A Review of Conditional Rare Event Simulation for Tail Probabilities of Heavy Tailed Random Variables

Leonardo Rojas-Nandayapa
School of Mathematics and Physics
The University of Queensland
4072 Brisbane, Australia
l.rojas@uq.edu.au.

Abstract

Approximating the tail probability of a sum of heavy-tailed random variables is a difficult problem. In this review we exhibit the challenges of approximating such probabilities and concentrate on a rare event simulation methodology capable of delivering the most reliable results: Conditional Monte Carlo. To provide a better flavor of this topic we further specialize on two algorithms which were specifically designed for tackling this problem: the Asmussen-Binswanger estimator and the Asmussen-Kroese estimator. We extend the applicability of these estimators to the non-independent case and prove their efficiencies.
Keywords: Rare-event simulation, conditional Monte Carlo, heavy-tails, efficiency, non-independent.
MSC: 60E05, 90-04.

1 Introduction

The term *rare event* is used to designate all those events whose probabilities are small, yet non-negligible and characterized by the difficulty of its calculation. Often, these rare events are extremely important in applications; for instance, consider the consequences of a natural disaster for an insurance company, or an economic crisis for a financial institution or the sudden arrival of huge number of jobs to a server as it often occurs in a web server. Many of the probability models employed for dealing with these problems contain multiple random variables (not necessarily independent) and the quantities of interest are given in terms of transformations such as sums, products or extremes. In consequence, the explicit calculation of a distribution of interest is often non-trivial and one must rely on approximation methods. Among these, the Monte Carlo method is considered to be one the most reliable, specially in cases where analytical approximations are not available.

In this review we mainly focus on the Monte Carlo method for approximating rare event probabilities, but we also discuss asymptotic approximations; the reason for this is that the implementation of efficient Monte Carlo estimators often requires to draw elements from asymptotic theory. In particular, we specialize on tail probabilities of a sum of random variables

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

When the involved random variables are light tailed, the approximation of such probabilities is dealt via Large Deviations theory. Notwithstanding, certain phenomena are better modeled with heavy-tailed distributions. However, the approximation of rare event probabilities in the

presence of heavy tails is often more involved and it has been considered a challenging problem among the applied probability community. The reason for this is that most classical methods require that the domain of convergence of the moment generating function contains an open set including the origin —a condition which is not satisfied by heavy-tailed random variables as these are characterized by the non-existence of their moment generating functions for positive values of the argument. Therefore, new methods have been called for tackling this problem, and as a result we have seen in the last fifteen years a very intense research activity devoted to Rare Event Simulation. In this review, we focus on the methodology called Conditional Monte Carlo, which has provided some of the most powerful and efficient estimators so far. Some of our contributions employ this technique [1–3]. Here we include some results which have not been previously published in peer-review journals. These are the generalizations of the so called Asmussen-Binswanger and Assmusen-Kroese estimators for the case of independent but non-identical random variables. These extensions are accompanied by their corresponding proofs of efficiency.

This paper is structured as follows. In section 2 we provide a discussion on independent heavy-tailed random variables. This theory is now at an advanced level and well-understood. Several alternative definitions for heavy tails are reviewed and their relations and main properties are studied. In particular, we pay attention to the rich class of subexponential distributions and we discuss how its defining property provides a useful insight into the occurrence of large values of a sum —a characteristic behavior known as the principle of the single big jump. Moreover, it has been recognized that the subexponential property goes beyond the independent case and it is now an area of active research. One of the main contributions of this author is in this front. The main result in [4] states that a sum of lognormals possesses the subexponential property even when the involved random variables are correlated via a Gaussian dependence structure.

A general overview of Monte Carlo methods is provided in Section 3 with a particular emphasis in the area known as *Rare Event Simulation*. The notions of *rare event* and *efficient estimator* are formalized here in order to provide the proper framework for analyzing Monte Carlo estimators for rare event probabilities. We discuss the classical tools such as importance sampling, exponential change of measure and conditional Monte Carlo. We discuss briefly the limitations of some standard methods when applied in a heavy-tailed setting. Section 4 is devoted exclusively to the approximation of tail probabilities of sums of heavy-tailed random variables; a recount of available methods is given there followed by a more detailed exposition on a set of estimators based on the Conditional Monte Carlo; these are known as the Asmussen-Binswanger [5] and the Asmussen-Kroese [6]. In particular, we provide extensions to the non-independent case and prove the efficiency of these estimators. We stress the fact that Theorems 4.1–4.3 are original contributions and their efficiency proofs can be found in the Appendix in Section 6. Finally, Section 5 contains some concluding remarks.

2 Heavy Tails

The term *heavy-tailed phenomena* [7], is often used to refer to real world phenomena where record values are characterized by its extreme behavior. Examples of this type of phenomena are abundant in insurance; for instance, consider the two record costliest (adjusted for inflation) hurricanes striking the United States during the period 1900-2010: Katrina (2005) and Andrew (1992) with damage costs of $105,840 and $45,561 respectively [8]. Both records have extreme values but the most striking feature is that the damage cost of hurricane Katrina more than doubles the damage cost of hurricane Andrew! Further examples occur in Finance and Telecommunications where economic losses or system breakdowns due to large data file sizes or long transmission lengths are of great concern.

The examples above sketch the huge relevance of heavy-tailed phenomena and stress the impor-

tance of having the right probabilistic distributions for modeling their behavior. In the rest of this section we will provide alternative definitions which lead to several classes of heavy-tailed distributions and will study their properties. In particular, we will pay attention to the distinctive behavior of convolutions of certain types of heavy-tailed distributions known as the *principle of the single big jump*. We also establish some contrasts with respect to light tailed distributions, which typically comprehend most of the classical models in probability and statistics. Our exposition follows closely [9] but we also draw elements from [10] and [11].

We say that a random variable X has a (right) *heavy-tailed distribution* if

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

Foss, Korshunov and Zachary [9] employ the term *exponential moment* to refer to the quantity $\mathbb{E}[\mathrm{e}^{\theta X}]$. Adopting this terminology, we say that X has a heavy-tailed distribution if it fails to have a positive exponential moment. In contrast, if a distribution with unbounded right support has a finite positive exponential moment, then we say that it has a *light-tailed distribution*. Further to this, we can easily verify that light-tailed distributions have moments of every order while a random variable with an infinite moment of any order will necessarily have a heavy-tailed distribution. The converse of the last statement is false in general; the classical example is that of a lognormal random variable which has finite moments of every order but it fails to have a positive exponential moment, and in consequence classifies as a heavy-tailed distribution.

In the definition above it is implicit that a right heavy-tailed distribution should have an unbounded right support. In fact, the defining property of a heavy-tailed distribution is inherently related to the rate of decay of its tail probability $\overline{F}(x) := 1 - F(x)$. Therefore, it is natural to obtain equivalent definitions of a heavy-tailed distribution in terms of its tail probability or its hazard rate function. In particular, we define the *hazard function* $\Lambda(x)$ as

$$\Lambda(x) := -\log \overline{F}(x).$$

Moreover, if the tail probability $\overline{F}$ of a distribution is differentiable, then we define the *hazard rate function* $\lambda(x) := \Lambda'(x)$. Hazard (rate) functions arise in a wide variety of applications in survival analysis and reliability where it is known under alternative names such as *survival* or *failure* (rate) functions. In addition, we say that an arbitrary nonnegative function f is *heavy-tailed* iff

$$\limsup_{x\to\infty} \frac{f(x)}{\mathrm{e}^{-\theta x}} = \infty, \qquad \forall \theta > 0.$$

Thus a function is heavy-tailed if it decays slower than an exponential function. The following theorem (cf. [9, Theorem 2.6]) ties together the properties of the tail and the hazard functions of a heavy-tailed distribution and provides alternative definitions.

Theorem 2.1. *Let F be a distribution function with unbounded right support. The following are equivalent:*

1. *F is a heavy-tailed distribution.*
2. *$\overline{F}$ is a heavy-tailed function.*
3. *$\liminf_{x\to\infty} \Lambda(x)x^{-1} = 0$.*

These definitions provide practical means for testing the *heaviness* of any given distribution. Classical examples of heavy-tailed distributions include the subfamily of regularly varying distributions (including Pareto, Loggamma, Burr), Weibull distribution with parameter $0 < \lambda < 1$, Cauchy and Lognormal. On the other hand, the exponential, gamma and normal random variables are examples of light-tailed distributions. In fact, from the definition of a heavy-tailed function it follows that an exponential transformation of a *light-tailed* random variable might

yield to a heavy-tailed random variable. In particular, the Pareto, loggamma and lognormal are the respective heavy-tailed distributions obtained from an exponential transformation of the exponential, gamma and normal distributions.

The class of heavy-tailed distributions as defined above is too general for deriving useful properties. However, by adding some regularity conditions one can obtain tractable subclasses of heavy-tailed distributions which posses attractive properties, yet remain general enough. One of such subclasses is that of long tailed distributions. This subclass is denoted $\mathcal{L}$ and defined by the following property. A distribution $F \in \mathcal{L}$ iff

$$\lim_{x\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(x)} = 1, \qquad \forall y \in \mathbb{R}.$$

An analogue definition exists for a general function f which is ultimately positive and posses the property listed above; in such case we say that f is a long-tailed function. The following result [9, Lemma 2.17] provides the connection between heavy and long tailed distributions.

Lemma 2.2. *If f is a long-tailed function then*

$$\lim_{x\to\infty} \frac{f(x)}{\mathrm{e}^{-\theta x}} = \infty, \qquad \forall \theta \geq 0.$$

In consequence, a long-tailed distribution F is necessarily heavy-tailed, but the converse is not always true. The reference [9] provides a counterexample of a heavy-tailed distribution which fails to be long-tailed. By adding the smoothness condition which defines the class of long-tailed distributions, we gain some useful properties in exchange of some generality. In particular, it is possible to prove that the class of long-tailed functions is closed under linear transformations (mixtures), products, maxima, minima, and convolutions. Moreover, the convolution of a long-tailed distribution with an arbitrary distribution is long-tailed.

A very useful characterization of long-tailed distributions is via their *insensitiveness* with respect to a function h. More precisely, we say that a function f is h-insensitive [9] iff

$$\sup_{|y|\leq h(x)} |f(x+y) - f(x)| = \mathrm{o}(f(x)), \qquad x \to \infty,$$

uniformly in $|y| \leq h(x)$. If the function f is monotone, then h-insensitivity reduces to having $f(x+h(x)) \sim f(x)$ as $x \to \infty$. Clearly, a long-tailed function is insensitive with respect a constant function. However, this property can be strengthened as shown in the following Lemma:

Lemma 2.3. *If F is a long-tailed distribution, then there exists a function $h(x) \to \infty$ such that $\overline{F}$ is an h-insensitive function.*

For instance, if $\overline{F}$ is regularly varying then it is $\mathrm{o}(x)$-insensitive, while the lognormal is $\mathrm{o}(x/logx)$-insensitive and the heavy-tailed Weibull with parameter $\lambda \in (0,1)$ is $\mathrm{o}(x^{1-\lambda})$-insensitive. Next, we discuss briefly the relationship between long-tailed distributions, integrated tails and the mean excess function. If a distribution function is such that $\int_0^\infty \overline{F}(x)\,\mathrm{d}x < \infty$, then we can define the *integrated tail distribution of F* as

$$\overline{F}_I(x) := \min\left\{1, \int_x^\infty \overline{F}(t)\,\mathrm{d}t\right\}.$$

The *mean excess function* can be defined for a distribution having a finite first moment as

$$e(x) := \mathbb{E}[X - x | X > x].$$

That is, the mean excess function is the expected value of the excess of a random variable over a given x, provided that it has exceeded this threshold value. The mean excess function is related to the integrated tail distribution via the relation $e(x) = \overline{F}_I(x)/\overline{F}(x)$. Moreover, the following lemma [9, Lemma 2.25] provides a useful characterization of long-tailed distributions in terms of mean-excess functions.

Lemma 2.4. *The integrated tail distribution F_I is long-tailed iff its associated mean excess function is such that $e(x) \to \infty$.*

In applications, the mean excess function is often used to diagnose the presence of heavy-tails. However, the previous lemma shows that if $\mathrm{e}(u) \to \infty$ we can only verify that the integrated tail distribution is long-tailed but we cannot say anything about the *heaviness* of the original distribution F. In fact, one can construct a counterexample of a light tailed distribution whose mean excess function goes to infinity. This is a case of a more general fact which says that if F is an absolutely continuous distribution, then its density f is a heavy-tailed function but the converse is false in general; that is, if a density function f is heavy-tailed, its distribution function is not necessarily heavy-tailed. In consequence, the mean excess function of a heavy-tailed function should increase to infinity as we let the threshold value $x \to \infty$, but it is not an absolutely reliable tool to diagnose a heavy tail because a distribution with mean excess function going to infinity is not necessarily heavy-tailed. Counterexamples for all these cases can be found in [9].

Convolutions of certain regular and nonnegative heavy-tailed distributions have a unique property which set them apart from light-tailed distributions: the *principle of the single big jump*. This property is extremely useful and most of the heavy-tailed distributions used in practice possess it; in fact, the family of distributions defined by this property forms a proper subclass of long-tailed distributions. To define it we concentrate exclusively on distributions with nonnegative values, but point out that some of the subclasses of heavy-tailed distributions defined below can be generalized to distributions supported over the reals.

We start with the following elementary property of a convolution which holds for all nonnegative distributions F with unbounded right support. Let F^{*n} the n-fold convolution of F and $\overline{F^{*n}}$ its corresponding tail distribution, then

$$\liminf_{x\to\infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} \geq n, \qquad \forall\, n \in \mathbb{N}.$$

The following theorem [9, Theorem 2.12] provides a sufficient condition for the liminf above to be equal to n. In fact, this will provide a very useful insight into the characteristic behavior of the convolution of a heavy-tailed distributions:

Theorem 2.5. *Let F be a nonnegative heavy-tailed distribution. Then*

$$\liminf_{x\to\infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} = n, \qquad \forall\, n \in \mathbb{N}. \tag{1}$$

The corresponding lim inf of most light-tailed nonnegative distributions like the exponential and gamma will be infinite. Hence, it is tempting to use the theorem above as an alternative definition of heavy-tails, but it turns out this is not possible as one can construct a light tailed nonnegative distribution for which the liminf of the ratio of convolutions as defined above is equal to n. A counterexample can be found in [9]. However, we can strengthen the condition above to obtain a subclass of long-tailed distributions: *subexponential distributions*, This subclass was originally introduced by Chistyakov in 1964 [12]. We say that a nonnegative distribution F belongs to the class of subexponential distributions, denoted $\mathcal{S}$, if it possesses the *subexponential property*; that is, the tail probability of the n-fold convolution of F is asymptotically equivalent to n times the tail probability $\overline{F}$. More precisely,

$$\lim_{x\to\infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} = n, \qquad \forall\, n \in \mathbb{N}. \tag{2}$$

Therefore, a heavy-tailed distribution requires and additional regularity condition to be subexponential. That condition is the existence of the limit (2). Moreover, it is possible to prove that

that subexponential distributions form a proper subclass of long-tailed distributions. That is, any subexponential distribution is long-tailed but not every long-tailed distribution will necessarily be subexponential. For a counterexample see [9]. Curiously, the name *subexponential* was originally employed to refer to the class of distributions satisfying $\lim_{x\to\infty} \overline{F}(x)\mathrm{e}^{\lambda x} < \infty$, but nowadays it is employed in the more restrictive sense described above. Subexponentiality is a property of the tail exclusively; however, it has a very interesting implication for the tail of an n-convolution —a characteristic which often goes under the name of the *principle of the single big jump*. Let us start by noting that the distribution of the maximum M_n of n arbitrary i.i.d. random variables (not necessarily subexponential) is given by $F^n(u)$. Hence

$$\mathbb{P}(M_n > x) = 1 - F^n(x) = 1 - (1 - \overline{F}(x))^n = 1 - \sum_{k=0}^{n} (-1)^k \binom{n}{k} \overline{F}^{n-k}(x) = \overline{F}(x)\big(n + \mathrm{O}(\overline{F}(x))\big).$$

In consequence, the tail probability of the maximum is asymptotically equivalent to the tail probability of the convolution, namely $\mathbb{P}(M_n > x) \sim \mathbb{P}(S_n > x)$, where $S_n := X_1 + \cdots + X_n$. Since the $X_i's$ are nonnegative, then $\{M_n > x\} \subset \{S_n > x\}$ and the conclusion above can be written in a conditional form as

$$\lim_{u\to\infty} \frac{\mathbb{P}(M_n > u)}{\mathbb{P}(S_n > u)} = \lim_{u\to\infty} \mathbb{P}(M_n > u | S_n > u) = 1.$$

This expression is very appealing as it says that the if the sum becomes large it is only likely due to the contribution of a single random variable. This behavior is completely opposite to that of lighted tails where the only likely way that a sum of i.i.d. random variables becomes large is as a consequence of several moderately large but otherwise proportionally sized contributions of two or more random variables. 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. Subexponential distributions inherit the properties of long-tailed distributions but also possess many of their own; for instance, the class of subexponential distributions is closed under maxima, minima, mixtures, convolutions and random translations. Also, most of the heavy-tailed distributions used in practice are subexponential such as the Pareto, loggamma, Burr, Weibull and Lognormal. Also, in the case of nonnegative distributions, there exists an upper bound for the expression $\overline{F^{*n}}(x)/\overline{F}(x)$. Such bound goes under the name of Kesten's bound [9, Theorem 3.34], and it is described in the following Theorem:

Theorem 2.6. *Let F be a subexponential distribution. Then, for every $\epsilon > 0$ there exists a constant c such that for all $x \geq 0$ and all $n \geq 1$ it holds that*

$$\frac{\overline{F^{*n}}(x)}{\overline{F}(x)} \leq c(1+\epsilon)^n.$$

Notice that the subexponential property is given for nonnegative, independent and identically distributed random variables, thus it would be desirable to extend this definition to more general sets of random variables and investigate more general conditions under which the principle of the single big jump holds. First we concentrate on distributions supported on the whole real line where it turns out that the defining property (2), which from now on we call *subexponential-type property*, is no longer a tail property. For instance, if we consider a distribution supported all over the reals which fulfills the subexponential-type property, then the distribution $F^+(x) := F(x)\mathbb{I}_{x\geq 0}$ will not necessarily be subexponential and the principle of the single big jump does not hold anymore; this argument motivates the following alternative definition. We say that a distribution F is *whole-line subexponential* if F^+ is subexponential. Alternatively, one could obtain an equivalent definition by consider the distribution $G(x) = \mathbb{P}(X \leq x | X \geq 0)$ instead of F^+. The following theorem [9, Lemma 3.4 and Theorem 3.6] summarize two alternative equivalent definitions which provide useful insights into the class of whole-line subexponential distributions.

Theorem 2.7. *The following assertions are equivalent:*

1. *F is whole-line subexponential.*
2. *F is long-tailed and it possesses the subexponential-type property.*
3. *F is long-tailed and there exists a function $h(x) \to \infty$ for which F is h-insensitive and such that for any two independent random variables $X_1, X_2 \sim F$ it holds that*

$$\mathbb{P}(X_1 + X_2 > x, X_1 > h(x), X_2 > h(x)) = \mathrm{o}(F(x)), \qquad x \to \infty.$$

This theorem says that for a distribution to be whole-line subexponential it is not enough to just have the *subexponential-type property* but we also require a long-tail. Clearly, whole-line subexponential distributions form a proper subclass of long-tailed and heavy-tailed distributions. Moreover, the third part of the theorem above shows that the principle of the single big jump holds as it says that it is unlikely to observe a large value of the sum as a consequence of two (or more) random variables taking moderately large values.

The following Lemma generalizes the principle of the single big jump to independent but non-identically distributed random variables [9, Corollary 3.18].

Lemma 2.8. *Let F be a whole-line subexponential distribution and $F_1, \ldots, F_n$ be a collection of distributions such that $\lim_{x\to\infty} \overline{F}(x)/\overline{F}_i(x) = c_i \geq 0$. Then it holds that*

$$\lim_{x\to\infty} \frac{\overline{F_1 * F_2 * \cdots * F_n}(x)}{\overline{F}(x)} = \sum_{i=1}^{n} c_i.$$

If $\lim_{x\to\infty} F_1(x)/F_2(x) = c > 0$, then we say that F_1 and F_2 are *tail equivalent*. It is straightforward to prove that if F_2 is tail equivalent to a long-tailed distribution, then F_2 is long tailed as well.

The last subclass of heavy-tailed distributions that we will discuss are the subfamilies of distributions with regularly varying tails with index α, denoted $\mathcal{R}(\alpha)$, and defined as the family of nonnegative random variables whose tail probability can be written as $\overline{F}(x) = L(x)x^{-\alpha}$ with $x, \alpha > 0$, and $L(x)$ is a *slowly varying function*. That is, $L(x)$ is a measurable function satisfying

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

In particular, it holds that F is regularly varying distribution iff

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

The class $\mathcal{R}_\alpha$ is often understood as those distributions with a tail behavior similar to a power function with exponent α while 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. 13]. Many authors consider that regularly varying is a synonym of heavy-tails [7]; moreover, this subfamily play fundamental roles in the theory of fluctuations of sums and extremes of independent random variables. The Pareto, Burr, α-stable and loggamma are typical examples of regularly varying distributions. The theory of regularly varying distributions is quite extensive, so we decide to omit most of it but we enunciate a few properties. One of the most remarkable results is Karamata's theorem which is as follows: Let $L \in \mathcal{R}_0$ be bounded in $[x_0, \infty)$ and $\alpha > 1$. Then

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

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 $u \to \infty$. 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 $\mathcal{S}$.

3 Rare Event Simulation

As discussed in the previous section, most of the heavy-tailed distributions used in practice belong to the class of whole-line subexponential distribution. Therefore, the subexponential type-property can be used to approximate the tail probability of a sum of heavy-tailed random variables. Such approximation is very precise in the asymptotic regions of the tail distribution; however, this approximation can loose some precision for moderately large values. Hence, it is desirable to obtain sharper approximations and a natural choice is to recourse to the Monte Carlo method. The elementary version for calculating the tail probability $p_x := \mathbb{P}(X_1 + \cdots + X_n > x)$ is the so called *Crude Monte Carlo* and consists in simulating R identical copies of the random vector $(X_1, \ldots, X_n)$, say $\{(X_{1,r}, \ldots, X_{n,r}),\, r = 1, 2, \ldots\}$; calculating the sums $S_r := X_{1,r} + \cdots + X_{n,r}$; defining the (Bernoulli) random variables $W_{r,x} := \mathbb{I}(S_r > x)$ and returning the arithmetic average

$$\widehat{p}_{x,R} := \frac{1}{R} \sum_{r=1}^{R} W_{r,x}.$$

The law of large numbers implies that for a fixed x, the sequence of random variables $\widehat{p}_{x,R}$ converges to p_x as $R \to \infty$. Moreover, since the random variables $\{W_r : r = 1, 2, \ldots\}$ have bounded variance, the Central Limit Theorem implies that a measure for the random error is the *margin of error* of the Crude Monte Carlo estimator

$$\mathrm{me}(\widehat{p}_{x,R}) := \sqrt{\frac{p_x(1-p_x)}{R}}.$$

This formula exhibits the natural trade-off between precision and computational effort that is required for obtaining an estimate. While in theory we could attain any desirable level of precision by simply increasing the number of replications, in is not so uncommon to end up with very long running times which make unfeasible to attain a certain desired precision. In fact, the crude version of the Monte Carlo is fated to deliver poor approximations when used to estimate *rare event probabilities*. More precisely, we say that an indexed family of events $\{A_x : x \in \mathbb{R}\}$ is a sequence of rare events if $p_x := \mathbb{P}(A_x) \to 0$ as $x \to \infty$. Crude Monte Carlo is considered to deliver poor estimates for rare event probabilities because the asymptotic order of the margin of error is larger than the asymptotic order of the probability of interest as the events becomes rarer. For instance, the *relative error* of the Crude Monte Carlo estimator goes to infinity as the event becomes rarer:

$$\lim_{x \to \infty} \frac{\mathrm{me}(\widehat{p}_{x,R})}{p_x} = \lim_{p_x \to 0} \sqrt{\frac{1-p_x}{p_x R}} = \infty.$$

This implies that the number of replications needed to achieve certain relative precision grows to infinity as the event becomes rarer. This discussion makes obvious the two following facts. 1) The margin of error is not an appropriate precision measure for rare event simulation; instead, we should look at the relative error as defined above (or equivalently to the *coefficient of variation* defined as the square of the margin of error). 2) We need to turn our attention to alternative collections of Monte Carlo estimators requiring a finite number of replications for achieving certain relative precisions no matter how rare the event is.

3.1 Efficiency criteria in rare event simulation

First we discuss the efficiency criteria employed in rare event simulation. In a rare event framework, we say that a Monte Carlo estimator $\widetilde{p}_{x,R}$ is *strongly efficient* or has *bounded relative error* if the (single-replicate) estimator has the following property

$$\limsup_{x\to\infty} \frac{\mathbb{V}\mathrm{ar}\widetilde{p}_{x,1}}{p_x^2} < \infty.$$

This efficiency property says that the number of replications required to estimate p_x with certain fixed relative precision remains bounded as $p_x \to 0$. However, it is often difficult to construct such estimators and/or prove that the limsup above remains bounded. For that reason it is common to employ an alternative weaker criterion denominated *logarithmic efficiency*. This is defined as

$$\limsup_{x\to\infty} \frac{\mathbb{V}\mathrm{ar}\widetilde{p}_{x,1}}{p_x^{2-\epsilon}} = 0, \qquad \forall \epsilon > 0.$$

This criterion implies that the number of replications needed for achieving certain relative precision grows at most at rate of order $|\log(p_x)|$. From a practical point of view, there is no substantial difference between these two criteria, but as mentioned before it is often much easier to prove logarithmic efficiency not only because it is a weaker criterion but also due to the equivalent definition given in the following result.

Lemma 3.1. *An estimator $\widetilde{p}_x$ is logarithmically efficient iff*

$$\liminf_{x\to\infty} \frac{\left|\log \mathbb{V}\mathrm{ar}\,\widetilde{p}_x\right|}{\left|\log p_x\right|} \geq 1.$$

The condition given in the previous Lemma often arises in Large Deviation theory, but are now standard in Rare-Event simulation. The proof of this Lemma is standard but to the best of the author's knowledge, it has seldom appeared in the rare event simulation literature. For sake of completeness, we provide an alternative proof.

Proof of Lemma 3.1. Let us first assume that

$$\liminf_{x\to\infty} \frac{\left|\log \mathbb{V}\mathrm{ar}\,\widetilde{p}_{x,1}\right|}{2\left|\log p_x\right|} \geq 1$$

Then for all $\epsilon > 0$ there exists x_0 such that the inequality on the left hand side on the following display holds for all $x \geq x_0$

$$\frac{\left|\log \mathbb{V}\mathrm{ar}\,\widetilde{p}_{x,1}\right|}{2\left|\log p_x\right|} > 1 - \epsilon/2 \qquad \Longleftrightarrow \qquad \frac{\mathbb{V}\mathrm{ar}\,\widetilde{p}_{x,1}}{p_x^{2-\epsilon}} < 1.$$

The inequality on the right hand side above is obtained by simple algebraic manipulations of the inequality in the left hand side. Taking lim sup we obtain

$$\limsup_{x\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widetilde{p}_{x,1}}{p_x^{2-\epsilon}} < 1.$$

The last inequality holds for all $\epsilon > 0$, hence the the limsup is necessarily smaller or equal to 0. The converse is proved in a similar way. Let us assume

$$\limsup_{x\to\infty} \frac{\mathbb{V}\mathrm{ar}\widetilde{p}_{x,1}}{p_x^{2-\epsilon}} = 0, \qquad \forall \epsilon > 0.$$

Hence, for all $0 < \delta < 1$ there exist x_0 such that for all $x \geq x_0$ the inequality on the left hand side of the following display holds

$$\frac{\mathbb{V}\mathrm{ar}\widetilde{p}_{x,1}}{p_x^{2-\epsilon}} < c \quad \Rightarrow \quad \left|\frac{\log \mathbb{V}\mathrm{ar}\widetilde{p}_{x,1}}{2\left|\log p_x\right|}\right| > \left|\frac{\log c}{\left|\log p_x\right|} - \frac{2-\epsilon}{2}\right|$$

The inequality on the right hand side follows from standard algebraic manipulations. Taking lim inf on both sides we obtain that

$$\liminf_{x\to\infty} \frac{|\log \mathbb{V}\mathrm{ar}\, \widetilde{p}_{x,1}|}{2\,|\log p_x|} > 1-\epsilon/2, \qquad \Longleftrightarrow \qquad \liminf_{x\to\infty} \frac{|\log \mathbb{V}\mathrm{ar}\, \widetilde{p}_{x,1}|}{2\,|\log p_x|} \geq 1.$$

This completes the proof. □

A stronger efficiency concept is achievable and has been proven for several algorithms in the recent literature. This criterion goes under several different names such as *asymptotically zero relative error* or *vanishing relative error*, and it is defined as follows. We say that an estimator $\widehat{p}_{x,1}$ has asymptotically zero relative error iff

$$\limsup_{x\to\infty} \frac{\mathbb{V}\mathrm{ar}\widetilde{p}_{x,1}}{p_x^2} = 0.$$

This criterion is stronger than bounded relative error. In fact, when an estimator has *asymptotically zero relative error*, it can theoretically produce a variance reduction such that the number of replications necessary to attain certain precision is of order $\mathrm{O}(p_x)$ as $x \to \infty$. That means that ultimately the number of replications needed for achieving certain relative precision will continue to decrease as the event becomes rarer until it will be necessary to have a single replication. In addition, there exist, on one hand, efficiency criteria which are 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 definition fills the gap between *logarithmic efficiency* and *Crude Monte Carlo efficiency* and it is often used to describe the improvement over Crude Monte Carlo. On the other hand, there are stronger efficiency concepts which take care of the moments of higher order of an estimator. These go under the name *bounded relative error of order k* [14].

Before we move on, we would like to remark that proving that a candidate estimator satisfies any of the efficiency properties listed above is often a very difficult problem. The reason for this is that the variance of an estimator (appearing on the numerator of the efficiency criteria discussed) is often unknown and one has to find an upper bound of the second moment of the estimator which is tight enough so it remains asymptotically bounded by the appropriate power of the first moment.

3.2 Variance reduction techniques

One of the the most important practical tasks in rare-event simulation is to propose estimators for a given sequence of rare events which may satisfy any of the efficiency properties discussed in the previous section. In a more general framework, the set of techniques employed used to produce estimators which improve the performance of the Crude Monte Carlo estimator go under the name of *variance reduction methods* (c.f. [15]). More precisely, a variance reduction method is an algorithm that modifies an existing estimator (or constructs a new one) in such a way that the resulting estimator remains unbiased and (hopefully) produces a reduction in variance when compared to Crude Monte Carlo. Among the most notorious variance reduction methods one could list *Importance Sampling*, *Control Variates*, *Stratification*, *Conditional Monte Carlo*, *Antithetic sampling*. In addition, variance reduction methods can be divided in *static* and *adaptive*. In a static method, every step of the algorithm is conducted independent of the outcome; in contrast, the evolution of an adaptive algorithm depends on previous outcomes of the algorithm.

While most of these methods are potentially able to produce smaller variances than Crude Monte Carlo, not all of these are well suited for rare event simulation. The main reason is that the demand of variance reduction in the presence of rare events is huge. As discussed in the previous section, one requires a variance which is of much lower order than the one provided by Crude Monte Carlo. Moreover, the quality of these methods is often assessed not only based on the variance reduction itself but also in the amount of computational resources consumed, the theoretical work required and the implementation effort invested.

Among the most powerful methods mentioned above, the most effective ones for rare event simulation are Importance Sampling and Conditional Monte Carlo. We discuss briefly these two methods, but before doing so we point that this review is dedicated to static algorithms; however, it is worth mentioning that adaptive techniques have attracted a considerable amount of attention in recent years due to its effectiveness [cf. 16, 17].

Let us start with Importance Sampling. Assume that all the random variables of interest are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This method relies on the existence of a Radon-Nykodym derivative of the original measure with respect to an alternative probability measure: the *importance sampling distribution*. More precisely, suppose that we are interested in estimating $\mathbb{E}[h(W)]$ where W is a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{E}$ is the expectation operator under the measure $\mathbb{P}$. If $\mathbb{Q}$ is an absolutely continuous measure with respect to $\mathbb{P}$, then it holds that

$$\mathbb{E}[g(W)] = \mathbb{E}^{\mathbb{Q}}[L\, g(W)],$$

where $\mathbb{E}^{\mathbb{Q}}$ is the expectation operator under the measure $\mathbb{Q}$ and $L := \mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$ is the Radon-Nykodym derivative of $\mathbb{P}$ with respect to $\mathbb{Q}$ (the last also goes under the name of *likelihood ratio* in the stochastic simulation literature). In particular, if the measures $\mathbb{P}$ and $\mathbb{Q}$ are absolutely continuous, then the Radon-Nykodym derivative/likelihood ratio is simply the ratio of the corresponding density functions.

The main idea of importance sampling is that if W is simulated according to the measure $\mathbb{Q}$, then the random variable $L\, h(W)$ has an expected value which is equal to $\mathbb{E}[h(W)]$, hence it is an unbiased estimator of the quantity of interest. The variance of the estimator is clearly altered as the second moment is given by

$$\mathbb{E}^{\mathbb{Q}}[L^2 h^2(W)] = \mathbb{E}[Lh(W)].$$

Observe that importance sampling does not always produce variance reduction as the expressions above are not necessarily bounded by the second moment of the estimator under the original measure. In fact, one can end up with an increased or even an infinite variance if one chooses the wrong importance sampling distribution. To the best of the author's knowledge there does not exist a generaliz methodology for choosing an appropriate importance sampling (there is however, a large number of strategies that can suggest good importance sampling distributions); most of the time the selection is based on the experience of the simulator or other additional information about the quantity of interest.

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{I}_A(X)] = \mathbb{E}^{\mathbb{Q}}[L; A] = \mathbb{P}(A)$$

However, when the focus is in estimating probabilities, there exists a distribution with *zero variance*, which is simply the original distribution restricted to the event of interest, that is $\mathbb{Q}(\mathrm{d}x) := \mathbb{I}_A(x)/\mathbb{P}(A)\mathbb{P}(dx)$. Clearly the Radon-Nykodym derivative/likelihood ratio is given by $L := \mathbb{P}(A)\mathbb{I}_A(\cdot)$. Thus we have

$$\mathbb{E}^{\mathbb{Q}}[L; A] = L\mathbb{E}^{\mathbb{Q}}[\mathbb{I}(A)] = \mathbb{P}(A).$$

while for the second moment of the estimator we obtain that

$$\mathbb{E}^{\mathbb{Q}}[L^2; A] = \mathbb{P}^2(A)\mathbb{E}^{\mathbb{Q}}[\mathbb{I}_A] = \mathbb{P}^2(A).$$

From the last expression it follows that this estimator has variance 0. At first sight, this observation might appear of no practical use as the implementation of the zero variance estimator is unfeasible since it requires the knowledge of the unknown probability of interest $\mathbb{P}(A)$. However, the zero variance distribution is of great theoretical interest as one can obtain partial information about it an serve as the ideal model when choosing an appropriate distribution; that is a distribution which is as "close" as possible to the zero variance distribution. Intuitively, we would like to choose a distribution in such way that the "important" event A is sampled with higher frequency with respect to the original distribution. However, there is a natural trade-off in the final value of the variance for the new estimator because if we increase the frequency of any subset it would also increase the values of its likelihood ratios. Therefore, the selection of the importance sampling requires a conscious analysis. In fact, a considerable amount of research effort in rare event simulation has been devoted to approximating the zero variance distribution. One of the most prominent cases is that of the Cross-Entropy method, which consists of an iterative method which selects an "optimal" distribution from a parametric family by minimizing the Kullback-Leibler distance with respect to the zero variance distribution. Another prominent case is that of *Exponential Change of Measure* or *Exponential Twisting* where the importance sampling distribution is selected from the so called exponential family generated by the original distribution. The later technique will be discussed in some detail in the following section.

The second variance reduction technique that will be discussed here is *Conditional Monte Carlo*. This is perhaps the most general variance reduction technique and the one requiring more theoretical effort. The intuitive idea behind it is that the variance of a given estimator can be reduced by *extracting* the variability coming from known information. If we add a little bit more of rigor to this idea we simply end up with conditional expectation. Let us consider again a random variable W defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, h an arbitrary function and $\mathcal{G}$ a simulatable sub-σ-algebra of $\mathcal{F}$. Then

$$\mathbb{E}[g(W)] = \mathbb{E}[\mathbb{E}[g(W)|\mathcal{G}]],$$

and in consequence $\mathbb{E}[g(W)|\mathcal{G}]$ is an unbiased estimator of the quantity of interest $\mathbb{E}[g(W)]$. This estimator is unbiased and, in most practical cases one could easily verify the conditions of the Rao-Blackwell Theorem; in such case, there variance is always smaller or equal than the one of the original estimator. The implementation of this algorithm is more involved when compared to other variance reduction methods as it requires two critical steps, 1) simulating from $\mathcal{G}$ and 2) computing explicitly the random variable $\mathbb{E}[g(W)|\mathcal{G}]$. Obviously it is also desired that the resulting estimator provides a substantial variance reduction, and for achieving that, the sub-σ-algebra $\mathcal{G}$ should contain as much information about the occurrence of the event of interest.

Most of the estimators discussed in this paper rely on Conditional Monte Carlo method. A variety of examples will be provided in the following section to help clarifying the use of this method.

4 Main Results

We will be interested in the tail probabilities of a sum of random variables. More precisely,

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

where $X_1, X_2, \ldots$ is a sequence of random variables and N is possibly random (most algorithms condition on the random number N and then employ a method for simulating the tail probability for a fixed number of random variables). The case of independent and *light-tailed* random variables is well understood via the theory of Large Deviations. Moreover, in terms of Monte Carlo simulation, the standard variance reduction method is Importance Sampling with an *exponential change of measure*. That consists in selecting an importance sampling distribution from a family of probability measures consisting of the normalized measures $F_\theta(\mathrm{d}x) := \mathrm{e}^{\theta x} F(\mathrm{d}x)$ for all possible values of θ in the domain of convergence $\Theta := \{\theta : \mathbb{E}[\mathrm{e}^{\theta X}] < \infty, X \sim F\}$. For estimating

$\mathbb{P}(X_1 + \cdots + X_n > u)$ where the X_i's are nonnegative and independent random variable with common distribution F, the importance sampling distribution and its associated parameter θ are chosen in such way that $\mathbb{E}_\theta[X] = u$. From our discussion of light and heavy tails it follows that such parameter always exists in the light-tailed case for all values of u. This selection is asymptotically optimal as Large Deviations results can be used to prove that it converges to the zero variance importance sampling [18–21]. Moreover, it is known that in the light-tailed case, an exponential change of measure delivers a logarithmically efficient estimator.

However, in the heavy-tailed case the domain of convergence Θ is reduced to the set of non positive values of θ, and therefore we only hope to find a solution of $\mathbb{E}_\theta[X] = u$ when $u \leq \mathbb{E}[X]$; hence, it is clear that it is not possible to implement an optimal exponential change of measure for large values of u. The (nowadays considered) seminal paper [22] presented a number of examples which further exhibited the inherent difficulty in designing good estimators for probabilities of rare events involving heavy-tailed random variables and the challenges of demonstrating their efficiencies. This paper triggered an intense research activity devoted to rare event simulation of heavy-tailed random variables; during the last fifteen years we have seen a wide variety of new developments including estimators for ad hoc applications, novel simulation methodologies targeting rare events and theoretical advances which allowed to simplify some efficiency proofs. As a result, the literature is quite vast. In the following we list a few notorious early works in the area which exemplify the main ideas that one could find in this expanding area of research. A prevalent idea in most of these works is the exploitation of the principle of the single large jump, either by proposing importance sampling distributions which increase the frequency of single big jumps or conditioning in such a way that the conditional probability of a single big jump can be explicitly calculated.

The first logarithmic efficient algorithm was proposed in [5] for the regularly varying case. This is a Conditional Monte Carlo estimator and based on order statistics. The reference [22] proposes a variant of the latter estimator and proves logarithmic efficiency for regularly varying and the lognormal case. A conditional algorithm, similar to that of [5] was proposed in [6]. That algorithm exploits a symmetry relation of random variables which are i.i.d. and the conditioning involves the lower order statistic. It is proved that this estimator has the stronger bounded relative error efficiency property in the regularly case and it is numerically superior to many similar algorithms. It was later proved in [23] that it also achieves bounded relative error in the lognormal case. An independent proof was provided in [1] and reported in [24]; an extension of this is given in Theorem 4.3 in this review. The performance of this algorithm has been improved over time but it continues to be used as a benchmark of performance for similar algorithms. On the importance sampling front, [25] developed a novel methodology where an importance sampling distribution is selected according to a criteria involving the hazard rate function; accordingly, this method is called *hazard rate twisting*. Early examples of adaptive algorithms include [17] which proposed a state-dependent algorithm for the regularly varying case having bounded relative error. Also [16] proposed a state-dependent algorithm in a queueing context and proved that their estimator has vanishing relative error when applied to a *GI/G/1* queue for a large class of heavy-tailed distributions.

Here, we will concentrate in the conditional algorithms proposed in [5] and [6]. At the end of this section we include some extended results. Moreover, several algorithms studied in this dissertation build on these early ideas.

4.1 Conditional Monte Carlo Methods Based on Order Statistics

In this subsection we discuss the algorithms designed by [5] and [6]. As mentioned previously, this algorithms exploit the principle of the single big jump by using order statistics. The idea is neat and simple as one can calculate explicitly the probability that the maximum alone is responsible for the large value of the sum by conditioning on the remaining order statistics. Although the

algorithms in [5] and [6] are both based on this idea, they differ in the way of conditioning. The original idea appeared in [5] but the modified version in [6] provided a more efficient and easier to implement algorithm. We complement these ideas by providing extensions with the corresponding proofs of efficiency. Let us first start with the Asmussen-Binswanger estimator:

Asmussen-Binswanger estimator We assume that $\{X_1, \ldots, X_n\}$ is a collection of i.i.d. heavy-tailed random variables. The idea is to simulate the first $n-1$ order statistics out of n. The procedure is simple as we just simulate $X_1, \ldots, X_n$ and discard the largest one. The clever idea here is that we can now calculate explicitly the conditional probability of the rare event $\{S_n \geq u\}$ given the order statistics. This comes out as the following random variable

$$\mathbb{P}\big(S_n > x \,\big|\, X_{(1)}, \ldots, X_{(n-1)}\big) \;=\; \frac{\overline{F}((x - S_{(n-1)}) \vee X_{(n-1)})}{\overline{F}(X_{(n-1)})},$$

where $S_{(n-1)} = X_{(1)} + \cdots + X_{(n-1)}$. This algorithm is logarithmic efficient in the regularly varying [26] and the lognormal case [5]. However, we can easily drop the *identically distributed* assumption. When simulating the order statistics, we just need to keep track of the (random) index of the largest random variable, say K. The conditioning will deliver instead

$$\mathbb{P}\big(S_n > x \,\big|\, X_{(1)}, \ldots, X_{(n-1)}\big) \;=\; \frac{\overline{F}_K\big(X_{(n-1)} \vee (x - S_{n-1})\big)}{\overline{F}_K\big(X_{(n-1)}\big)}, \tag{3}$$

where $F_k(\cdot)$ is the distribution function of the k-th random variable. Clearly, the random variable above is unbiased. Moreover, we prove that it has logarithmic efficiency when all random variables are independent lognormals but not necessarily identically distributed.

Theorem 4.1. *Let $X_1, .., X_n$ be independent lognormal random variables. Then the estimator* (3) *is logarithmic efficient.*

The proof of this theorem is slightly technical and relegated to the appendix. This result can be further extended to the case where the random variable with the heaviest tails is lognormal. The proof of this follows trivially by comparison of the tail asymptotics.

Asmussen-Kroese estimator. A slight tweak in the Asmussen-Binswanger estimator can result in a dramatic variance reduction. The main observation of [6] is that the algorithm above still has a large variability due to the fact that there is a significant large probability of having a big jump among the first $n-1$ order statistics. This probability is dramatically reduced by considering a symmetry argument. The idea is to calculate the probability of the event $\{S_n > x, X_k = M_n\}$ for $k = 1, \ldots, n$ and where $M_n = \max\{X_i : i = 1, \ldots, n\}$. By symmetry we obtain

$$\mathbb{P}(S_n > x) = n\,\mathbb{P}(S_n > x, X_d = M_n). \tag{4}$$

Conditioning on $\mathcal{F} = \sigma(X_1, \ldots, X_{d-1})$ and noting that

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

This algorithm has bounded relative error in the regularly varying case [6] and the lognormal [1, 23]. Moreover, the *identically distributed* assumption can be dropped by substituting the symmetric argument with

$$\mathbb{P}(S_n > x) = \sum_{k=1}^{n} \mathbb{P}(S_n > x, X_k = M_n).$$

This idea was empirically explored in our technical report [1]. An obvious approach consists in estimating individually each of the terms in the summation above. The resulting estimator

has good efficiency properties but requires more computational effort. The following alternative approach delivers much better results. The strategy described can be seen as an hybrid between conditional Monte Carlo and importance sampling where the importance sampling distribution is a mixture. Let $p_k := \mathbb{P}(X_k = M_d)$, the probability that X_k takes the largest value among the X_i's and q_k a discrete probability measure supported over. Hence $\{1, 2, \ldots, n\}$,

$$\begin{aligned}\mathbb{P}(S_n > x) = \sum_{k=1}^{n} \mathbb{P}(S_n > x | X_k = M_n) p_k \frac{q_k}{q_k} &= \sum_{k=1}^{n} \frac{\mathbb{P}(S_n > x, X_k = M_n)}{q_k} q_k \\ &= \mathbb{E}\left[\frac{\mathbb{I}(S_n > x, X_K = M_n)}{q_K}\right].\end{aligned}$$

where K is distributed according to q_k. Further if we condition with respect to the sub-σ-algebra $\mathcal{F} = \sigma(K, X_1, \ldots, X_{K-1}, X_{K+1}, \ldots, X_n)$ we obtain

$$\mathbb{P}(S_n > x) = \mathbb{E}\left[\frac{\overline{F}_K(M_{-K} \vee (x - S_{-K}))}{q_K} \mid \mathcal{F}\right], \tag{5}$$

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 q_k's should deliver a significant variance reduction. Intuitively, this should be minimized if we choose $q_k^* := P(X_k = M_n | S_n > u)$. That is, the probability that the k-th random variable is largest conditioned to the rare event. However, this probability is not available beyond the independent case. Our suggestion is to use

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

Empirically, we have verified that this proposal approaches the value of the p_k's as $u \to \infty$. Hence, it is conjectured that asymptotically these are equivalent. Moreover, this estimator

$$\frac{\overline{F}_K(M_{-K} \vee (x - S_{-K}))}{p_K} \tag{6}$$

delivers excellent numerical results with lower computational effort. Efficiency proofs for non-identical and independent random variables in the lognormal and regularly varying cases are given in the next two Theorems and their proofs can be found in the appendix.

Theorem 4.2. *Let $X_1, X_2, \ldots, X_n$ be independent lognormal random variables, K a discrete random variable supported over $\{1, \ldots, n\}$. Then (6) is an unbiased estimator of $\mathbb{P}(S_n > x)$ with bounded relative error.*

Theorem 4.3. *Let $X_1, X_2, \ldots, X_n$ be independent regularly varying random variables with indexes α_i respectively, K a discrete random variable supported over $\{1, \ldots, n\}$. Then (6) is an unbiased estimator of $\mathbb{P}(S_n > x)$ with bounded relative error.*

5 Conclusions

Calculating the tail probability of a sum of random variables is a fundamental problem in applied probability. In particular, having sharp approximations of these tail probabilities is of key importance in several disciplines. While in the most common cases this problem is tackled with standard methods, it turns out that it is very challenging to deal with random variables which posses heavy tails. This was the perfect excuse to provide a review on the standard theory of heavy-tails and subexponentiality; we studied the alternative definitions for heavy-tails and provided a glimpse of their main properties. In particular, we paid attention to the subexponential-type property which is characterized for the principle of the single big jump. One of my main contributions to this area was to demonstrate that this behavior goes beyond the

independent case by proving that a collection of correlated lognormals posses the subexponential-type property. This result provides an asymptotic equivalent expression for the tail probability of a sum of correlated lognormals which can be used as an approximation of the real probability.

However, in most applications it is desired to have a better precision. Some of the most precise and reliable methods to obtain approximations is via the Monte Carlo method. In particular, the subarea known as Rare-event simulation is devoted to develop the methodologies for delivering sharp approximations. The second part of this review is dedicated to these aspects. We discussed the main techniques and introduced the efficiency concepts used to assess the theoretical performance of estimators for rare event probabilities. This area of research is quite extensive. However, for this review I decided to focus on Conditional Monte Carlo for independent but not necessarily identically distributed random variables. The original estimators are now considered standard for the i.i.d. case but nevertheless we were able to extend these results by dropping the identically distributed assumption. The efficiency proofs in Theorems 4.1, 4.2 and 4.3 appear in my Phd thesis but otherwise this is unpublished material.

6 Appendix: Proofs

Proof of Theorem 4.1. In order to characterize the dominant tail behavior we define

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

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

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

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

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

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

$$\mathbb{E}\Big[\frac{\overline{F}_k^2(u/d)}{\overline{F}_k^2(X_{(d-1)})}; X_{(d-1)} < \frac{u}{d}\Big] + \mathbb{P}\big(X_{(d-1)} > u/d\big). \tag{7}$$

We concentrate on the expectation in the last term. Since $X_{(n-1)} < u/d$ we can apply Lemma 6.1 to get a bound for the quotient in the first expectation to obtain

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

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

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

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

$$\frac{\overline{F}_{(d-1)}(t)}{\overline{F}^2(t)} \le \frac{\sum_{i\neq j}\overline{F}_i(t)\overline{F}_j(t)}{\overline{F}^2(t)} = \mathrm{O}(1), \qquad (0,\infty). \tag{8}$$

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

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

where the last inequality was obtained by using the argument (8). So, to prove logarithmic efficiency we need

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

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 6.1. *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 $y \le x$ it holds that*

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

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

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

For proving this, we will start from the inequality

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

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 4.2. Recall that the condition for *asymptotic bounded relative error* is equivalent to

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

By subexponentiality we have that $\mathbb{P}(X_k > u) = \mathrm{O}(\mathbb{P}(S_n > u))$ 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

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

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

$$\limsup_{u\to\infty} \int_0^{u/2} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_i(y/d)dy + \int_{u/2d}^{\infty} \frac{\overline{F}_k^2(y)}{\overline{F}_k^2(u)} f_i(y)dy < \infty, \qquad i,k = 1,\ldots,d. \tag{9}$$

We prove separately that each of this two integrals remain bounded as $u \to \infty$. The first integral remains bounded due to Lemma 6.2. 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 6.2. *Under the hypothesis of the Theorem 4.2 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 $\overline{F}_k(u) = \exp\big\{-\int_0^u \lambda(t)dt\big\}$, 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

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

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 $\big\{y : \log(2y) > \frac{y\log u}{u-y}\big\} = (g(u), u/2)$ 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 rancom 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 4.3. Note that in the proof of Theorem 4.2 we did not make use of the hypothesis about the distribution up to (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

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

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

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

Putting together these two expressions we complete the proof. □

Acknowledgments

The author would like to thank the editors and the referees ???. The author is supported by the ARC grant DE130100819.

References

[1] S. Asmussen and L. Rojas-Nandayapa, "Sums of dependent lognormal random variables: Asymptotics and simulation," in *Thiele Research Reports*, (Aarhus, DK), Thiele Centre, 2006.

[2] L. Rojas-Nandayapa and S. Asmussen, "Efficient simulation of finite horizon problems in queueing and insurance risk," *Queueing Systems*, vol. 57, pp. 85–97, 2007.

[3] J. H. Blanchet and L. Rojas-Nandayapa, "Efficient simulation of tail probabilities of sums of dependent random variables," *Journal of Applied Probability*, vol. 48A, pp. 147–164, 2011.

[4] S. Asmussen and L. Rojas-Nandayapa, "Asymptotics of sums of lognormal random variables with Gaussian copula," *Stat. Probab. Lett.*, vol. 78, 2008. To appear.

[5] S. Asmussen and K. Binswanger, "Simulation of ruin probabilities for subexponential claims," *ASTIN Bulletin*, vol. 27, pp. 297–318, 1997.

[6] S. Asmussen and D. P. Kroese, "Improved algorithms for rare event simulation with heavy tails," *Advances in Applied Probability*, vol. 38, pp. 545–558, 2006.

[7] S. I. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer-Verlag, 2006.

[8] E. S. Blake, C. W. Landsea, and E. J. Gibney, "The deadliest, costliest, and most intense united states tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts)," tech. rep., National Oceanic and Atmospheric Administration, 2011.

[9] S. Foss, D. Korshunov, and S. Zachary, *An introduction to heavy-tailed and subexponential distributions*. Springer, 2011.

[10] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*. New York: Springer-Verlag, 1997.

[11] T. Mikosch, *Non-Life Insurance Mathematics*. Berlin: Springer-Verlag, 2009.

[12] V. P. Chistyakov, "A theorem of sums of independent positive random variables and its applications to branching processes," *Theory of Probability and its Applications*, vol. 9, pp. 640–648, 1964.

[13] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular Variation*. Cambridge: Cambridge University Press, 1987.

[14] P. L'Ecuyer, J. Blanchet, B. Tuffin, and P. W. Glynn, "Asymptotic robustness of estimators in rare-event simulation," *ACM TOMACS*, 2008. To appear.

[15] S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*. New York: Springer-Verlag, 2007.

[16] J. Blanchet and P. W. Glynn, "Efficient rare-event simulation for the maximum of a heavy-tailed random walk," *Ann. of Appl. Probab.*, vol. 18, pp. 1351–1378, 2008.

[17] P. Dupuis, K. Leder, and H. Wang, "Importance sampling for sums of random variables with regularly varying tails," *ACM TOMACS*, vol. 17, pp. 1–21, 2006.

[18] D. Siegmund, "Importance sampling in the Monte Carlo study of sequential tests," *Ann. Stat.*, vol. 3, pp. 673–684, 1976.

[19] J. A. Bucklew, J. A. Ney, and P. Sadowsky, "Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains," *J. Appl. Prob.*, vol. 27, pp. 44–59, 1990.

[20] J. S. Sadowsky and J. A. Bucklew, "On large deviations theory and asymptotically efficient monte carlo estimation," *IEEE Transactions on Information Theory*, vol. 36, pp. 579–588, 1990.

[21] J. S. Sadowsky, "On Monte Carlo estimation of large deviations probabilities," *Ann. Appl. Prob.*, vol. 6, pp. 399–422, 1996.

[22] S. Asmussen, K. Binswanger, and B. Højgaard, "Rare events simulation for heavy-tailed distributions," *Bernoulli*, vol. 6, pp. 303–322, 2000.

[23] J. Hartinger and D. Kortschak, "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, 2006.

[24] L. Rojas-Nandayapa, *Risk probabilities: Asymptotics and Simulation.* PhD thesis, Aarhus University, 2008.

[25] S. Juneja and P. Shahabuddin, "Simulating heavy-tailed processes using delayed hazard rate twisting," *ACM TOMACS*, vol. 12, pp. 94–118, 2002.

[26] K. S. Binswanger, *Rare Events and Insurance.* PhD thesis, ETH, Zürich, 1997. 12233.