*Applied Probability Trust* (8 February 2011)

# EFFICIENT SIMULATION OF TAIL PROBABILITIES OF SUMS OF DEPENDENT RANDOM VARIABLES

**Abstract**

We study asymptotically optimal simulation algorithms for approximating the tail probability of $\mathbb{P}(e^{X_1} + \cdots + e^{X_d} > u)$ as $u \to \infty$. The first algorithm proposed is based on Conditional Monte Carlo and assumes that $(X_1, \ldots, X_d)$ has an elliptical distribution with very mild assumptions on the radial component. This algorithm is applicable to a large class of models in finance as we demonstrate with examples. In addition, we propose an importance sampling algorithm for an arbitrary dependence structure that is shown to be asymptotically optimal under mild assumptions on the marginal distributions and basically, that one can simulate efficiently $(X_1, \ldots, X_d | X_j > b)$ for large $b$. Extensions that allow us to handle portfolios of financial options are also discussed.

JOSÉ H. BLANCHET,* *Columbia University*

LEONARDO ROJAS-NANDAYAPA,** *The University of Queensland*

*Keywords:* Rare-event simulation; Efficiency; Dependence; Heavy-tailed distributions; Log-elliptical distributions; Tail probabilities; Variance reduction; Importance Sampling; Conditional Monte Carlo.

2000 Mathematics Subject Classification: Primary 60G50,60W40
Secondary 60G70

## 1. Introduction

Efficient rare-event simulation for sums of heavy-tailed random variables is a challenging problem of significant relevance in several disciplines such as queueing theory, insurance and finance. This research area was fundamentally shaped by the contributions of Søren Asmussen and his collaborators; the first class of provably efficient algorithms in this type of settings was proposed in [6]. The difficulties inherent to rare-event simulation with heavy tails were further fleshed out in [7]. Since then, this research area has attracted a considerable amount of interest and has rapidly grown into a major subject in rare-event simulation.

In order to discuss our results and put them in perspective relative to the existing literature, we consider the following mathematical formulation of the problem. Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a $d$-dimensional multivariate random vector and define $\alpha(u) := \mathbb{P}(e^{X_1} + e^{X_2} + \cdots + e^{X_d} > u)$. An important number of multivariate models in finance and insurance applications possessing heavy-tailed marginal distributions arise from the exponential transformations of standard light-tailed multivariate distributions [33].

* Postal address: Department of Industrial Engineering and Operations Research, Columbia University, S. W. Mudd Building, 500 West 120th Street, New York, NY, USA. Email address jose.blanchet@columbia.edu

** Postal address: Department of Mathematics, The University of Queensland, Brisbane, Queensland 4072, Australia, Email address: l.rojasnandayapa@uq.edu.au

In this paper we consider the problem of developing *asymptotical optimal* Monte Carlo estimators of $\alpha(u)$, but also additional functions beyond the sum (see Example 2). Recall that a collection of estimators $(Z_u : u \geq 0)$ is said to be asymptotically optimal if $Z_u$ is an unbiased estimator for $\alpha(u)$ and if $\sup_{u>0} E(Z_u^2)/\alpha(u)^{2-\varepsilon} < \infty$ for all $\varepsilon > 0$. Jensen's inequality implies that the second moment of an asymptotically optimal estimator achieves the best possible rate of decay to zero in logarithmic scale. In other words, asymptotic optimality amounts to showing that

$$\frac{\log E(Z_u^2)}{2\log(\alpha(u))} \to 1, \qquad u \to \infty. \tag{1}$$

Most of the recent literature on provably efficient rare-event simulation for $\alpha(u)$ concentrates on the sum of independent and identically distributed heavy-tailed random variables. This setting is motivated by classical queueing models (tail of the delay in an $M/G/1$ queue) and insurance applications (ruin probabilities in the classical risk model); c.f. Asmussen [4, 5].

The first provable efficient estimator for $\alpha(u)$, based on Conditional Monte Carlo (CMC), was given in [6] for regularly varying increment distributions (power-law heavy-tails). A more recent paper Asmussen and Kroese [10] proposed refined CMC algorithms that are also applicable to Weibull-type and lognormal tails. Other provably efficient estimators based on hazard-rate tilting ideas include those of Juneja and Shahabuddin [28] and Boots and Shahabuddin [16]. Dupuis, Leder and Wang [21] proposed a mixture-based importance sampling for regularly varying increment distributions and proved that their sampler is strongly optimal (i.e. the optimality criterion above holds with $\varepsilon = 0$). The paper of Blanchet and Li [15] provides an estimator that can be shown to be strongly optimal assuming only that the increment distributions are subexponential. Asymptotically vanishing relative error (i.e. $\limsup_{u\to\infty} E(Z_u^2)/\alpha(u)^2 = 1$) has been established in a few instances, mostly in the setting of independent heavy-tailed increments, see for instance [13, 24, 27].

In the case where the $X_i$'s exhibit dependence has been substantially less studied. Asmussen, Blanchet, Juneja and Rojas-Nandayapa [8] consider the case in which $\mathbf{X}$ is a multidimensional Gaussian vector. This setting is motivated by considering $d$ correlated asset prices, each following a Black-Scholes dynamic in which individual stock price are lognormal. In that paper, several other Monte Carlo estimators based on importance sampling are proposed and shown to be asymptotically optimal; one of those estimators is actually shown to have asymptotically vanishing relative error as $u \nearrow \infty$. A related paper that discovered independently one of the strategies suggested in Blanchet, Juneja and Rojas-Nandayapa [14] is Klöppel, Reda and Schachermayer [29] where an exponential tilting of the radial component of $\mathbf{X}-\mathbb{E}\mathbf{X}$ in polar coordinates is proposed. Related conditional Monte Carlo strategies have been studied in [14, 17, 18].

On the side of asymptotic approximations, many authors have obtained results for the sum of dependent heavy-tailed random variables (for a recent account see [3, 11, 23, 30, 34] and references therein). A standard approach consists in taking advantage of conditional independence structure to reduce the problem to the (well understood) case of independent components. We point out that even when asymptotic approximations are available, efficient Monte Carlo methods provide a good complement because the error present in any type of approximation can be reduced at the price of increasing the number of replications. Asymptotic optimality then provides reassurance that such number of replications will scale graciously as the event of interest becomes more rare.

In this paper we develop a methodology applicable to a wide class of models beyond the Gaussian case treated in [14] and [29]. Our contributions are as follows:

A. Let $\mathbf{X}$ follow an elliptical distribution with radial component $R$. Assume that the density $f_R(\cdot)$ of $R$ is eventually positive and satisfies

$$\lim_{x\to\infty} \frac{x f_R(x)}{\mathbb{P}(R > x)^{1-\varepsilon}} = 0, \qquad \forall \varepsilon > 0. \tag{2}$$

We propose an efficient CMC estimator for $\alpha(u)$; see Theorem 1.

B. Assume that for every $i = 1, \ldots, d$ and all $c > 0$ it holds that

$$\frac{\log \mathbb{P}(X_i > b - c)}{\log \mathbb{P}(X_i > b)} \to 1, \qquad b \to \infty. \tag{3}$$

Assume also that an asymptotically optimal importance sampling estimator is available for $\mathbb{E}[\sum_{i=1}^d I(X_i > u)]$ as $u \nearrow \infty$. Based on such estimator we construct an importance sampling estimator that is asymptotically optimal for $\alpha(u)$; see Theorem 2.

C. These results can be applied to several situations of interest. To illustrate our contributions we apply the results in A to the Variance-Gamma process and to portfolios containing call options. For the results in B we consider the Kou model [31] for modeling asset prices. All these applications are discussed in Section 3.

Typically, both estimators in A and B are easy to implement as we shall discuss in Sections 3 and 4. The first estimator requires the implementation of a numerical algorithm for finding the roots of a function depending on the spherical component. However, the underlying function has nice regularity properties that make the root finding procedure fast and reliable; implementation details are given in Section 4. The second estimator is applicable whenever one can compute or estimate efficiently $P(X_i > b)$ for each $i = 1, ..., d$ as well as to be able to sample $(X_1, ..., X_d | X_i > b)$ efficiently. These requirements, which involve only marginal computations and marginal conditioning, can often be satisfied using exponential tilting as we shall illustrate in Section 3. Moreover, in some important cases one can obtain asymptotically vanishing relative error estimators by using the ideas underlying the second estimator. Such is the case of jointly Gaussian $X_i$'s which was studied in [14].

The result described in A allows us to deal with virtually any type of tail behavior for the marginal distributions (within the elliptical framework) as long as the radial component satisfies the mild assumption (2). The price to pay, of course, is a restrictive dependence structure. In contrast, the result in B is helpful to deal with more general dependence structures. The required condition (3) is satisfied if the tails of $e^{X_i}$ are suitably heavy-tailed: lognormal-type tails, Pareto or power-law tails. Moreover, since the estimator in B is based on importance sampling it can also be used to easily estimate conditional expectations of $\mathbf{X}$ given the event $\{e^{X_1} + e^{X_2} + \cdots + e^{X_d} > u\}$; see the related discussion in [2]. Conditional expectations (such as the conditional overshoot over level $u$) are of importance, for instance, in quantitative risk management; see for instance [32]. Evaluating such conditional expectations is more complicated when one uses a CMC estimator as in A. The reason is that when applying CMC one needs to analytically evaluate the expectation of interest given the conditioning. We can do so in our case because such conditional expectation involves finding at most two roots as we shall see in Section 2.

Thus one may think that estimators in B are preferable given their level of generality in terms of dependence. However, an advantage of the estimator in A is that it is guaranteed to give variance reduction for all values of $u$ whereas the asymptotic optimality proved for the class of estimators in B only guarantees optimal performance for relatively large values of $u$. It is often the case that one can apply both estimators to the same problem instance; in such situation, we recommend using both for cross validation.

The rest of the paper is organized as follows. In Section 2 we provide the statements of our main results along with their proofs. In Section 3 we discuss examples. Finally, 4 contains numerical experiments and additional discussion on several implementation issues.

## 2. Assumptions and Asymptotic Optimality Results

This section is divided into three parts. Subsection 2.1 is devoted to our CMC estimator applicable to elliptical distributions while Subsection 2.2 concerns our importance sampling estimator. Finally, Subsection 2.3 contains the proof of efficiency of our CMC estimator, which is somewhat technical.

### 2.1. Conditional Monte Carlo for the Sum of Log-elliptical Distributions

**Definition 1.** We say that a vector $\mathbf{X} = (X_1, \ldots, X_d)^T$ follows an elliptical distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^d$, non-negative definite dispersion matrix $\mathbf{\Sigma}$, and radial (cumulative) distribution $F_R(\cdot)$ supported on $[0, \infty)$, if $\mathbf{X}$ admits the stochastic representation

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + R\mathbf{A}\Theta,$$

where $\mathbf{A}\mathbf{A}^T = \mathbf{\Sigma}$, $R$ is random variable with distribution $F_R(\cdot)$, $\Theta$ is a random vector with a uniform distribution on the unit sphere $\mathcal{S}_d$ in $\mathbb{R}^d$ and independent of $R$. In such case we write $\mathbf{X} \sim \mathcal{E}(\boldsymbol{\mu}, \mathbf{\Sigma}, F_R)$.

Additionally, we will say that the random vector $(\mathrm{e}_1^X, \ldots, \mathrm{e}^{X_d})$ follows a log-elliptical distribution with parameters $(\boldsymbol{\mu}, \mathbf{\Sigma}, F_R)$.

A number of important models in the financial literature can be cast in the framework of elliptical distributions [32]. Our first estimator allows us to estimate the tail probability of a sum of log-elliptical random variables $\mathrm{e}^{X_1} + \cdots + \mathrm{e}^{X_d}$ under very mild assumptions. In particular, we shall only impose assumption (2), which is verified by using L'Hôpital's rule in virtually any model with a continuous distribution for $R$ supported in the whole real line. Interesting examples are discussed in Section 3. Assumption (2), however, rules out distributions with compact support.

Motivated by applications in the financial literature, one might also be interested in estimate tail probabilities of more general functions of the vector $(X_1, ..., X_d)$. For instance, one might be interested in the tail of a portfolio of call options when the underlying assets follow log-elliptical distributions (see Example 2 below). In order to formulate a result that is general enough to handle these types of applications we shall represent our random variable of interest as $G(R, \Theta)$. So, for example, in the case of the sum of log-elliptical distributions we have that

$$G(R,\Theta) = \sum_{i=1}^{d} \exp\left(\mu_i + R\left\langle \mathbf{A}_i, \Theta\right\rangle\right) = \sum_{i=1}^{d} \exp(X_i), \tag{4}$$

where $\mathbf{A}_i$ is the $i$-th row of the matrix $\mathbf{A}$. More generally, we shall consider any function $G(r,\theta)$ that for large values of $r$ and all $\theta \in \mathcal{S}_d$ behaves like the function in (4). Thus, we shall impose the following assumptions.

**Assumptions on** $G(\cdot)$

1. Suppose that $(G(r,\theta) : r \geq 0, \theta \in \mathcal{S}_d)$ is a positive function, continuous in both variables and differentiable in $r$.
2. For any $\delta > 0$ and $s \in \mathcal{S}_d$ define $\mathcal{D}(\delta, s) = \{\theta \in \mathcal{S}_d : \|\theta - s\| < \delta\}$ and assume that there exists $\delta_0 > 0$, $s_* \in \mathcal{S}_d$, $r_0 > 0$ and $v > 0$ such that for all $0 < \delta \leq \delta_0$ and all $r > r_0$
$$\sup_{\theta \in \mathcal{S}_d} G(r,\theta)^{1-v\delta} \leq \inf_{\theta \in \mathcal{D}(\delta, s^*)} G(r,\theta). \tag{5}$$
3. Suppose also that
$$\sup_{r>r_0, \theta \in \mathcal{D}(\delta_0, s^*)} G(r,\theta) = \sup_{r>r_0, \theta \in \mathcal{S}_d} G(r,\theta). \tag{6}$$
4. Finally, suppose that $\delta_1 \in (0,1)$ can be chosen in such way that for all $r > r_0$ and $\theta \in \mathcal{D}(\delta_0, s^*)$ it holds that
$$\delta_1 \leq \frac{\partial \log G(r,\theta)}{\partial r} \leq \frac{1}{\delta_1}. \tag{7}$$

These assumptions are verified in the specific case of (4) in Example 1 in Section 3. Furthermore, sums of call options with log-elliptical underlying assets also satisfy Assumptions 1 to 4 as we shall see in Example 2 in Section 3 .

We now are ready to state our result in the setting of our conditional Monte Carlo estimator.

**Theorem 1.** *Let $G(\cdot)$ satisfy Assumptions 1 to 4 and suppose that $f_R(\cdot)$ satisfies*

$$\lim_{x\to\infty} \frac{x f_R(x)}{\mathbb{P}(R > x)^{1-\varepsilon}} = 0 \qquad \forall \varepsilon > 0. \tag{8}$$

*Then, for every $\varepsilon > 0$ there exists $u_0 > 0$ such that if $u \geq u_0$*

$$\sup_{\theta \in \mathcal{S}_d} \mathbb{P}(G(R,\theta) > u)^{1-\varepsilon} \leq \mathbb{P}(G(R,\Theta) > u) \leq \sup_{\theta \in \mathcal{S}_d} \mathbb{P}(G(R,\theta) > u). \tag{9}$$

*Consequently, the conditional Monte Carlo estimator*

$$L(\Theta, u) = \mathbb{P}(G(R,\Theta) > u | \Theta) \tag{10}$$

*is asymptotically optimal.*

Let us discuss how to implement the estimator $L(\Theta, u)$ in (10) in the setting of $G(\cdot)$ defined as in (4). The implementation is done in two steps. First, we need to simulate $\Theta$ uniformly on $\mathcal{S}_d$; a standard procedure is to sample a $d$-dimensional vector of standard Gaussian random variables and normalize it by its Euclidian norm (for further details, see [9]). Second, given $\Theta = \theta$, we need to compute $L(\theta, u)$ which by independence is simply $\mathbb{P}(R \in \mathcal{A}_\theta(u))$ where

$$\mathcal{A}_\theta(u) := \{r \geq 0 : G(r, \theta) > u\}.$$

Typically, a root-finding numerical procedure such as Newton's method is required to determine $\mathcal{A}_\theta(u)$. For instance, for the function $G(r, \theta)$ as defined in (4) there are three possibilities depending on the sign of $\langle \mathbf{A}_i, \theta \rangle$, $i = 1, \ldots, d$.

1. $G(\cdot, \theta)$ is decreasing, which occurs if $\langle \mathbf{A}_i, \theta \rangle \leq 0$ for all $i$.
2. $G(\cdot, \theta)$ is increasing, which occurs if $\langle \mathbf{A}_i, \theta \rangle \geq 0$ for all $i$.
3. $G(\cdot, \theta)$ is strictly convex with a global minimum, which occurs if there exists $i \neq j$ such that $\langle \mathbf{A}_i, \theta \rangle < 0$ and $\langle A_{j,\cdot}, \theta \rangle > 0$.

Given these cases it is easy to show that the sets $\mathcal{A}_\theta(u)$ can only take the form $(0, r_-) \cup (r_+, \infty)$ with $0 \leq r_- \leq r_+ \leq \infty$. For instance, the case $\mathcal{A}_\theta(u) = \varnothing$, which is possible if $G(\cdot, \theta)$ is decreasing and $u$ is sufficiently large, is formally represented by choosing $r_- = 0$ and $r_+ = \infty$. Therefore, for the implementation it is only necessary to evaluate the cumulative distribution function of the random variable $R$ in at most two points; that is

$$\mathbb{P}(R \in \mathcal{A}_\theta(u)) = F_R(r_-) + 1 - F_R(r_+).$$

We provide further discussion on how to initialize the root-finding algorithm and locate $r_-$ and $r_+$ in our last section. As we shall see, the form $G(\cdot, \theta)$ is useful to guarantee fast global convergence.

### 2.2. Importance Sampling for a Class of Heavy-tailed Sums with Arbitrary Dependence

Although elliptical distributions have become popular models in practice, the fact is that the dependence structure in such models is limited. So, in order to cope with more general models we present a second result which involves a technique that allows to translate asymptotically optimal estimators for the tails of the marginal components into asymptotically optimal estimators for the tail of the sum. The decomposition implied by the number of components that exceeds a large threshold (a sum of pieces each involving marginal tail probabilities) facilitates the design of asymptotically optimal estimators; this will be illustrated with an example in the next section. However, we need to impose a suitable condition on the tail behavior of the marginal components. This condition is given in terms of the next definition.

**Definition 2.** We say that $Z$ is *logarithmically long-tailed* if for each $c \in (0, \infty)$

$$\lim_{b \to \infty} \frac{\log \mathbb{P}(Z > b - c)}{\log \mathbb{P}(Z > b)} = 1.$$

The term *logarithmically long-tailed* is borrowed from the literature on heavy-tailed random variables; cf. [22]. In that context a random variable $Z$ is said to be long-tailed if and only if $\lim_{b\to\infty}[\mathbb{P}(Z>b-c)/\mathbb{P}(Z>b)]=1$ for all $c>0$. Clearly, long-tailed random variables are logarithmically long-tailed and, in turn, every subexponential distribution is long-tailed. The class of logarithmically long-tailed distributions includes virtually any heavy-tailed distribution used in practice but also a large class of light-tailed distributions. In particular, the Gaussian and Gamma distributions are logarithmically long-tailed as well as their mixtures. However, one should point out that, although logarithmically long-tailed distributions provide substantial generality, not all distributions that arise naturally in practice can be cast in the framework of Definition 3. For instance, if $\mathrm{e}^{Z}$ is Weibull, then $Z$ is not logarithmically long-tailed

Our second result involves the use of importance sampling. Let $\widehat{\mathbb{P}}$ satisfy the absolute continuity condition which states that for every Borel set $A$, $\widehat{\mathbb{P}}(\mathbf{X}\in A,\mathrm{e}^{X_1}+\cdots+\mathrm{e}^{X_d}>u)=0$ implies $\mathbb{P}(\mathbf{X}\in A,\mathrm{e}^{X_1}+\cdots+\mathrm{e}^{X_d}>u)=0$. Then one can define the importance sampling estimator

$$\frac{d\mathbb{P}}{d\widehat{\mathbb{P}}}I\left(\mathrm{e}^{X_1}+\cdots+\mathrm{e}^{X_d}>u\right),$$

which is clearly unbiased. We now provide a precise statement of our result and a short and instructive proof.

**Theorem 2.** *Suppose that the $X_i$'s are logarithmically long-tailed. Then*

$$\frac{\log\alpha(u)}{\log\max_{i=1}^{d}\mathbb{P}(\mathrm{e}^{X_i}>u)}\longrightarrow 1,\qquad u\nearrow\infty. \tag{11}$$

*Moreover, if*

$$\widetilde{L}(\mathbf{X},b)=\frac{d\mathbb{P}}{d\widehat{\mathbb{P}}}(\mathbf{X},b)\sum_{i=1}^{d}I(X_i>b),$$

*is an asymptotically optimal estimator for $\mathbb{E}[\sum_{i=1}^{d}I(X_i>b)]$ as $b\nearrow\infty$, then by letting $b:=b(u)=\log(u)-\log(d)$ we obtain that the estimator*

$$L'(\mathbf{X},b(u))=\frac{d\mathbb{P}}{d\widehat{\mathbb{P}}}(\mathbf{X},b(u))\,I\left(\mathrm{e}^{X_1}+\cdots+\mathrm{e}^{X_d}>u\right) \tag{12}$$

*is asymptotically optimal for $\alpha(u)$.*

*Proof.* Throughout the proof we shall use frequently the following observation,

$$\{\mathrm{e}^{X_1}+...+\mathrm{e}^{X_d}>u\}\subseteq\bigcup_{i=1}^{d}\{\mathrm{e}^{X_i}>u/d\}. \tag{13}$$

In simple words, if $\mathrm{e}^{X_1}+\cdots+\mathrm{e}^{X_d}>u$ there is at least one $X_i$ such that $\mathrm{e}^{X_i}>u/d$. Using this observation and the Bonferroni inequality we obtain

$$\begin{aligned}\max_{i=1,\ldots,d}\mathbb{P}(X_i>\log u)\le\alpha(u)\le\mathbb{P}\left(\cup_{i=1}^{d}\{\mathrm{e}^{X_i}>u/d\}\right)&\le\sum_{i=1}^{d}\mathbb{P}(X_i\ge\log u-\log d)\\&\le d\max_{i=1,\ldots,d}\mathbb{P}(X_i>\log u-\log d).\end{aligned}$$

Since the $X_i$'s are logarithmically long-tailed the limit in (11) follows. Now we examine the performance of the simulation estimator induced by $\widetilde{L}(\mathbf{X},b)$. First, if $b(u) = \log(u) - \log(d)$ the estimator $L'(\mathbf{X},b(u))$ is well defined in the sense that the required absolute continuity condition is satisfied by virtue of (13). Hence we can write

$$L'(\mathbf{X},b(u)) = \widetilde{L}(\mathbf{X},b(u))\, I\left(\mathrm{e}^{X_1} + \cdots + \mathrm{e}^{X_d} > u\right).$$

Note that $\sum_{i=1}^{d} I(X_i > b(u))$ has disappeared from the left hand side due to (13). Now, by (1) all we need to verify in order to prove asymptotic optimality is that

$$\liminf_{u\to\infty} \frac{\log \widehat{\mathbb{E}}[L'(\mathbf{X},b(u))^2]}{2\log\alpha(u)} \geq 1.$$

However, for $u$ large enough it holds that $1 > \widehat{\mathbb{E}}[\widetilde{L}(\mathbf{X},b(u))^2] \geq \widehat{\mathbb{E}}[L'(\mathbf{X},b(u))^2]$. Therefore

$$\frac{\log \widehat{\mathbb{E}}[L'(\mathbf{X},b(u))^2]}{2\log\alpha(u)} \geq \frac{\log \widehat{\mathbb{E}}[\widetilde{L}(\mathbf{X},b(u))^2]}{2\log \sum\limits_{i=1}^{d} \mathbb{P}(X_i \geq b(u))} \times \frac{\log \sum\limits_{i=1}^{d} \mathbb{P}(X_i \geq b(u))}{\log\alpha(u)}. \tag{14}$$

By assumption we have that $\widetilde{L}(\mathbf{X},b(u))$ is asymptotically optimal. Therefore,

$$\liminf_{u\to\infty} \frac{\log \widehat{\mathbb{E}}[\widetilde{L}(\mathbf{X},b(u))^2]}{2\log \sum_{i=1}^{d} \mathbb{P}(X_i \geq b(u))} = 1,$$

and because the $X_i$'s are logarithmically long-tailed we obtain that

$$\lim_{u\to\infty} \frac{\log \sum_{i=1}^{d} \mathbb{P}(X_i \geq b(u))}{\log\alpha(u)} = 1.$$

By combining these two observations after taking limits in (14) we obtain the result. □

### 2.3. The proof of Theorem 1.

*Proof of Theorem 1.* To simplify the notation define $H(r,\theta) = \log G(r,\theta)$. The upper bound in (9) follows directly by independence:

$$\mathbb{P}(G(R,\Theta) > u) = \mathbb{E}[\mathbb{P}(G(R,\Theta) > u|\Theta)] \leq \sup_{\theta\in\mathcal{S}_d} \mathbb{P}(G(R,\theta) > u).$$

Now we proceed to prove the lower bound in (9). We claim that for large enough $u$,

$$\sup_{\theta\in\mathcal{S}_d} \mathbb{P}(G(R,\theta) > u) = \sup_{\theta\in\mathcal{D}(\delta_0,s^*)} \mathbb{P}(G(R,\theta) > u). \tag{15}$$

To see this, note that since $G(\cdot,\theta)$ is eventually monotone and increasing for $\theta \in \mathcal{D}(\delta_0,s^*)$ and due to (7) we can define the inverse $G^{-1}(\cdot,\theta)$ for each $\theta \in \mathcal{D}(\delta_0,s^*)$ over an interval $[u_0, \infty)$ by selecting $u_0$ sufficiently large. Therefore, for all $u > u_0$,

$$
\begin{aligned}
\sup_{\theta\in\mathcal{S}_d} \mathbb{P}(G(R,\theta) > u) \leq \mathbb{P}\Big(\sup_{\theta\in\mathcal{S}_d} G(R,\theta) > u\Big) &= \mathbb{P}\Big(\sup_{\theta\in\mathcal{D}(\delta_0,s^*)} G(R,\theta) > u\Big) \\
&= \mathbb{P}\Big(R \in \bigcup_{\theta\in\mathcal{D}(\delta_0,s^*)} \{r : G(r,\theta) > u\}\Big) \\
&= \mathbb{P}\Big(R > \inf_{\theta\in\mathcal{D}(\delta_0,s^*)} G^{-1}(u,\theta)\Big) \\
&= \sup_{\theta\in\mathcal{D}(\delta_0,s^*)} \mathbb{P}(G(R,\theta) > u),
\end{aligned}
$$

where the first equality in the first line follows from (6). The reverse inequality is immediate and therefore the claim follows.

The identity (15) allows us to concentrate in developing a lower bound in the region $\mathcal{D}(\delta_0, s^*)$ as we do now. By virtue of (6) and (5) it follows that for all $\delta \leq \min(\delta_0, 1/(2v))$ there exists a $u_0$ such that for all $u > u_0$ and for all $\theta \in \mathcal{D}(\delta_0, s^*)$ it holds that

$$
\begin{aligned}
\mathbb{P}\left(G\left(R,\Theta\right) > u\right) &\geq \mathbb{P}\left(G\left(R,\Theta\right) > u, ||\Theta - s^*|| \leq \delta\right) \\
&\geq \mathbb{P}(G\left(R,\theta\right)^{1-v\delta} > u, ||\Theta - s^*|| \leq \delta) \\
&\geq c\mathbb{P}(G\left(R,\theta\right)^{1-v\delta} > u)\delta^{d-1} \\
&= c\mathbb{P}(H\left(R,\theta\right) > b/(1-v\delta))\delta^{d-1} \\
&\geq c\mathbb{P}(H\left(R,\theta\right) > b + 2\delta v b)\delta^{d-1}.
\end{aligned}
$$

The third inequality is obtained by independence and the fact that $\Theta$ is uniformly distributed over $\mathcal{S}_d$. Let $\Lambda_\theta\left(\cdot\right)$ be the hazard function of $H(R,\theta)$, i.e., $\mathbb{P}(H(R,\theta) > b) = \exp(-\Lambda_\theta(b))$, and select $\gamma_\theta(b)$ satisfying

$$
\Lambda_\theta(b) - \Lambda_\theta(b + \gamma_\theta(b)) = -1.
$$

Since the density of $R$ exists and is eventually positive and $H(\cdot,\theta)$ is eventually continuously differentiable and strictly increasing (due to (7)), then the hazard function is not only eventually strictly increasing but it is also eventually differentiable. Let $\delta := \delta_\theta(b) = \min\{\gamma_\theta(b)/[2vb], 1/[2v], \delta_0\}$ and obtain that

$$
\begin{aligned}
\mathbb{P}(H\left(R,\theta\right) > b + 2\delta b)\delta^{d-1} &\geq \exp\big\{-\Lambda\big(b + \gamma_\theta(b)\big)\big\}\left(\delta_\theta\left(b\right)\right)^{d-1} \\
&= \exp\{-1 - \Lambda_\theta(b)\}\left(\delta_\theta\left(b\right)\right)^{d-1}.
\end{aligned}
$$

Now we claim that for each $\varepsilon > 0$ there exists $b_0 > 0$ such that

$$
\frac{\gamma_\theta(b)}{b} = \frac{\Lambda_\theta^{-1}(\Lambda_\theta\left(b\right) + 1) - b}{b} \geq \exp\left(-\varepsilon\Lambda_\theta\left(b\right)\right). \tag{16}
$$

for all $b \geq b_0$ and every $\theta \in \mathcal{D}(\delta_0, s^*)$. If we let $\Lambda_\theta\left(b\right) = y$ then it suffices to establish (due to the inequality $\exp\left(-\varepsilon y\right) + 1 \leq \exp\left(\exp\left(-\varepsilon y\right)\right)$) that there exists $y_0$ such that

$$
\frac{\Lambda_\theta^{-1}(y+1)}{\Lambda_\theta^{-1}(y)} \geq \exp\left(\exp\left(-\varepsilon y\right)\right), \tag{17}
$$

for all $y \geq y_0$ and all $\theta \in \mathcal{D}(\delta_0, s^*)$. Now, let us write $f_\theta(x)$ and $\overline{F}_\theta(x)$ for the density and the tail distribution respectively of $H(R,\theta)$ evaluated at $x$. Let

$$\beta_\theta(y) := \frac{\partial}{\partial y} \log \Lambda_\theta^{-1}(y) = \frac{\overline{F}_\theta\left(\Lambda_\theta^{-1}(y)\right)}{\Lambda_\theta^{-1}(y)\, f_\theta\left(\Lambda_\theta^{-1}(y)\right)}.$$

We can select $y_0$ independent of $\theta$ so that

$$\exp\left(\int_{y_0}^{y} \beta_\theta(s)\, ds\right) = \frac{\Lambda_\theta^{-1}(y)}{\Lambda_\theta^{-1}(y_0)}.$$

Therefore, in order to conclude (17) it suffices to show that

$$\exp(\varepsilon y) \int_y^{y+1} \beta_\theta(s)\, ds \geq \exp(-\varepsilon) \int_y^{y+1} \exp(\varepsilon s)\, \beta_\theta(s)\, ds \geq 1. \tag{18}$$

for all $y \geq y_0$. In fact, we will show that the function $\exp(\varepsilon y)\beta_\theta(y)$ can be made arbitrarily large as $y \nearrow \infty$. Notice that for $y > y_0$

$$\exp(\varepsilon y)\, \beta_\theta(y) = \frac{\exp(\varepsilon y)\, \overline{F}_\theta\left(\Lambda_\theta^{-1}(y)\right)}{\Lambda_\theta^{-1}(y)\, f_\theta\left(\Lambda_\theta^{-1}(y)\right)} = \exp(\varepsilon \Lambda_\theta(b))\, \frac{\overline{F}_\theta(b)}{b f_\theta(b)} = \frac{\overline{F}_\theta^{1-\epsilon}(b)}{b f_\theta(b)}$$

Let $\Gamma_\theta(\cdot)$ be defined such that $\Gamma_\theta(H(b,\theta)) = b$ for all $b > b_0$, then according to (7) it follows that for all $\theta \in \mathcal{D}(\delta_0, s^*)$ the following inequalities hold

$$\frac{f_R(\Gamma_\theta(b))}{\delta_1} \geq f_\theta(b) = \frac{f_R(\Gamma_\theta(b))}{\dot{H}(\Gamma_\theta(b), \theta)} \geq \delta_1 f_R(\Gamma_\theta(b)), \tag{19}$$

where $\dot{H}(r,\theta) := \partial H(r,\theta)/\partial r$. Now, using (19) and (7) once again we obtain that

$$\overline{F}_\theta(b) = \exp(-\Lambda_\theta(b)) \geq \delta_1 \int_b^\infty f_R(\Gamma_\theta(s))\, ds \geq \delta_1^2 \int_{\Gamma_\theta(b)}^\infty f_R(u)\, du = \delta_1^2 \overline{F}_R(\Gamma_\theta(b)).$$

The second inequality was obtained by noting that (7) implies that $\partial\Gamma_\theta(b)/\partial b \in [\delta_1, \delta_1^{-1}]$ if $b$ is sufficiently large. Therefore,

$$\frac{\overline{F}_\theta^{1-\epsilon}(b)}{b f_\theta(b)} \geq \frac{\Gamma_\theta(b)}{b} \times \frac{\delta_1^{3+2\varepsilon}\overline{F}_R(\Gamma_\theta(b))^{1-\varepsilon}}{\Gamma_\theta(b)\, f_R(\Gamma_\theta(b))} \geq \frac{c\delta_1^{3+2\varepsilon}\overline{F}_R(\Gamma_\theta(b))^{1-\varepsilon}}{\Gamma_\theta(b)\, f_R(\Gamma_\theta(b))},$$

for some constant $c \in (0,\infty)$, due to (7). It follows from (2) that the right hand side in the previous inequality can be made arbitrarily large if $y_0$ is sufficiently large and therefore, in particular, inequality (16) follows. We thus conclude that for all $\varepsilon > 0$ there exists a $u_0 > 0$ such that for all $\theta \in \mathcal{D}(\delta_0, s^*)$ and all $u \geq u_0$ it holds that

$$\mathbb{P}(G(R,\Theta) > u) \geq \kappa \mathbb{P}(G(R,\theta) > u)^{1-\varepsilon},$$

where $\kappa$ is a suitable constant depending on $u_0$ but not on $\theta$. Taking the supremum on the right hand side over $\theta \in \mathcal{D}(\delta_0, s^*)$ we obtain (9). Finally, the fact that $L(\Theta, u)$

is asymptotically optimal is almost immediate. Namely, if $u$ is sufficiently large and $\varepsilon > 0$ is small enough, then

$$\frac{\mathbb{E}L\left(\Theta,u\right)^{2}}{\mathbb{P}\left(G\left(R,\Theta\right)>u\right)^{2-\varepsilon}}\leq\frac{\sup\limits_{\theta\in\mathcal{S}_{d}}\mathbb{P}\left(G\left(R,\theta\right)>u\right)^{2}}{\sup\limits_{\theta\in\mathcal{S}_{d}}\mathbb{P}\left(G\left(R,\theta\right)>u\right)^{(2-\varepsilon)(1-\varepsilon)}}=\sup_{\theta\in\mathcal{S}_{d}}\mathbb{P}\left(G\left(R,\theta\right)>u\right)^{2\varepsilon-\varepsilon^{2}}\leq 1,$$

thereby concluding the result.

## 3. Applications and Examples

We now discuss how our results can be applied to a number of models that are popular in applications to finance and risk management. In Subsection 3.1 we illustrate the results of Theorem 1 with Examples while in Subsection 3.2 we provide an example for the use of the results in Theorem 2.

### 3.1. Illustrating Conditional Monte Carlo: Theorem 1.

**Example 1.** (*Sums of Log-elliptical Distributions.*) Let $G\left(r,\theta\right)$ as defined in (4). We will verify that conditions (5), (6) and (7) are satisfied. First, since $\theta \in \mathcal{S}_d$, the Cauchy-Schwartz inequality implies that $|\langle \mathbf{A}_i, \theta\rangle| \leq ||\mathbf{A}_i||$ where the equality is attained by choosing $\theta_i = \mathbf{A}_i / ||\mathbf{A}_i||$. Therefore, we select $s^* = \theta_{i^*}$ with $i^*$ such that $\|\mathbf{A}_{i^*}\| = \max_{i=1}^{d} \|\mathbf{A}_i\|$. If $||\theta - s^*|| \leq \delta$, then again by Cauchy-Schwartz inequality,

$$\begin{aligned}\exp\left(\mu_i + r\left\langle \mathbf{A}_i, \theta\right\rangle\right) &= \exp\left(\mu_i + r\left\langle \mathbf{A}_i, s^*\right\rangle + r\left\langle \mathbf{A}_i, \theta - s^*\right\rangle\right)\\ &\geq \exp\left(\mu_i + r\left\langle \mathbf{A}_i, s^*\right\rangle - r\delta\|\mathbf{A}_i\|\right)\\ &\geq \exp\left(\mu_i + r\left\langle \mathbf{A}_i, s^*\right\rangle - r\delta\|\mathbf{A}_{i^*}\|\right).\end{aligned}$$

Therefore, if $||\theta - s^*|| \leq \delta$

$$G(r,\theta) = \sum_{i=1}^{d} \exp\left(\mu_i + r\left\langle \mathbf{A}_i, \theta\right\rangle\right) \geq \exp(-r\delta\|\mathbf{A}_{i^*}\|)G(r,s^*).$$

However, clearly there is a large enough $r_0 > 0$, chosen independent of $\delta$ and $\theta \in \mathcal{S}_d$, such that $G(r,\theta) \leq G(r,s^*)$ if $r \geq r_0$. We then conclude that if $\delta > 0$, then

$$\inf_{\theta\in\mathcal{D}(\delta,s^*)} G\left(r,\theta\right) \geq \exp(-r\delta\|\mathbf{A}_{i^*}\|) \sup_{\theta\in\mathcal{S}_d} G\left(r,\theta\right).$$

Now we need to show that one can choose $v > 0$ such that

$$\exp(-r\delta\|\mathbf{A}_{i^*}\|)G(r,s^*)^{v\delta} \geq 1,$$

for $r$ large enough, but this follows easily by choosing $v > \|\mathbf{A}_{i^*}\|$. So, the parameter $\delta_0 > 0$ can be selected arbitrary for (5). Similarly, bounds (6) and (7) are easily seen to be satisfied.

**Example 2.** (*Sums of Call Options with Log-elliptical Underlying.*)

A call option gives the owner the right to buy an underlying asset at a so-called strike price $K > 0$ and at some maturity time $T$ in the future. The profit at maturity time is therefore $(S_T - K)^+$, where $S_t$ is the price of the underlying asset at time

$0 \le t \le T$. It is well known that the price at time $0 \le t \le T$ satisfies (assuming zero interest rates for simplicity) $\mathbb{E}[(S_T - K)^+|S_t]$, for a suitably defined expectation; see for example chapter 5 of [20].

Several popular models in finance, such as the Black-Scholes model, allow to express $S_T = S_t \exp(Z)$, where $Z$ is a random variable independent of $S_t$, but clearly depending on $T - t$. Thus, in these types of situations, if one has $d$ underlying assets following a joint log-elliptical distribution, the value of a portfolio at some time $t > 0$ containing $d$ call options, with strike prices $K_1, ..., K_d$, can be expressed as $\sum_{i=1}^d \mathbb{E}[(\exp(X_i) Z_i - K_i)^+|X_i]$, where the $Z_i$'s are positive random variables with finite mean and following a distribution that depends on the maturity time of each of the contracts. Assume in what follows that the $Z_i$'s have a continuous distribution with infinite support.

Using the representation $X_i = \mu_i + R\langle \mathbf{A}_i, \Theta\rangle$ we then conclude that in order to analyze the tail of the distribution of a portfolio of call options with log-elliptical underlying price assets we can apply Theorem 1 to the function

$$G(r,\theta) = \sum_{i=1}^{d} \mathbb{E}[\,(\exp(\mu_i + r\langle \mathbf{A}_i, \theta\rangle) Z_i - K_i)^+] \qquad \theta \in \mathcal{S}_d. \tag{20}$$

Note that

$$G(r,\theta) \sim \sum_{i=1}^{d} \exp(\mu_i + r\langle \mathbf{A}_i, \theta\rangle)\, \mathbb{E}[Z_i]$$

as $r \to \infty$ if and only if $\theta$ is such that

$$\sum_{i=1}^{d} \exp(\mu_i + r\langle \mathbf{A}_i, \theta\rangle) \longrightarrow \infty \tag{21}$$

as $r \to \infty$. Therefore, (5) and (6) are completely analogous to Example 1. In order to verify (7) one can use dominated convergence (here we use that $Z_i$ has a continuous distribution) to conclude that

$$\begin{aligned} \frac{d \log G(r,\theta)}{dr} &= \sum_{j=1}^{d} \frac{\langle \mathbf{A}_j, \theta\rangle\, \mathbb{E}[\exp(\mu_j + r\langle \mathbf{A}_j, \theta\rangle) Z_j I(\exp(\mu_i + r\langle \mathbf{A}_j, \theta\rangle) Z_j \ge K_j)]}{\sum_{i=1}^{d} \mathbb{E}[\,(\exp(\mu_i + r\langle \mathbf{A}_i, \theta\rangle) Z_i - K_i)^+]} \\ &\sim \sum_{j=1}^{d} \frac{\langle \mathbf{A}_j, \theta\rangle \exp(\mu_j + r\langle \mathbf{A}_j, \theta\rangle)\, \mathbb{E}[Z_j]}{\sum_{i=1}^{d} \exp(\mu_i + r\langle \mathbf{A}_i, \theta\rangle)\, \mathbb{E}[Z_i]}. \end{aligned}$$

as $r \to \infty$ if and only if $\theta$ is such that (21) holds as $r \to \infty$. So, (7) also holds as in Example 1.

A problem that arises in the implementation of the corresponding conditional Monte Carlo estimator in this setting is that one must be able to evaluate in closed form $\mathbb{E}[(\exp(X_i) Z_i - K_i)^+|X_i]$. This can be done, for instance, in the Black-Scholes model. In this setting, the root finding procedure necessary to evaluate $L(\Theta, u)$ in (10) is entirely analogous to that explained at the end of Subsection 2.1.

**Example 3.** (*Symmetric Generalized Hyperbolic Distributions.*)

A random variable $W$ is said to have a Generalized Inverse Gaussian distribution (GIG) with parameters $(\lambda,\chi,\psi)$ in the set defined by

$$\Lambda := \begin{cases} \lambda\in\mathbb{R}, \chi>0, \psi>0 \\ \lambda<0, \chi>0, \psi=0 \\ \lambda>0, \chi=0, \psi>0 \end{cases}$$

if its density function is given by

$$f_W(w) = \frac{(\psi\chi)^{d/2}}{2K_\lambda(\sqrt{\psi\chi})} w^{\lambda-1}\exp\left\{-\frac{1}{2}(\chi w^{-1}+\psi w)\right\}, \qquad w>0.$$

where $K_\lambda$ is the modified Bessel function of third kind with index $\lambda$. We denote it $W\sim\mathcal{N}^-(\lambda,\chi,\psi)$. Important cases of the GIG family are the limiting cases of the Gamma ($\lambda>0$,$\chi=0$,$\psi>0$) and the Inverse Gamma ($\lambda<0$,$\chi>0$,$\psi=0$). The Normal Inverse Gaussian NIG occurs when $\lambda=-1/2$ and the Hyperbolic when $\lambda=1$. Notice that for parameter values not contained in $\Lambda$ the function $f_W$ is not a density function. For further details see, for example, [26].

The family of elliptical distributions generated by a radial random variable with stochastic representation

$$R \stackrel{d}{=} \sqrt{\tau\chi_d^2},$$

with $\tau$ having a GIG distributions is known as *symmetric generalized hyperbolic* (SGH). The density of the radial component of a SGH distribution is given by

$$f_R(r) := \begin{cases} \frac{\chi^{-\lambda/2}\psi^{d/4}}{2^{d/2-1}\Gamma(d/2)K_\lambda\left(\sqrt{\chi\psi}\right)}\frac{r^{d-1}K_{d/2-\lambda}\left(\sqrt{\psi(\chi+r^2)}\right)}{(\chi+r^2)^{d/4-\lambda/2}} & \lambda\in\mathbb{R}, \chi>0, \psi>0 \\ \frac{2\chi^{-\lambda}}{\mathsf{Beta}(-\lambda,\mathsf{d}/2)}\, r^{d-1}(\chi+r^2)^{\lambda-d/2} & \lambda<0, \chi>0, \psi=0 \\ \frac{\psi^{\lambda/2+d/4}}{2^{\lambda+d/2-2}\Gamma(\lambda)\Gamma(d/2)} r^{\lambda+d/2-1}K_{d/2-\lambda}\left(\sqrt{\psi}r\right) & \lambda>0, \chi=0, \psi>0 \end{cases}$$

where $\Gamma$ and $\mathsf{Beta}$ are the Gamma and Beta functions. Next, we prove that the density of the radial component of a SGH distribution satisfies (8) in each of the three cases above. For doing so, the asymptotic expansion of

$$K_\lambda(w) = (\pi/2w)^{-1/2}\mathrm{e}^{-w}(1+o(w^{-1})) \tag{22}$$

is used. The facts that $K_\lambda(w)=K_{-\lambda}(w)$ and $K'_\lambda(w)=\lambda K_\lambda(w)/w - K_{\lambda+1}(w)$ are also used (for further details see [1], page 374).

1. **Case 1 ($\lambda\in\mathbb{R}$, $\psi>0$, $\chi>0$):** Two of the most prominent examples are the multivariate symmetric hyperbolic distribution ($\lambda\in\mathbb{N}$) and the multivariate Normal Inverse Gaussian (NIG) distribution ($\lambda=1/2$). Using L'Hôpital's rule and (22) we verify that

$$\lim_{r\to\infty}\frac{rf_R(r)}{\overline{F}_R^{1-\epsilon}(r)} = -\lim_{r\to\infty}\frac{1-d+r^2\sqrt{\frac{\psi}{\chi+r^2}}}{(1-\epsilon)\overline{F}_R^{-\epsilon}(r)} = -\frac{\sqrt{\psi}}{1-\epsilon}\left(\lim_{r\to\infty}\frac{\overline{F}_R(r)}{r^{-1/\epsilon}}\right)^{\epsilon}$$

The limit inside the brackets is equivalent to

$$\lim_{r\to\infty} \epsilon\; k\; r^{d+1/\epsilon}(\chi+r^2)^{\lambda/2-d/4}\sqrt{\frac{\pi}{2\sqrt{\psi(\chi+r^2)}}}\mathrm{e}^{-\sqrt{\psi(\chi+r^2)}} = 0$$

2. **Case 2 ($\lambda < 0$, $\chi > 0$, $\psi = 0$):** This boundary case occurs when the mixing random variable $W$ has an Inverse Gamma distribution. A classical example is that of the multivariate $t$ distribution ($\lambda = -\nu/2, \chi = 1, \psi = 0$). The limit is given by

$$\lim_{r\to\infty}\frac{rf_R(r)}{\overline{F}_R^{1-\epsilon}(r)} = \lim_{r\to\infty}\frac{1-\big(d\chi+2r^2\lambda\big)/(\chi+r^2)}{(1-\epsilon)\overline{F}_R^{-\epsilon}(r)} = \frac{1-2\lambda}{1-\epsilon}\lim_{r\to\infty}\overline{F}_R^{\epsilon}(r) = 0$$

3. **Case 3 ($\lambda > 0$,$\chi = 0$, $\psi > 0$):** The second boundary case corresponds to multivariate distributions known as Laplace, Bessel or Variance-Gamma and occurs when $W$ has a Gamma distribution. This model has been applied in finance as an alternative to the Black-Scholes model and has become popular because, among other features, it allows to incorporate heavier tails in the log-returns; a stylized feature that has been observed in financial data [32]. In the multivariate Variance-Gamma process for price dynamics, discussed in Section 5 of Cont and Tankov [19], the vector $(X_1,\ldots,X_d)$ of the log-prices of $d$ assets follows a multivariate Variance-Gamma distribution. It turns out that

$$\lim_{r\to\infty}\frac{rf_R(r)}{\overline{F}_R^{1-\epsilon}(r)} = -\frac{1}{1-\epsilon}\lim_{r\to\infty}\frac{\sqrt{\chi}r-d+1}{\overline{F}^{-\epsilon}(r)} = -\frac{\sqrt{\chi}}{1-\epsilon}\left(\lim_{r\to\infty}\frac{\overline{F}_R(r)}{r^{-1/\epsilon}}\right)^{\epsilon}.$$

The limit inside the brackets is equal to

$$\lim_{r\to\infty}\frac{k\epsilon r^{\lambda+d/2-1}K_{d/2-\lambda}(\sqrt{\psi}r)}{r^{-1-1/\epsilon}} = \lim_{r\to\infty} k\epsilon r^{\lambda+d/2+1/\epsilon}\sqrt{\frac{\pi}{2\sqrt{\psi}r}}\mathrm{e}^{-\sqrt{\psi}r} = 0$$

### 3.2. Illustrating Importance Sampling: Theorem 2

**Example 4.** (*Kou model.*)

The SGH distributions form a subfamily of a larger class of distributions known as Generalized Hyperbolic distributions (GH) which were introduced by Barndorff-Nielsen in [12]. A random vector is said to have a GH distribution if it has the following stochastic representation

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \tau\mathbf{m} + \sqrt{\tau\chi_d^2}\,\mathbf{C}\,\Theta,$$

where $\boldsymbol{\mu},\mathbf{m}\in\mathbb{R}^d$, $\mathbf{C}\in\mathbb{R}^{d\times d}$, $\tau$ is random variable with a GIG distribution, $\chi_d^2$ is a chi-square random variable with $d$ degrees of freedom and $\Theta$ is a random vector uniformly distributed on $\mathcal{S}_d$. In particular, if $\boldsymbol{\mu} = \mathbf{0}$ and $\tau\sim\exp(1)$, then $\mathbf{X}$ is said to follow a $d$-dimensional *Asymmetric Laplace* distribution with parameters $\mathbf{m}$ and $\mathbf{G} := \mathbf{C}\mathbf{C}^T$ and is denoted by $\mathcal{AL}_d(\mathbf{m},\mathbf{G})$.

The following is a multivariate asset pricing model proposed in [25] as an extension of the popular Kou model [31] (see also [19]). In order to specify the model, we introduce

$\boldsymbol{\mu} \in \mathbb{R}^d$ and a positive definite matrix $\boldsymbol{\Sigma}$ with decomposition $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$. Under such a model the vector of log-returns evaluated at time $t$ takes the form

$$\mathbf{X}(t) = \mathbf{X}(0) + (\boldsymbol{\mu} - \mathbf{D}/2)t + \mathbf{A}\mathbf{B}(t) + \sum_{j=1}^{N_0(t)} \mathbf{Y}_j + \sum_{i=1}^{d} \sum_{j=1}^{N_i(t)} \mathbf{e}_i W_{i,j}, \tag{23}$$

where $\mathbf{D}_i = \Sigma_{i,i}$, $\mathbf{e}_i$ is the $i$-th canonical vector, $\{\mathbf{B}(t), t \geq 0\}$ is a $d$-dimensional standard Brownian motion, $\{N_k(t) : t \geq 0\}$, $k = 0, \ldots, d$, are $d+1$ independent homogenous Poisson processes with parameters $\{\lambda_k : k = 0, \ldots, d\}$, $\{\mathbf{Y}_j : j \geq 1\}$ is a sequence of independent $d$-dimensional random vectors with common distribution $\mathcal{AL}_d(\mathbf{m}, \mathbf{G})$ and $\{W_{i,j} : j \geq 1\}$ are sequences of i.i.d. random variables with common distribution $\mathcal{AL}_1(\nu_i, \beta_i)$, $i = 1, \ldots, d$.

We are interested in using the estimator (12) in order to approximate

$$\alpha(u) = \mathbb{P}\left(\mathrm{e}^{X_1(t)} + \cdots + \mathrm{e}^{X_d(t)} > u\right),$$

where $X_i(t)$ is the $i$-th component of the vector $\mathbf{X}(t)$ for some fixed time $t$. The idea is to apply exponential tilting to the vector (23) in order to estimate $\mathbb{P}(X_i(t) > b)$. Then, we propose an importance sampling distribution of the form

$$\frac{d\widehat{\mathbb{P}}}{d\mathbb{P}}(\mathbf{X}(t)) = \sum_{i=1}^{d} w_i \exp(\gamma_i X_i(t) - \psi(\gamma_i \mathbf{e}_i))$$

where the weights $w_i > 0$ are such that $\sum_{i=1}^d w_i = 1$, $\psi(\theta) := \log \mathbb{E} \exp\{\langle \theta, \mathbf{X}(t) \rangle\}$ is the log-moment generating function of $\mathbf{X}(t)$ and the $\gamma_i$'s are constants in the domains of convergence of the Laplace transforms of the $X_i$'s and are chosen as follows. Notice that

$$\psi(\gamma_i \mathbf{e}_i) = X_i(0)\,\gamma_i + t\left[\left(\mu_i - \frac{\Sigma_{i,i}}{2}\right)\gamma_i + \frac{\Sigma_{i,i}}{2}\gamma_i^2 - \lambda_0 - \lambda_i \right.$$
$$\left. + \frac{\lambda_0}{1 - \gamma_i^2 G_{i,i}/2 - \gamma_i m_i} + \frac{\lambda_i}{1 - \gamma_i^2 \beta_i/2 - \gamma_i \nu_i}\right].$$

Thus, it is easy to establish that for any selection of weights $w_i > 0$ an asymptotically efficient estimator for $\sum_i \mathbb{P}(X_i(t) > b)$ as $b \nearrow \infty$ can be obtained by solving

$$\frac{\partial}{\partial \gamma_i}\psi(\gamma_i \mathbf{e}_i) = b$$

for $\gamma_i > 0$, $i = 1, \ldots, d$ in the domain of convergence of the Laplace transforms of the $X_i$'s (cf. [9]). In fact, it can be recognized that the proposed importance sampling corresponds to a model such as (23) with modified parameters

$$\mu_i^* := \mu + \gamma_i \Sigma_{i,\cdot}$$

$$\lambda_{0,i}^* := \frac{\lambda_0}{1 - \gamma_i^2 G_{i,i}/2 - \gamma_i m_i} \qquad \lambda_{i,i}^* := \frac{\lambda_i}{1 - \gamma_i^2 \beta_i/2 - \gamma_i \nu_i}$$

$$G_i^* := \frac{\lambda_{0,i}^*}{\lambda_0} G \qquad m_i^* := \frac{\lambda_{0,i}^*}{\lambda_0}(m + \gamma_i G_{i,\cdot})$$

$$\beta_{i,i}^* := \frac{\lambda_{i,i}^*}{\lambda_0}\beta_i \qquad \nu_{i,i}^* := \frac{\lambda_{i,i}^*}{\lambda_0}(\nu_i + \gamma_i \beta_i)$$

Notice that the parameters $\{\beta^*_{i,j} : j \neq i\}$ and $\{\nu^*_{i,j} : j \neq i\}$ are left unchanged. By standard large deviations techniques, and taking advantage of the change of measure suggested above for $X_i(t)$, it follows that $\mathbb{P}(X_i(t) > b) = \exp(-\gamma_i^* b + o(b))$ and therefore $X_i(t)$ is logarithmically long-tailed. In Section 4 we show a numerical example (for more details, see [25]).

## 4. Implementation and Numerical Examples

In this section we discuss the implementation of (10) for the case

$$G_1(r,\theta) := \sum_{i=1}^{d} \exp\{\mu_i + r\langle \mathbf{A}_i, \theta\rangle\}.$$

Notice that for every replication of the estimator we generate $\Theta = \theta$ and then we solve $G_1(r,\theta) = u$ using a numerical algorithm. The main issue here is that most iterative methods are not guaranteed to converge and their performance is largely affected by the shape of the function and the initial guess.

In the problem at hand, the functions $\{G_1(\cdot,\theta) : \theta \in \mathcal{S}_d\}$ are smooth; under this setting a root-finding algorithm (such as Newton-Raphson) will converge rather quickly provided that the initial guess is chosen close to the solution but more importantly that their successive iterations do not lie in a region where the derivative of the function is too close to 0. In general, a fixed initial guess will deliver poor results since we cannot assure that it will be a *good* initial guess for all functions in $\{G_1(\cdot,\theta) : \theta \in \mathcal{S}_d\}$. Indeed, in our numerical implementations, when we used a fixed initial value we observed that the algorithm failed to converge for several values of $\theta \in \mathcal{S}_d$; this occurred more often when all values of $\langle \mathbf{A}_i, \theta\rangle$ were close to 0.

Taking into consideration this observation we propose a set of initial values which help to improve dramatically the speed of convergence. The idea behind this proposal is that the tail probability of the maximum of $d$ positive random variables can be used to approximate the tail probability of the convolution. We define

$$G_2(r,\theta) = \max_{i=1,\ldots,d} \exp\{\mu_i + r\langle \mathbf{A}_i, \theta\rangle\}.$$

It is straightforward to prove that the CMC estimator in (10) for $\mathbb{P}(G_2(R,\Theta) > u)$ is given by

$$F_R(m_-) + 1 - F_R(m_+)$$

where

$$m_- := \sup_i \left\{ \frac{\log u - \mu_i}{\langle \mathbf{A}_i, \theta\rangle} : \langle \mathbf{A}_i, \theta\rangle < 0 \right\}, \quad m_+ := \inf_i \left\{ \frac{\log u - \mu_i}{\langle \mathbf{A}_i, \theta\rangle} : \langle \mathbf{A}_i, \theta\rangle < 0 \right\},$$

with the usual conventions $\sup\emptyset = -\infty$ and $\inf\emptyset = \infty$. Moreover, it is easy to verify that $m_- \leq r_-$ and $r_+ \leq m_+$. The idea is that the values $m_-$ and $m_+$ are not only close to $r_-$ and $r_+$ respectively but also lie in a region where the derivatives are not too close to 0. The procedure is described next.

1. If $\langle \mathbf{A}_i, \theta\rangle \leq 0$ for all $i = 1, \ldots, d$, then $G_1(r,\theta)$ is strictly decreasing with exactly one root. We use $m_-$ as initial value to find $r_-$ and we set $r_+ = \infty$.

2. If $\langle \mathbf{A}_i, \theta \rangle \geq 0$ for all $i = 1, \ldots, d$, then $G_1(r, \theta)$ is strictly increasing with exactly one root. We use $m_+$ as initial value to find $r_+$ and we set $r_- = 0$.

3. If there exists $i \neq j$ such that $\langle \mathbf{A}_i, \theta \rangle < 0$ and $\langle \mathbf{A}_j, \theta \rangle > 0$ and the global minima is smaller than $u$ then there exist two roots. In such case we run the root-algorithm twice; each time with the initial values $m_-$ and $m_+$.

   Notice that if $u$ is large enough (a common feature in a rare-event setting) it is enough to check that $G(0, \theta) < u$ in order to verify that the global minima is smaller than $u$.

We illustrate this algorithm with the following example.

**Example 5.** (*Variance-Gamma.*) We implemented the first algorithm for estimating the probability of $\alpha(u) := \mathbb{P}(\mathrm{e}^{X_1} + \cdots + \mathrm{e}^{X_d} > u)$ where $\mathbf{X}$ follows a multivariate Variance-Gamma distribution. That is $\mathbf{X} \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_R)$ where $F_R$ is such that $R^2 \sim \mathcal{N}^-(\lambda, 0, \psi)$ with $\lambda > 0$ and $\psi > 0$. The parameters used are $\mu_i = -i$, $\Sigma_{i,i} = 11 - i$, $\Sigma_{i,j} = 0.4\sqrt{\Sigma_{i,i}\Sigma_{j,j}}$ for $i \neq j$ and $R^2 \sim \mathcal{N}^-(1, 0, 4)$. A total of $10^5$ replications were used to obtain the estimations.

The numerical results are summarized in Table 1. Estimated values of the expected value, standard deviation and variation coefficient of the estimator (10) for the corresponding values of $u$ are given. The results are complemented with cpu times necessary the $10^5$ replications of the estimator.

TABLE 1: Statistics for the estimator of $\mathbb{P}(\mathrm{e}^{X_1} + \cdots + \mathrm{e}^{X_{10}} > u)$, where $\mathbf{X}$ follows a multivariate Variance-Gamma process as described above.

| u | Sample Mean | Standard Error | Coeff. of Variation | Time (secs) |
|---|---|---|---|---|
| $1 \times 10^5$ | $2.41 \times 10^{-6}$ | $1.53 \times 10^{-5}$ | 6.37 | 92 |
| $2 \times 10^5$ | $1.73 \times 10^{-6}$ | $1.20 \times 10^{-5}$ | 6.93 | 92 |
| $3 \times 10^5$ | $1.22 \times 10^{-6}$ | $8.52 \times 10^{-6}$ | 6.96 | 92 |
| $4 \times 10^5$ | $8.10 \times 10^{-7}$ | $5.92 \times 10^{-6}$ | 6.81 | 92 |
| $5 \times 10^5$ | $8.51 \times 10^{-7}$ | $5.72 \times 10^{-6}$ | 6.73 | 92 |

Notice that the coefficient of variation increases slowly and even decreases as $u$ becomes large. This feature, common in efficient algorithms, shows that accurate estimates of $\alpha(u)$ for larger values of $u$ can be obtained with an affordable increment of the number of replications. The cpu time is relatively high since for each replication a root-finding algorithm is run. However, the times remain fairly constant as $u \to \infty$.

**Example 6.** (*Kou Model.*) For the Kou model we need to solve $\frac{\partial}{\partial \gamma_i}\psi(\gamma_i \mathbf{e}_i) = b$ where

$$\frac{\partial}{\partial \gamma_i}\psi(\gamma_i \mathbf{e}_i) = X_i(0) + t\Bigg[\mu_i - \frac{\Sigma_{i,i}}{2} + \gamma_i \Sigma_{i,i} + \frac{\lambda_0 (G_{i,i}\gamma_i + m_i)}{(1 - \gamma_i^2 G_{i,i}/2 - \gamma_i m_i)^2} + \frac{\lambda_i(\beta_i \gamma_i + \nu_i)}{(1 - \gamma_i^2 \beta_i/2 - \gamma_i \nu_i)^2}\Bigg].$$

Observe that the expression on the right hand side possesses vertical asymptotes and possibly more than one positive root. Remember that the equality above holds in the region of convergence of $\psi(\cdot)$ and therefore we should pick the smallest positive root. However, one must be careful to verify that the root-finding algorithm returns this root.

For the implementation of the estimator (12) for the Kou model the parameters used are as in (23) and given as follows: $\mathbf{X}(0) = (\log(70), \log(52))$, $\boldsymbol{\mu} = (0.05, 0.05)$, $\Sigma = (0.09, 0.06; 0.06, 0.25)$, $\lambda_0 = 3$, $m = (-0.5, 0.1)$, $G = (0.16, 0; 0, 0.36)$, $\lambda_1 = 0.5$, $\lambda_2 = 1.5$, $\nu_1 = -0.2$, $\nu_2 = 0.4$, $\beta_1 = 0.25$ and $\beta_2 = 0.09$. A total of $10^7$ were used to obtain the estimations.

Table 2 summarizes the results of our numerical experiments. The coefficient of variation increases very slowly as $u \to \infty$, showing that the algorithm produces accurate estimations for very small probabilities, in this case of the order of $10^6$.

TABLE 2: Statistics for the estimator of $\mathbb{P}(e^{X_1} + e^{X_2} > u)$ for the Kou model.

| u | Sample Mean | Standard Error | Coeff. of Variation | Time (secs) |
|---|---|---|---|---|
| $1 \times 10^6$ | $1.54 \times 10^{-5}$ | $4.48 \times 10^{-5}$ | 2.91 | 142 |
| $2 \times 10^6$ | $5.90 \times 10^{-6}$ | $1.77 \times 10^{-5}$ | 3.00 | 140 |
| $3 \times 10^6$ | $3.35 \times 10^{-6}$ | $1.02 \times 10^{-5}$ | 3.06 | 147 |
| $4 \times 10^6$ | $2.23 \times 10^{-6}$ | $6.89 \times 10^{-6}$ | 3.10 | 148 |
| $5 \times 10^6$ | $1.62 \times 10^{-6}$ | $5.07 \times 10^{-6}$ | 3.12 | 155 |

## Acknowledgements

We are grateful to the Referees and the Associate Editor for their handling of our paper and their very valuable comments. The first author acknowledges support from the NSF grants DMS-0902075 and CMMI-0846816.

## References

[1] ABRAMOWITZ, M. AND STEGUN, I. A. (1965). *Handbook of Mathematical Functions.* Dover, New York.

[2] ADLER, R., BLANCHET, J. AND LIU, J. C. (2010). Efficient Monte Carlo for high excursions of Gaussian random fields. arXiv:1005.0812v3.

[3] ALBRECHER, H., ASMUSSEN, S. AND KORTSCHAK, D. (2006). Tail asymptotics for the sum of two heavy-tailed dependent risks. *Extremes* 107–130.

[4] ASMUSSEN, S. (2000). *Ruin Probabilities.* World Scientific Publishing, Singapore.

[5] ASMUSSEN, S. (2003). *Applied Probability and Queues.* Springer-Verlag, New York.

[6] ASMUSSEN, S. AND BINSWANGER, K. (1997). Simulation of ruin probabilities for subexponential claims. *ASTIN Bulletin* **27,** 297–318.

[7] Asmussen, S., Binswanger, K. and Højgaard, B. (2000). Rare events simulation for heavy-tailed distributions. *Bernoulli* **6,** 303–322.

[8] Asmussen, S., Blanchet, J., Juneja, S. and Rojas-Nandayapa, L. (2009). Efficient simulation of tail probabilities of sums of correlated lognormals. *Annals of Operations Research*. To appear.

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

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

[11] Asmussen, S. and Rojas-Nandayapa, L. (2008). Asymptotics of sums of lognormal random variables with Gaussian copula. *Stat. Probab. Lett.* **78,** 2709–2714.

[12] Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London* **353,** 409–419.

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

[14] Blanchet, J., Juneja, S. and Rojas-Nandayapa, L. (2008). Efficient tail estimation for sums of correlated lognormals. In *Proceedings of the 2008 Winter Simulation Conference.* IEEE, Miami, FL., USA. pp. 607–614.

[15] Blanchet, J. and Li, C. (2011). Efficient rare-event simulation for heavy-tailed compound sums. *ACM TOMACS* **21,**. Forthcoming.

[16] Boots, N. K. and Shahabuddin, P. (2001). Simulating ruin probabilities in insurance risk processes with subexponential claims. In *Proceedings of the 2001 Winter Simulation Conference.* IEEE, Arlington, VA., USA. pp. 468–476.

[17] Chan, J. C. C. and Kroese, D. (2009). Rare-event probability estimation with Conditional Monte Carlo. *Annals of Operations Research*. To appear.

[18] Chan, J. C. C. and Kroese, D. (2010). Efficient estimation of large portfolio loss probabilities in $t$-copula models. *European Journal of Operational Research* **205,** 361–367.

[19] Cont, R. and Tankov, P. (2003). *Financial Modelling with Jump Processes.* Chapman & Hall, New York.

[20] Duffie, D. (2001). *Dynamic Asset Pricing Theory.* Princeton University Press, New Jersey.

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

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

[23] Foss, S. and Richards, A. (2010). On sums of conditionally independent subexponential random variables. *Mathematics of Operations Research* **35,** 102–119.

[24] Hartinger, J. and Kortschak, D. (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.* RESIM, Bamberg, Germany.

[25] Huang, Z. and Kou, S. (2006). First passage times and analytical solutions for options on two assets with jump risks. Technical Report.

[26] Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution* vol. 9 of *Lecture Notes in Statistics.* Springer-Verlag, New York.

[27] Juneja, S. (2007). Estimating tail probabilities of heavy-tailed distributions with asymptotically zero relative error. *QUESTA* **57,** 115–127.

[28] Juneja, S. and Shahabuddin, P. (2002). Simulating heavy-tailed processes using delayed hazard rate twisting. *ACM TOMACS* **12,** 94–118.

[29] Klöppel, S., Reda, R. and Scharchermayer, W. (2010). A rotationally invariant technique for rare event simulation. *Risk Magazine* **22,** 90–94.

[30] Kortschak, D. and Albrecher, H. (2008). Asymptotic results for the sum of dependent non-identically distributed random variables. *Methodology and Computing in Applied Probability* **11,** 279–306.

[31] Kou, S. (2002). A jump diffusion model for option pricing. *Management Science* **48,** 1086–1101.

[32] McNeil, A., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press, New Jersey.

[33] Mikosch, T. (2009). *Non-Life Insurance Mathematics: An introduction with the Poisson process.* Springer-Verlag, Berlin.

[34] Mitra, A. and Resnick, S. I. (2009). Aggregation of rapidly varying risks and asymptotic independence. *Advances in Applied Probability* **41,** 797–828.