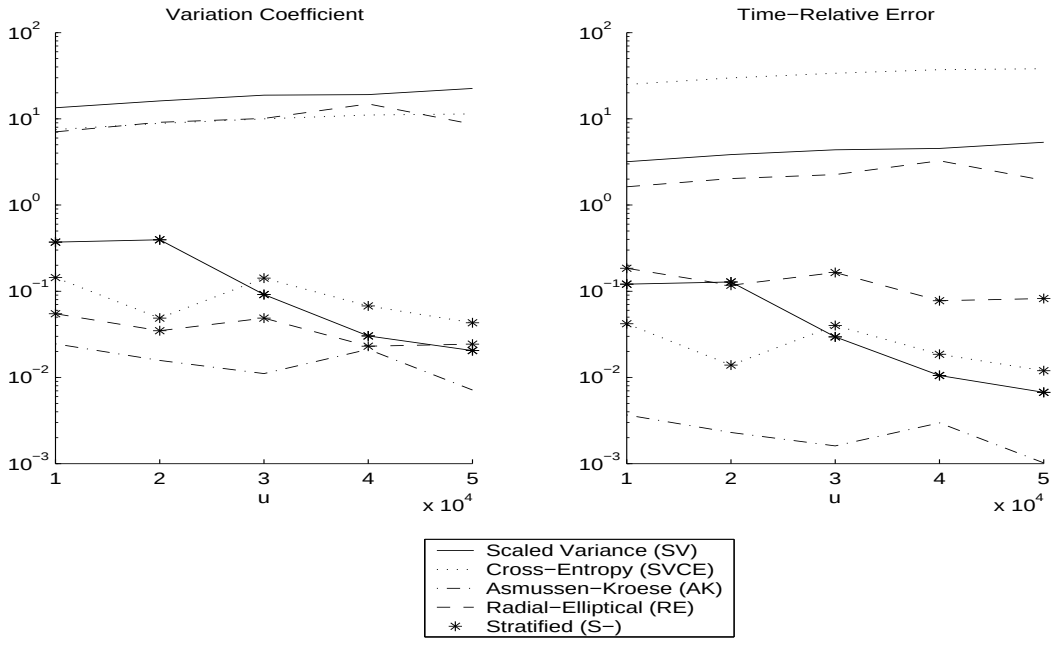


Figure D.4: Monte Carlo estimates for the tail probability of a sum of independent lognormal random variables.

Example D.5. For this example we have used $n = 10$ correlated lognormal random variables such that the i-th random variable has a $LN(-i, i)$ distribution and common correlation parameter $\rho = 0.4$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
SV	9.838832e-04	1.544085e-02	1.569378e 01	0.2400
SVCE	1.132847e-03	9.824999e-03	8.672837e 00	3.1170
AK	1.044250e-03	3.649722e-03	3.495064e 00	0.1520
RE	1.052730e-03	5.787320e-03	5.497438e 00	0.2200
S-SV	1.053528e-03	2.850599e-03	2.705765e 00	2.1970
S-SVCE	1.042666e-03	6.258719e-04	6.002610e-01	2.1590
S-AK	1.014621e-03	3.923703e-05	3.867159e-02	2.0320
S-RE	1.049404e-03	4.191588e-04	3.994253e-01	5.0560

Table D.12: $\mathbb{P}(S_{10} > 20000)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
SV	5.475374e-04	1.142151e-02	2.085979e 01	0.2630
SVCE	4.663640e-04	5.075465e-03	1.088305e 01	3.1490
AK	4.744569e-04	1.928983e-03	4.065665e 00	0.1520
RE	4.742387e-04	3.310505e-03	6.980673e 00	0.2200
S-SV	4.696919e-04	5.916554e-04	1.259667e 00	2.2970
S-SVCE	4.768497e-04	5.454482e-04	1.143857e 00	2.2580
S-AK	4.606356e-04	1.368913e-05	2.971791e-02	2.1310
S-RE	4.720674e-04	1.812297e-04	3.839063e-01	5.1550

Table D.13: $\mathbb{P}(S_{10} > 40000)$.

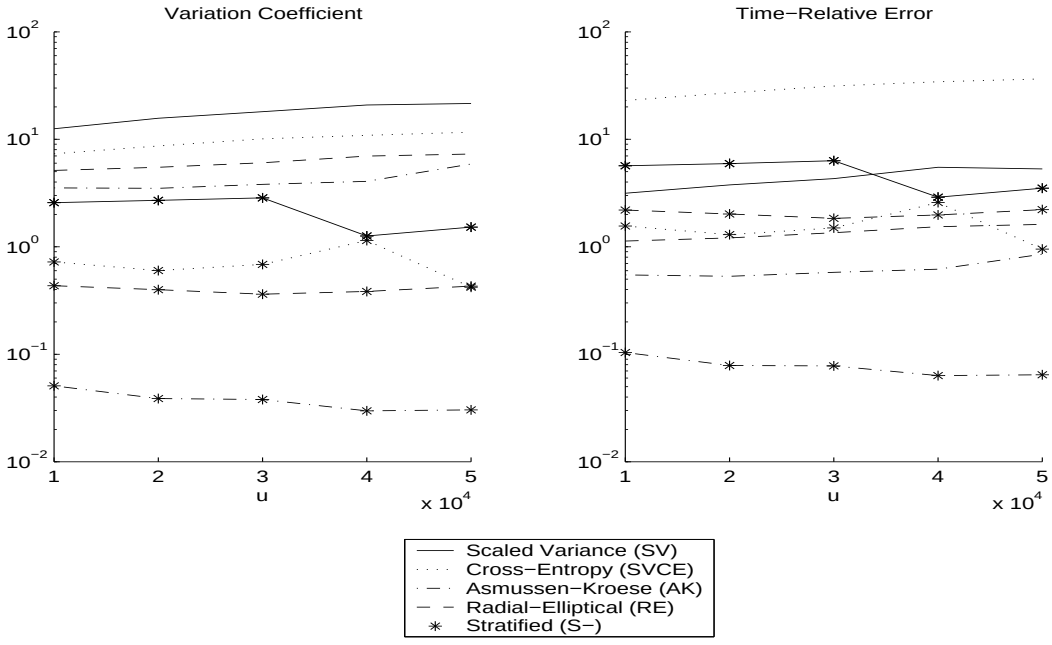


Figure D.5: Monte Carlo estimates for the tail probability of a sum of correlated lognormal variables (medium correlation).

Example D.6. For this example we have used $n = 10$ correlated lognormal random variables such that the i-th random variable has a $LN(-i, i)$ distribution and common correlation parameter $\rho = 0.9$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
SV	1.056406e-03	1.770722e-02	1.676176e 01	0.2210
SVCE	1.103081e-03	1.010449e-02	9.160247e 00	2.5140
AK	1.296232e-03	2.414212e-02	1.862484e 01	0.1450
RE	1.118677e-03	5.263256e-03	4.704890e 00	0.2270
S-SV	1.112304e-03	9.385298e-03	8.437705e 00	2.2250
S-SVCE	1.141068e-03	2.945288e-03	2.581167e 00	2.1900
S-AK	9.021938e-04	1.515841e-04	1.680173e-01	2.0530
S-RE	1.137749e-03	2.258961e-03	1.985465e 00	4.3380

Table D.14: $\mathbb{P}(S_{10} > 20000)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
SV	8.891581e-04	1.731132e-02	1.946933e 01	0.2380
SVCE	7.326583e-04	7.435824e-03	1.014910e 01	2.4210
AK	7.206390e-04	1.596332e-02	2.215162e 01	0.1430
RE	7.143845e-04	3.607646e-03	5.050006e 00	0.2180
S-SV	7.667911e-04	8.964689e-03	1.169117e 01	2.4220
S-SVCE	7.337671e-04	1.959899e-03	2.671010e 00	2.3850
S-AK	5.762806e-04	9.371145e-05	1.626142e-01	2.2430
S-RE	7.365516e-04	1.699989e-03	2.308037e 00	4.5300

Table D.15: $\mathbb{P}(S_{10} > 40000)$.

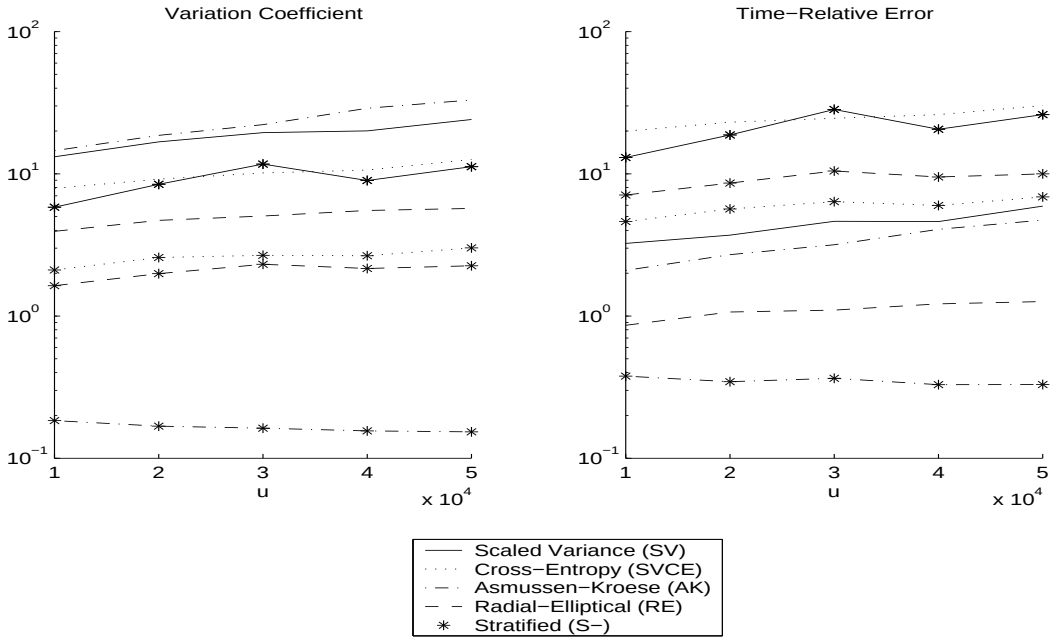


Figure D.6: Monte Carlo estimates for the tail probability of a sum of correlated lognormal random variables (high correlations).

D.2 Ruin Probabilities in a Finite Time Horizon

The following set of examples show some numerical results of the methods for approximating the ruin probability in a finite time horizon which where discussed in the chapter 2. These are:

1. The Asmussen-Klüppelberg approximation (**AK**) given in 2.3 on page 28.
2. The Foss-Palmowski-Zachary approximation (**FPZ**) given in 2.4 on page 28.
3. The conditional Monte Carlo estimate designed for the Cramér-Lundberg risk model (**CL**) discussed in 2.3.1 on page 29.
4. The conditional Monte Carlo estimate designed for the Lévy risk model (**L**) discussed in 2.3.2 on page 32.

The results are displayed in tables and graphics. The tables show the value of the approximations and in the case of the estimators it also shows the standard deviation, the variation coefficient and the cpu time used to generate the whole amount of replications. Here we have used $R = 5000$.

We have chosen a Cramér-Lundberg risk model where the intensity of the compound Poisson process is λ and the claims follow a Pareto distribution with parameter α. Therefore the *stability condition* is translated into

$$\frac{\lambda}{\alpha - 1} < 1.$$

In the first two examples we have considered medium time horizon, so we have taken $t(u) = u$ while in the third and fourth examples we consider a shorter time horizon and we have chosen $t(u) = u^\gamma$ where $\gamma = \alpha/(1 + \alpha)$. We made this choice accordingly to the restriction in Theorem 2.7 on page 34 where we proved that the proposed estimator has bounded relative error if $\gamma < \alpha/(1 + \alpha)$.

Also, in the first and third example we have experimented with a very heavy-tail, while in the second and fourth we have taken a moderately heavy-tail. Recall that Pareto distributions becomes heavier as long as α gets smaller.

Example D.7. In this example we choose the intensity of the Poisson process $\lambda = 1/3$, a Pareto distribution with index $\alpha = 3/2$ and a time horizon $t(u) = u$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	8.473255e-04			
FPZ	1.054076e-05			
CL	8.757389e-04	5.757320e-03	6.574242e 00	102.6180
L	8.496901e-04	1.235509e-04	1.454071e-01	164.0500

Table D.16: $\Psi(100000, 100000)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	6.918395e-04			
FPZ	7.315562e-06			
CL	7.215334e-04	5.158963e-03	7.149999e 00	154.5790
L	6.950262e-04	1.000262e-04	1.439172e-01	260.6370

Table D.17: $\Psi(150000, 150000)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	5.991511e-04			
FPZ	5.649756e-06			
CL	5.788303e-04	4.352878e-03	7.520129e 00	219.7720
L	5.984339e-04	8.338152e-05	1.393328e-01	384.7240

Table D.18: $\Psi(200000, 200000)$.

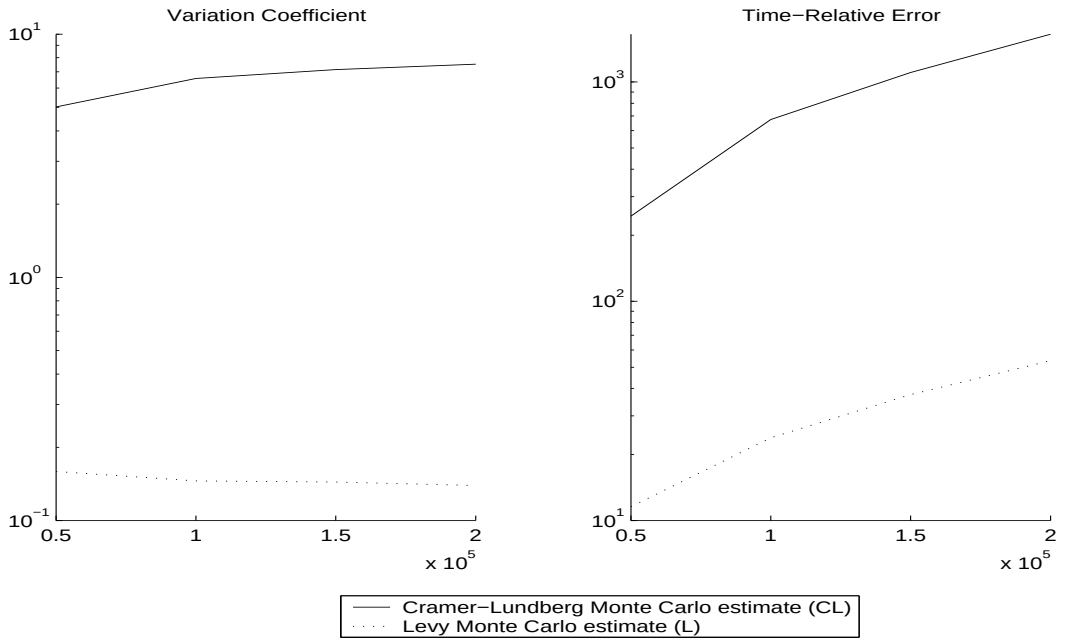


Figure D.7: Approximation for the ruin probability in a medium time horizon for a Cramér-Lundberg risk model with very heavy Pareto claims.

Example D.8. In this example we choose the intensity of the Poisson process $\lambda = 3/2$, a Pareto distribution with index $\alpha = 3$ and a time horizon $t(u) = u$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	4.152249e-04			
FPZ	2.148494e-04			
CL	5.472923e-04	1.775701e-03	3.244521e 00	0.6540
L	5.103081e-04	3.593041e-04	7.040925e-01	0.7700

Table D.19: $\Psi(50, 50)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	1.869806e-04			
FPZ	8.542608e-05			
CL	2.281770e-04	8.409871e-04	3.685677e 00	0.8650
L	2.166948e-04	1.313529e-04	6.061655e-01	1.1950

Table D.20: $\Psi(75, 75)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	1.058719e-04			
FPZ	4.658832e-05			
CL	1.097071e-04	4.627397e-03	4.217953e 00	1.0190
L	1.181029e-04	5.551590e-04	4.700637e-01	1.5310

Table D.21: $\Psi(100, 100)$.

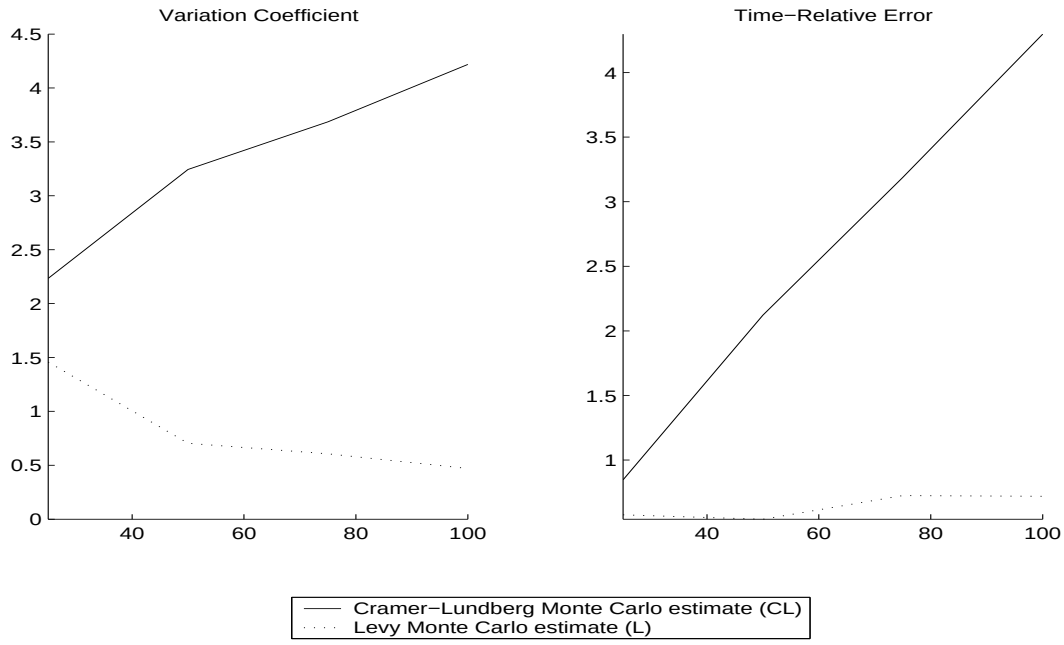


Figure D.8: Approximation for the ruin probability in a medium time horizon for a Cramér-Lundberg risk model with moderately heavy Pareto claims.

Example D.9. In this example we choose the intensity of the Poisson process $\lambda = 1/3$, a Pareto distribution with index $\alpha = 3/2$ and a time horizon $t(u) = u^{3/5}$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	1.198293e-03			
FPZ	1.967680e-05			
CL	1.899330e-05	4.342040e-05	2.286089e 00	0.7140
L	1.963627e-05	1.402237e-06	7.141056e-02	1.0970

Table D.22: $\Psi(100000, 1000)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	8.473255e-04			
FPZ	1.054076e-05			
CL	1.011332e-05	2.432811e-05	2.405551e 00	1.0260
L	1.052123e-05	5.809919e-07	5.522087e-02	1.4760

Table D.23: $\Psi(150000, 1275)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	6.918395e-04			
FPZ	7.315562e-06			
CL	6.999786e-06	1.889320e-05	2.699111e 00	1.2850
L	7.300557e-06	3.554226e-07	4.868431e-02	1.8030

Table D.24: $\Psi(200000, 1516)$.

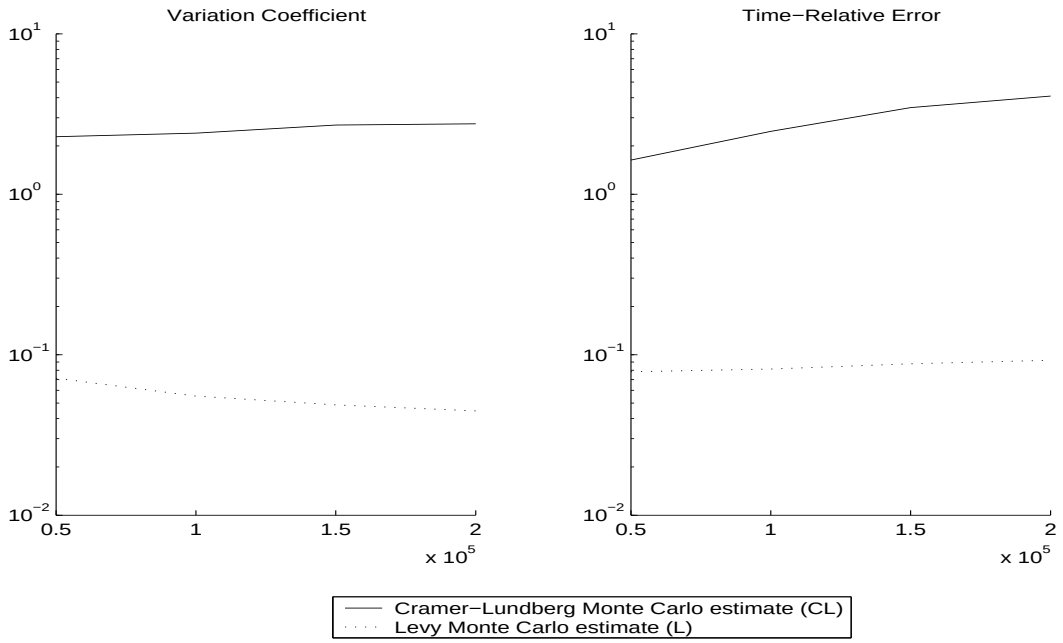


Figure D.9: Approximation for the ruin probability in a short time horizon for a Cramér-Lundberg risk model with very heavy Pareto claims.

Example D.10. In this example we choose the intensity of the Poisson process $\lambda = 3/2$, a Pareto distribution with index $\alpha = 3$ and a time horizon $t(u) = u^{3/4}$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	4.152249e-04			
FPZ	2.148494e-04			
CL	2.225946e-04	4.817529e-04	2.164260e 00	0.3200
L	2.142031e-04	1.095131e-04	5.112580e-01	0.2940

Table D.25: $\Psi(50, 19)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	1.869806e-04			
FPZ	8.542608e-05			
CL	8.195059e-05	1.866049e-04	2.277042e 00	0.4090
L	8.282271e-05	2.866981e-05	3.461587e-01	0.4150

Table D.26: $\Psi(75, 25)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
AK	1.058719e-04			
FPZ	4.658832e-05			
CL	4.596047e-05	1.268075e-04	2.759056e 00	0.4490
L	4.480293e-05	1.447567e-05	3.230965e-01	0.4980

Table D.27: $\Psi(100, 32)$.

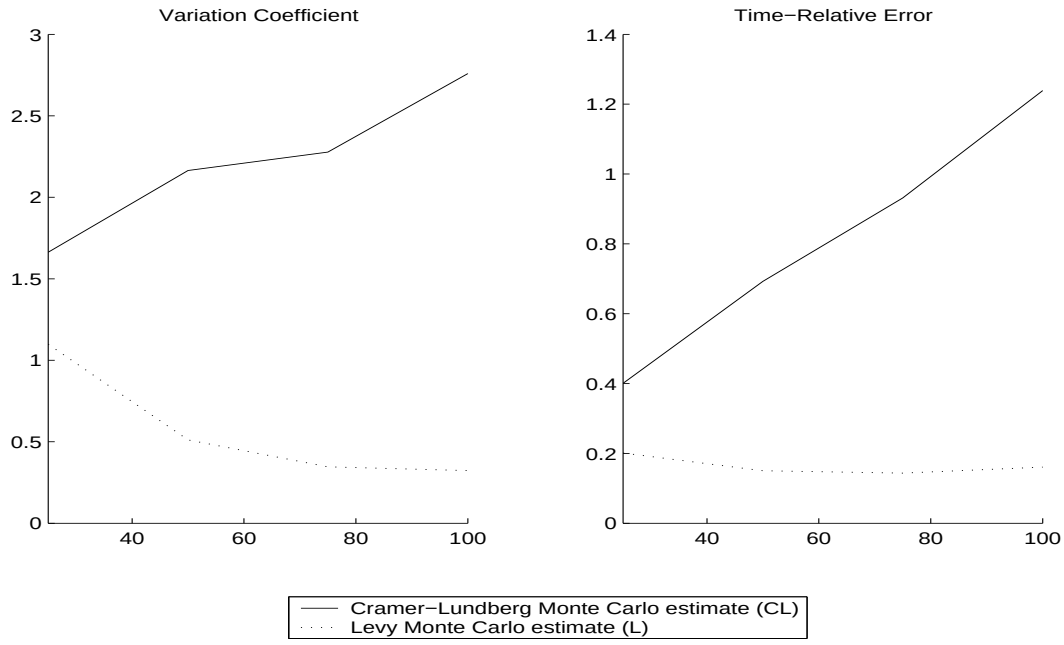


Figure D.10: Approximation for the ruin probability in a short time horizon for a Cramér-Lundberg risk model with moderately heavy Pareto claims.

D.3 Small Tail Probabilities

The following is a numerical example of the Monte Carlo estimators for the small tail probability $\mathbb{P}(S_n < nu)$ developed in the Chapter 5. The results are displayed in a similar fashion as in the previous sections of this Appendix. As for the setting, n was set equal to 20 and the X_i's are i.i.d. lognormal random variables with common distribution $LN(0, 1)$. The estimators displayed are:

1. The Importance Sampling based on the approximated optimal exponential change of measure (**ECM-A**) developed on page 78.

2. The Importance Sampling based on the lognormal distribution which was chosen as an approximation of the optimal exponential change of measure developed on page 79

Example D.11. In this example we choose the intensity of the Poisson process $\lambda = 3/2$, a Pareto distribution with index $\alpha = 3$ and a time horizon $t(u) = u^{3/4}$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
ECM-A	7.259476e-05	1.375445e-04	1.894689e 00	2.5020
LN-A	8.118687e-05	1.626060e-04	2.002861e 00	0.6820

Table D.28: $\mathbb{P}(S_20 < 12)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
ECM-A	8.227712e-04	1.405558e-03	1.708322e 00	2.0010
LN-A	8.233487e-04	1.560080e-03	1.894799e 00	0.6880

Table D.29: $\mathbb{P}(S_20 < 14)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
ECM-A	4.278660e-03	6.769241e-03	1.582093e 00	1.7070
LN-A	4.435283e-03	7.609912e-03	1.715766e 00	0.6800

Table D.30: $\mathbb{P}(S_20 < 16)$.

Method	Estimator	Standard Deviation	Variation Coefficient	Time
ECM-A	1.564343e-02	2.186577e-02	1.397760079696e 00	1.5280
LN-A	1.545749e-02	2.398020e-02	1.551365031917e 00	0.6770

Table D.31: $\mathbb{P}(S_20 < 18)$.

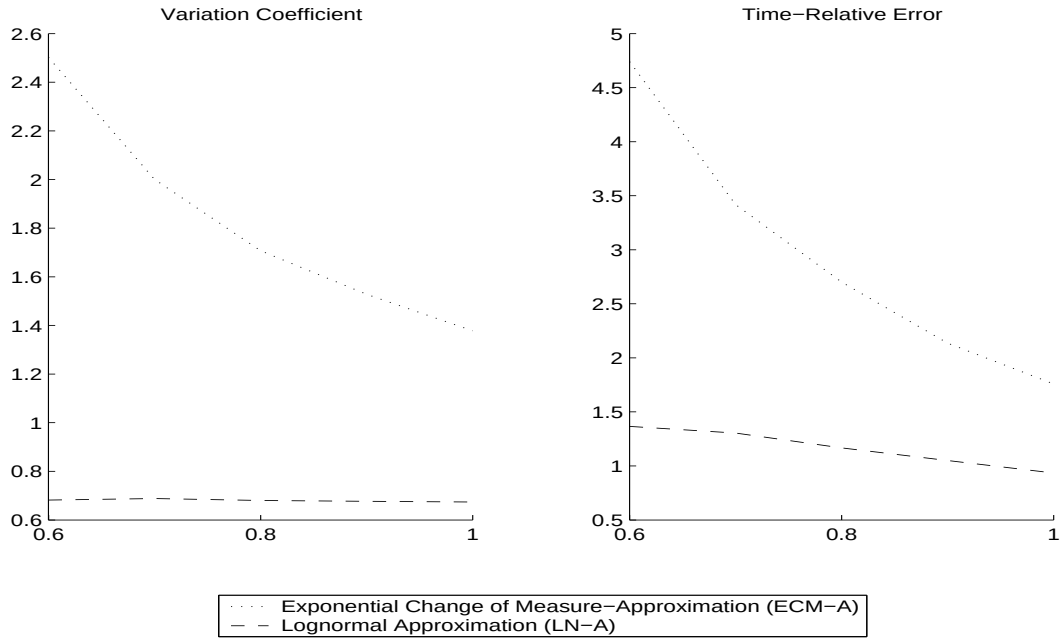


Figure D.11: Approximation for the ruin probability in a short time horizon for a Cramér-Lundberg risk model with moderately heavy Pareto claims.

List of algorithms

Bibliography

Adler, R., J. Blanchet, and J. C. Liu (2008). Fast simulation of Gaussian random fields. In *Proceedings of the 2008 Winter Simulation Conference*, Miami, FL., USA. To appear.

Adler, R. J., R. Feldman, and M. S. Taqqu (Eds.) (1996). *A Practical Guide to Heavy-Tails*. Boston: Birkhäuser.

Albrecher, H., S. Asmussen, and D. Kortschak (2006). Tail asymptotics for the sum of two heavy-tailed dependent risks. *Extremes*, 107–130.

Alink, S., M. Löwe, and M. V. Wüthrich (2004). Diversification of aggregate dependent risks. *Insurance: Mathematics and Economics 35*.

Alink, S., M. Löwe, and M. V. Wüthrich (2005). Analysis of the expected shortfall of aggregate dependent risks. *Astin Bulletin 35*.

Alink, S., M. Löwe, and M. V. Wüthrich (2007). Diversification for general copula dependence. *Statistic Neerlandica 61*.

Anantharam, V. (1989). How large delays build up in a GI/G/1 queue. *Queueing Systems 5*, 345–368.

Andersen, T. G., T. Bollerslev, D. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

Asmussen, S. (2000). *Ruin Probabilities*. Singapore: World Scientific Publishing.

Asmussen, S. (2003). *Applied Probability and Queues*. New York: Springer-Verlag.

Asmussen, S. and K. Binswanger (1997). Simulation of ruin probabilities for subexponential claims. *ASTIN Bulletin 27*, 297–318.

Asmussen, S., K. Binswanger, and B. Højgaard (2000). Rare events simulation for heavy-tailed distributions. *Bernoulli 6*, 303–322.

Asmussen, S., J. Blanchet, S. Juneja, and L. Rojas-Nandayapa (2008). Efficient simulation of tail probabilities of sums of correlated lognormals. *Annals of Operations Research*. Submitted.

Asmussen, S. and P. W. Glynn (2007). *Stochastic Simulation: Algorithms and Analysis*. New York: Springer-Verlag.

Asmussen, S. and C. Kluppelberg (1996). Large deviations results for subexponential tails, with applications to insurance risk. *Stochastic Processes and their Applications 64*, 103–125.

Asmussen, S. and D. P. Kroese (2006). Improved algorithms for rare event simulation with heavy tails. *Advances in Applied Probability 38*, 545–558.

Asmussen, S. and S. S. Petersen (1989). Ruin probabilities expressed in terms of storage processes. *Adv. Appl. Probab. 20*, 913–916.

Asmussen, S. and L. Rojas-Nandayapa (2006). Sums of dependent lognormal random variables: Asymptotics and simulation. In *Thiele Research Reports*, Aarhus, DK. Thiele Centre.

Asmussen, S. and L. Rojas-Nandayapa (2008). Asymptotics of sums of lognormal random variables with Gaussian copula. *Stat. Probab. Lett. 78*. To appear.

Asmussen, S. and R. Rubinstein (1995). Steady state rare events simulation in queueing models and its complexity properties. In J. Dshalalow (Ed.), *Advances in Queueing*. CRC Press.

Barbe, P., L. Fougéres, and C. Genest (2006). On the tail behavior of sums of dependent risks. *Astin Bulletin 36*, 361–373.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008a). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*. To appear.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008b). Subsampled realized kernels. *Journal of Econometrics*. To appear.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society 64*, 253–280.

Bassamboo, A., S. Juneja, and A. Zeevi (2008). Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research 56*, 593–606.

Bee, M. (2008). Importance sampling for sums of lognormal distributions with applications to operational risk. Submitted.

Bertoin, J. (1996). *Lévy Processes*. Cambridge: Cambridge University Press.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1987). *Regular Variation*. Cambridge: Cambridge University Press.

Binswanger, K. S. (1997). *Rare Events and Insurance*. Ph. D. thesis, ETH, Zürich. 12233.

Björk, T. (2004). *Arbitrage Theory in Continuous Time* (Second ed.). Oxford University Press.

Blanchet, J. and P. W. Glynn (2008). Efficient rare-event simulation for the maximum of a heavy-tailed random walk. *Ann. of Appl. Probab. 18*, 1351–1378.

Blanchet, J., P. W. Glynn, and J. Liu (2007). Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue. *QUESTA 57*, 99–113.

Blanchet, J., S. Juneja, and L. Rojas-Nandayapa (2008). Efficient tail estimation for sums of correlated lognormals. In *Proceedings of the 2008 Winter Simulation Conference*, Miami, FL., USA. To appear.

Blanchet, J. and C. Li (2008). Efficient rare-event simulation for heavy-tailed compound sums. *Submitted to ACM TOMACS*.

Boots, N. K. and P. Shahabuddin (2001). Simulating ruin probabilities in insurance risk processes with subexponential claims. In *Proceedings of the 2001 Winter Simulation Conference*, Arlington, VA., USA, pp. 468–476. IEEE.

Bucklew, J. A. (2004). *Introduction to Rare Event Simulation*. New York: Springer.

Bucklew, J. A., J. A. Ney, and P. Sadowsky (1990). Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *J. Appl. Prob. 27*, 44–59.

Chan, J. C. and D. P. Kroese (2008). A skew t-copula model for portfolio credit risk with efficient simulation. *Theory of Probability and its Applications*. Submitted.

Chistyakov, V. P. (1964). A theorem of sums of independent positive random variables and its applications to branching processes. *Theory of Probability and its Applications 9*, 640–648.

Corless, R. M., G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth (1996). On the Lambert W function. *Advances in Computational Mathematics 5*, 329–359.

Cossette, H., M. Denuit, and E. Marceau (2002). Distributional bounds for functions of dependent risks. *Schweiz. Aktuarver. Mitt. 1*, 45–65.

DeBruijin, N. G. (1958). *Asymptotic Methods in Analysis*. Amsterdam: North Holland Publishing Co.

Dembo, A. and O. Zeitouni (1998). *Large Deviations Techniques and Applications* (Second ed.). New York: Springer.

Denuit, M., C. Genest, and E. Marceau (1999). Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics 25*, 85–104.

Dupuis, P., K. Leder, and H. Wang (2006). Importance sampling for sums of random variables with regularly varying tails. *ACM TOMACS 17*, 1–21.

Dupuis, P., A. Sezer, and H. Wang (2007). Dynamic importance sampling for queueing networks. *Ann. Appl. Probab. 17*, 1306–1346.

Dupuis, P. and H. Wang (2004). Importance sampling, large deviations, and differential games. *Stoch. and Stoch. Reports 76*, 481–508.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. New York: Springer-Verlag.

Embrechts, P. and G. Puccetti (2006a). Bounds for functions of dependent risks. *Finance and Stochastics 10*, 341–352.

Embrechts, P. and G. Puccetti (2006b). Bounds for functions of multivariate risks. *Journal of multivariate analysis 97*, 526–547.

Embrechts, P. and N. Veraverbeke (1982). Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics 1*, 55–72.

Epps, T. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association 74*, 291–298.

Fang, K. T., S. Kotz, and K. W. Ng (1987). *Symmetric Multivariate and Related Distributions*. London: Chapman & Hall.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (Second ed.), Volume 2. New York: Wiley.

Fleming, J., C. Kirby, and B. Ostdiek (2001). The economic value of volatility timing. *The Journal of Finance 56*, 329–352.

Fleming, J., C. Kirby, and B. Ostdiek (2003). The economic value of volatility timing using realized volatility. *Journal of Financial Economics 67*, 473–509.

Foss, S., Z. Palmowski, and S. Zachary (2005). The probability of exceeding a high boundary on a random time interval for a heavy-tailed random walk. *Ann. Appl. Probab. 15*, 1936–1945.

Foss, S. and A. Richards (2008). On sums of conditionally independent subexponential random variables. Submitted.

Geluk, J. and K. Ng (2006). Tail behavior of negatively associated heavy tailed sums. *Journal of Applied Probabilty 3*, 525–554.

Geluk, J. and Q. Tang (2008). Asymptotic tail probabilities of sums of dependent subexponential random variables. *Journal of Theoretical Probability*. To appear.

Glasserman, P. (2000). *Monte Carlo Methods in Finance*. New York: Springer.

Glasserman, P., P. Heidelberger, and P. Shahabuddin (2000). Variance reduction techniques for estimating Value-At-Risk. *Management Science 46*, 1349–1364.

Glasserman, P., P. Heidelberger, and P. Shahabuddin (2002). Portfolio Value-At-Risk with heavy tailed risk factors. *Mathematical Finance 12*, 239–269.

Glasserman, P. and J. Li (2005). Importance samplin for portfolio credit risk. *Management Science 51*, 1643–1656.

Goovaerts, M., R. Kaas, Q. Tang, and R. Vernic (2004). The tail probability of discounted sums of Pareto-like losses in insurance. In *Proceedings of the 8th International Workshop on Insurance: Mathematics and Economics*, Rome.

Hansen, P. R. and A. Lunde (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics 3*, 525–554.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure. *Journal of Business and Economic Statistics 24*, 127–161.

Hartinger, J. and D. Kortschak (2006). On the efficiency of the Asmussen-Kroese-estimator and its applications to stop-loss transform. In *Proceedings of the 6th International Workshop on Rare Event Simulation*, Bamberg, Germany. RESIM.

Hayashi, T. and N. Yoshida (2005). On covariance estimation of non-synchronously observed difusion processes. *Bernoulli 11*, 359–379.

Henderson, S. G. and B. L. Nelson (Eds.) (2006). *Simulation*. Handbooks in Operations Research and Management Science. Amsterdam: Elsevier.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions* (2 ed.), Volume 1. New York: Wiley.

Juneja, S. (2007). Estimating tail probabilities of heavy-tailed distributions with asymptotically zero relative error. *QUESTA 57*, 115–127.

Juneja, S. and P. Shahabuddin (2002). Simulating heavy-tailed processes using delayed hazard rate twisting. *ACM TOMACS 12*, 94–118.

Juneja, S. and P. Shahabuddin (2006). Rare event simulation techniques: An introduction and recent advances. In S. G. Henderson and B. L. Nelson (Eds.), *Simulation*, Handbooks in Operations Research and Management Science, pp. 291–350. Amsterdam, The Netherlands: Elsevier.

Klüppelberg, C., A. E. Kyprianou, and R. Maller (2004). Ruin probabilties and overshoots for general Lévy insurance risk process. *Ann. Appl. Probab. 14*, 1766–1801.

Ko, B. and Q. Tang (2008). Sums of dependent nonnegative random variables with subexponential tails. *Journal of Applied Probability 45*, 85–94.

Kortschak, D. and H. Albrecher (2008). Asymptotic results for the sum of dependent non-identically distributed random variables. *Methodology and Computing in Applied Probability*. To appear.

Kyprianou, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Universitext. Berlin: Springer.

L'Ecuyer, P., J. Blanchet, B. Tuffin, and P. W. Glynn (2008). Asymptotic robustness of estimators in rare-event simulation. *ACM TOMACS*. To appear.

McNeil, A., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. New Jersey: Princeton University Press.

Mesfioui, M. and J. F. Quessy (2006). Bounds on the Value-At-Risk for the sum of possibly dependent risks. *Insurance: Mathematics and Economics 9*, 3–20.

Mikosch, T. (2003). *Non-Life Insurance Mathematics*. Berlin: Springer-Verlag.

Mikosch, T. (2006). Copulas: Tales and facts. *Extremes 9*, 3–20.

Mikosch, T. and A. Nagaev (2001). Rates in approximations to ruin probabilities for heavy-tailed distributions. *Extremes 4*, 67–78.

Nelsen, R. B. (2006). *An Introduction to Copulas*. New York: Springer-Verlag.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. New York: Springer-Verlag.

Resnick, S. I. (1992). *Adventures in Stochastic Processes*. Boston: Birkhäuser.

Resnick, S. I. (2006). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer-Verlag.

Robert, P. (2003). *Stochastic Networks and Queues*. Berlin: Springer.

Rojas-Nandayapa, L. (2006). *Tail Probabilities of Sums of Lognormal Random Variables*. Aarhus. Progress Report.

Rojas-Nandayapa, L. and S. Asmussen (2007). Efficient simulation of finite horizon problems in queueing and insurance risk. *QUESTA 57*, 85–97.

Rossberg, A. G. (2008). Laplace transforms of probability distributions and their inversions are easy on logarithmic scales. *Journal of Applied Probability 45*, 531–541.

Rossberg, A. G., T. Amemiya, and K. Itoh (2008). Accurate and fast approximations of moment-generating functions and their inversion for log-normal and similar distributions.

Rubinstein, R. Y. and D. P. Kroese (2004). *The Cross-Entropy Method*. New York: Springer.

Rubinstein, R. Y. and D. P. Kroese (2008). *Simulation and the Monte Carlo Method*. New York: John Wiley.

Sadowsky, J. S. (1996). On Monte Carlo estimation of large deviations probabilities. *Ann. Appl. Prob. 6*, 399–422.

Sadowsky, J. S. and J. A. Bucklew (1990). On large deviations theory and asymptotically efficient monte carlo estimation. *IEEE Transactions on Information Theory 36*, 579–588.

Sato, K. I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge, UK: Cambridge University Press.

Siegmund, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Stat. 3*, 673–684.

Szyszkowicz, S. and H. Yanikomerloglu (2007). On the tail probability of the sum of lognormals. In *IEEE Transactions on Comunications*, Glasgow, Scotland. IEEE.

Szyszkowicz, S. and H. Yanikomerloglu (2008). Limit theorems on the sum of identically distributed equally and positively correlated joint lognormals. *IEEE Transactions on Comunications*. Submitted.

Tang, Q. (2004). Uniform estimates for the tail probability of maxima over finite horizons with subexponential tails. *Prob. Eng. Inf. Sci. 18*, 71–86.

Tang, Q. (2006). Insensitivity to negative dependence of the asymptotic behavior of precise large deviations. *Electric Journal of Probability 11*, 107–120.

Tang, Q. (2008). Insensitivity to negative dependence of the asymptotic tail probabilities of sums and maxima of sums. *Stoch. Anal. Appl.*. To appear.

Tang, Q. and D. Wang (2006). Tail probabilities of randomly weighted sums of random variables with dominated variation. *Stochastic Models 22*, 253–272.

Vanduffel, S., X. Chen, J. Dhaene, M. Goovaerts, L. Henrard, and R. Kaas (2008). Optimal approximations for risk measures of sums of lognormals based on conditional expectations. *Journal of Computational and Applied Mathematics*. To appear.

Voev, V. and A. Lunde (2007). Integrated covariance estimation using high-frequency data in the presence of noise. *Journal of Financial Econometrics 5*, 68–104.

Wang, D. and Q. Tang (2004). Maxima of sums and random sums for negatively associated random variables with heavy tails. *Statistic and Probability Letters 68*, 287–295.

Wüthrich, M. V. (2003). Asymptotic Value-At-Risk estimates for sums of dependent random variables. *Astin bulletin 33*, 75–92.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the Americal Statistical Association 100*, 1394–1411.

Index