Efficient simulation of tail probabilities of sums of correlated lognormals.

Søren Asmussen
Department of Mathematical Sciences
University of Aarhus
asmus@imf.au.dk

José Blanchet
Department of IEOR
Columbia University
jose.blanchet@columbia.edu

Sandeep Juneja
School of Technology and Computer Sciences
Tata Institute of Fundamental Research
juneja@tifr.res.in

Leonardo Rojas-Nandayapa
Department of Mathematical Sciences
University of Aarhus
leorojas@imf.au.dk

May 23, 2009

Abstract

We consider the problem of efficient estimation of tail probabilities of sums of correlated lognormals via simulation. This problem is motivated by the tail analysis of portfolios of assets driven by correlated Black-Scholes models. We propose two estimators that can be rigorously shown to be efficient as the tail probability of interest decreases to zero. The first estimator, based on importance sampling, involves a scaling of the whole covariance matrix and can be shown to be asymptotically optimal. A further study, based on the Cross-Entropy algorithm, is also performed in order to adaptively optimize the scaling parameter of the covariance. The second estimator decomposes the probability of interest in two contributions and takes advantage of the fact that large deviations for a sum of correlated lognormals are (asymptotically) caused by the largest increment. Importance sampling is then applied to each of these contributions to obtain a combined estimator with asymptotically vanishing relative error.

Keywords: Black-Scholes model, correlated lognormals, Importance sampling, Cross-Entropy method, efficiency, rare-event simulation, vanishing relative error.

1 Introduction

We consider the problem of efficient estimation of tail probabilities of sums of random variables that are correlated and possess heavy tails. As a motivating example, one could consider the problem of computing the probability of large losses or high returns on a portfolio of correlated asset prices. A very popular model in the financial literature is the so-called Black-Scholes model, in which stock prices follow a lognormal distribution which are usually considered to have significative correlations. Motivated by these types of financial risk problems, we shall concentrate on efficient tail estimation of sums of correlated lognormals. More precisely, let $\mathbf{Y} = (Y_1, Y_2, ..., Y_d)^T$ be a d-dimensional vector distributed jointly Gaussian with mean $\boldsymbol{\mu} = (\mu_1, ..., \mu_d)^T$ and covariance matrix $\mathbf{\Sigma}$ (we say that $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$). Finally, define $X_i = \exp(Y_i)$ and set $S_d = X_1 + ... + X_d$. We are interested in the efficient estimation of $\alpha(b) = \mathbb{P}(S_d > b)$ as $b \nearrow \infty$.

Recall that an *unbiased* estimator Z_b for $\alpha(b)$ is said to be *weakly efficient*, *logarithmically efficient* or *asymptotically optimal* if $\log EZ_b^2 / \log \alpha(b) \longrightarrow 2$ as $b \nearrow \infty$. Equivalently, weak efficiency can be stated in terms of the requirement that $\sup_{b \geq 0} EZ_b^2 / \alpha(b)^{2-\varepsilon} < 0$ for each $\varepsilon > 0$. Moreover, an estimator is *strongly efficient* or is said to have *bounded relative error* if $\sup_{b \geq 0} EZ_b^2 / \alpha(b)^2 < \infty$. These notions are standard in rare event simulation, see for instance Asmussen and Glynn (2007); Bucklew (2004); Juneja and Shahabuddin (2006). Finally, a notion that has been recently introduced (see Juneja, 2007) is that of *asymptotically vanishing relative error*, which goes beyond strong efficiency and requires the second moment of the estimator to achieve the best possible asymptotic performance, namely $\lim_{b \longrightarrow \infty} EZ_b^2 / \alpha(b)^2 = 1$.

Most of the literature on efficient rare-event simulation for heavy-tailed systems has focused on random walk-type models (see, for instance, Asmussen and Kroese, 2006; Juneja and Shahabuddin, 2002; Dupuis, Leder and Wang, 2006; Blanchet and Glynn, 2007; Blanchet, Glynn and Liu, 2007; Juneja, Karandikar and Shahabuddin, 2007). In contrast, we consider a rare-event simulation problem that involves the sum of dependent increment distributions. The dependence structure makes the available rare-event simulation algorithms for tails of sums of i.i.d. heavy-tailed increments difficult to apply in our current setting because they rely heavily on the i.i.d. assumption.

We mentioned before that our current setting relates to applications in finance, in the context of tail probabilities of assets driven by correlated Black-Scholes models. In this context, a popular approach that is often suggested is approximating the prices by a t-distributed model. Such approximation is motivated by means of a Taylor expansion which is often called a Delta approximation, if it involves the first derivative only or Delta-Gamma approximation, if the first and second derivatives are considered (Glasserman, 2000). The use of t-distributions is appealing in these settings in order to capture the heavy-tailed behavior which is present in the original lognormal model (which is approximated by means of the Delta-Gamma development). Efficient rare-event simulation procedures are then designed for the Delta-Gamma approximation with t-distributed factors or quadratic forms of Gaussian factors (Glasserman, Heidelberger and Shahabuddin, 1998; Glasserman, 2000). The simulation estimators that we propose and analyze here avoid the need for a Delta-Gamma approximation by working directly with the lognormal factors in an efficient way. So, we do not incur bias errors that are inherent to the use of the Delta-Gamma approximation and, at the same time, efficiency of the estimators is preserved.

Our contributions are as follows. We analyze and propose two importance sampling estimators for $\alpha(b)$. The first estimator is closely related to the use of Cross-Entropy methods for finding the best tuning for the importance distribution. Interestingly, such tuning can be related to an appropriate exponential change-of-measure, but *not* directly to the underlying Gaussian distributions, but to the radial component expressed in polar coordinates. Such change-of-measure turns out to be equivalent to scaling the covariance matrix by a factor that grows at a suitable slow speed as $b \nearrow \infty$. Since the sampler involves a simple scaling, the estimator is straightforward to implement and it can be shown to be asymptotically optimal as $b \nearrow \infty$. The second of our estimators takes advantage of the fact that the largest of the increments dominates the large deviations behavior of the sums of correlated lognormals. The strategy is to decompose the tail event of interest in two contributions, a dominant piece corresponding to the tail of the maximum and a remaining contribution. The dominant contribution is analyzed by means of a strongly efficient estimator for the maximum of multivariate Gaussians and the remaining contribution is independently handled using the importance sampling strategy utilized in the design of the first estimator. We show that our second estimator is strongly efficient and, under additional mild conditions, it even possesses *asymptotically vanishing relative error.*

The rest of the paper is organized as follows: Basic large deviations results for sums of correlated lognormals are briefly discussed in Section 2. The description and analysis of the first proposed importance sampling estimator is given in Section 3. Section 4 contains the analysis of the proposed strongly efficient estimator. Finally, numerical examples are given in Section 5, the last section.

2 Tail Asymptotics for Sums of Lognormals

We first introduce some notation. Let $\sigma_i^2 = \mathbf{\Sigma}_{i,i}$, $\sigma_{i,j} = \mathbf{\Sigma}_{i,j}$ for $i \neq j$ and $\rho_{i,j} = \sigma_{i,j}/(\sigma_i \sigma_j)$; these three notions correspond to the variance of the i-th Gaussian component and the covariance and correlation between the i-th and j-th components respectively. We reserve the use of boldface to denote matrices and vectors (which by convention will be in column form). The use of capital letter is mostly reserved for random variables and the corresponding lower-case version is used to denote specific realizations. Finally, we also use the notation $f(t) = O(g(t))$ if there exists a constant $m_1 \in (0,\infty)$ such that $|f(t)| \leq m_1 g(t)$; if, in addition, $|f(t)| \geq m_2 g(t)$ for some $m_2 \in (0,\infty)$, then $f(t) = \Theta(g(t))$. Finally, we say that $f(t) = o(g(t))$ as $t \nearrow \infty$ if $f(t)/g(t) \longrightarrow 0$ as $t \nearrow \infty$.

As indicated in the Introduction, $S_d = X_1 + \ldots + X_d$, and we also write $M_d = \max\{X_i : 1 \leq i \leq d\}$. In addition, we let

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

The parameters σ^2 and μ allow to characterize the dominant tail behavior among the X_j's or, equivalently, among the Y_j's – recall from the introduction that $X_j = \log Y_j$. In order to see this, let us recall the following well known asymptotic relation (often refered to as Mill's ratio, cf. Resnick, 1992); if $Y_i \sim N\left(\mu_i, \sigma_i^2\right)$ then as $y \nearrow \infty$

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

Note that, in particular, that approximation (1) indicates that $\mathbb{P}(X_j > b) = o(\mathbb{P}(X_i > b))$ if $\sigma_i^2 > \sigma_j^2$, or $\sigma_i^2 = \sigma_j^2$ and $\mu_i > \mu_j$.

The following assumption is useful to our analysis:

Assumption A: The correlations $\rho_{k\ell}$ are less than 1, whenever $\sigma_k^2 = \sigma_\ell^2 = \sigma^2$.

The following result will be necessary for the efficiency analysis of our estimators.

Theorem 1 (Asmussen and Rojas-Nandayapa (2008))**.** *Suppose* $0 < \gamma(b) \rightarrow \gamma_* \in (0, \infty)$ *as* $b \nearrow \infty$*. Define* $\mathbf{Y}(b) = (Y_1(b), ..., Y_d(b)) \sim N(\boldsymbol{\mu}, \gamma(b)\boldsymbol{\Sigma})$*,* $X_j(b) = \exp(Y_j(b))$ *and put* $S_d(b) = X_1(b) + ... + X_d(b)$*. Then, under Assumption A, we have*

$$\lim_{b\rightarrow\infty} \frac{\mathbb{P}(S_d(b) > b)}{\mathbb{P}(\mu + \sigma\gamma(b)N(0,1) > \log b)} = m_d,$$

where with a slight abuse of notation we write $N(0,1)$ *for the random variable as well as the distribution.*

Remark 1. Throughout the paper, in order to shorten the size of some displays in the technical development, we sometimes drop the argument b in the random variables above. So, for instance, $X_k(b)$ is simply denoted by X_k.

The previous result for $\gamma(b) = 1$ is proved in Asmussen and Rojas-Nandayapa (2008). The extension to $\gamma(b) \rightarrow \gamma_* \in (0, \infty)$, is required in our future development and follows exactly as in Asmussen and Rojas-Nandayapa (2008). We omit the details here.

Theorem 1 is an extension of the subexponential property for sums of i.i.d. lognormal random variables which states that (see Embrechts, Klüppelberg and Mikosch, 1997)

$$\mathbb{P}(S_d > b) \sim \sum_{j=1}^{d} \mathbb{P}(X_j > b) = d\mathbb{P}(X_j > b), \qquad b \nearrow \infty.$$

It follows from (1) (following the notation in Theorem 1) that if $\mathbf{Y}(b) \sim N(\boldsymbol{\mu}, \gamma(b)\boldsymbol{\Sigma})$ then

$$\sum_{j=1}^{d} \mathbb{P}(X_j(b) > b) \sim m_d\mathbb{P}(\mu + \sigma\gamma(b)N(0,1) > \log b), \qquad b \nearrow \infty.$$

In the i.i.d. case, it is straightforward to show that $\mathbb{P}(M_d > b) \sim d\,\mathbb{P}(X_j > b)$ as $b \nearrow \infty$ and therefore, we obtain that $\mathbb{P}(M_d > b|\, S_d > b) \longrightarrow 1$ as $b \nearrow \infty$. In turn, we can intuitively interpret this result by saying that sum of i.i.d. lognormals are large due to the contribution of a single large increment, namely, the maximum. Proposition 1 below provides useful intuition behind the occurrence of the event $\{S_d(b) > b\}$. In particular, it indicates that, just as in the i.i.d. case, when dealing with correlated lognormals we also have that $\mathbb{P}(M_d(b) > b|\, S_d(b) > b) \longrightarrow 1$ as $b \nearrow \infty$ and therefore, the same intuition as before, namely, that the sum is large due to the contribution of the maximum increment, remains valid.

Proposition 1. Under the assumptions of Theorem 1 we have that

$$\lim_{b\rightarrow\infty} \frac{\mathbb{P}(M_d(b) > b)}{\mathbb{P}(\mu + \sigma\gamma(b)N(0,1) > \log b)} = m_d.$$

where $M_d(b) = \max\{X_k(b) : 1 \leq k \leq d\}$.

Proof. Note that

$$
\begin{aligned}
m_d &= \lim_{b\to\infty} \frac{\mathbb{P}(S_d(b) > b)}{\mathbb{P}(\mu + \sigma\gamma(b)\, N(0,1) > \log b)} \\
&\geq \lim_{b\to\infty} \frac{\mathbb{P}(M_d(b) > b)}{\mathbb{P}(\mu + \sigma\gamma(b)\, N(0,1) > \log b)} \\
&\geq \lim_{b\to\infty} \frac{\sum_{i=1}^{d} \mathbb{P}(X_i(b) > b) - \sum_{j\neq i} \mathbb{P}(X_i(b) > b, X_j(b) > b)}{\mathbb{P}(\mu + \sigma\gamma(b)\, N(0,1) > \log b)}.
\end{aligned}
$$

We claim that the last line in the previous display is asymptotically equivalent to

$$
\lim_{b\to\infty} \frac{\sum_{i=1}^{d} \mathbb{P}(X_i(b) > b)}{\mathbb{P}(\mu + \sigma\gamma(b)\, N(0,1) > \log b)} = m_d.
$$

To see this, write

$$
\mathbb{P}(X_i(b) > b, X_j(b) > b) = \begin{cases} \mathbb{P}(X_i(b) > b \mid X_j(b) > b)\, \mathbb{P}(X_j(b) > b) \\ \\ \mathbb{P}(X_j(b) > b \mid X_i(b) > b)\, \mathbb{P}(X_i(b) > b). \end{cases}
$$

If $\mathbb{P}(X_k(b) > b) = o(\mathbb{P}(\mu + \sigma\gamma(b)\, N(0,1) > \log b))$ as $b \nearrow \infty$ for $k = i$ or $k = j$, then the claim holds immediately, so, the interesting case is obtained when the $X_i(b)$ and $X_j(b)$ are identically distributed with $\mathbb{P}(X_k(b) > b) = \mathbb{P}(\mu + \sigma\gamma(b)\, N(0,1) > \log b)$. However, it is well known (see, for instance, McNeil, Frey and Embrechts, 2005) that if Z_1 and Z_2 are jointly standard Gaussian r.v.'s then $\mathbb{P}(Z_1 > b \mid Z_2 > b) \longrightarrow 0$ as $b \nearrow \infty$ whenever $\mathbb{C}\mathrm{orr}(Z_1, Z_2) < 1$. A straightforward adaptation of this result to the case of Gaussian random variables with scaled covariance structure allows us to conclude the previous claim in this case and, in turn, the result. □

3 Asymptotically Optimal IS via Variance Scaling

A popular principle in financial risk analysis is that high variance or volatility is associated with high risk. Of course, one has to be careful when applying this principle in light of what is meant by risk. Typically, the notion of risk is associated to tail behavior and, in general, variance has little to do with tail behavior. However, as we saw in Section 2, more precisely by means of approximation (1), in the case of Gaussian random variables, the variance controls the tail behavior of the underlying factors.

Using the previous principle, a natural importance sampling strategy that one might consider for computing $\alpha(b)$ is one that induces high variances. This motivates considering as importance sampler a distribution such as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/(1-\theta))$ for some $0 < \theta < 1$; in other words, relative to the nominal (original) probability distribution, we just inflate the covariance matrix by the factor $1/(1-\theta)$. This importance sampling distribution is denoted by $\mathbb{P}_\theta(\cdot)$ and we shall use the notation $\mathbb{E}_\theta(\cdot)$ for the associated expectation operator. An interesting observation, which follows from the fact that inflating the variance is equivalent to scaling the squared of the radial component in polar coordinates, is that this importance sampling strategy corresponds to an exponential tilting applied to the squared of the radial component.

The estimator induced by this simple strategy is

$$
\begin{aligned}
Z_1(b) &= \frac{\mathbb{I}(S_d > b) \exp\left(-(\mathbf{Y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})/2\right) / \det(\boldsymbol{\Sigma})^{1/2}}{\exp\left(-(\mathbf{Y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})(1-\theta)/2\right) \det(\boldsymbol{\Sigma}/(1-\theta))^{1/2}} \\
&= \mathbb{I}(S_d > b) \frac{\exp\left(-\theta(\mathbf{Y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})/2\right)}{(1-\theta)^{d/2}}.
\end{aligned}
$$

The next lemma summarizes a useful representation for the second moment of $Z_1(b)$ under the importance sampling distribution. However, in order to state such representation we introduce another family of probability measures (in addition to the $\mathbb{P}_\theta$'s), which we shall denote by $(Q_\theta : 0 \leq \theta \leq 1)$. We use $Q_\theta(\cdot)$ to denote a probability measure under which $\mathbf{Y}$ is $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/(1+\theta))$.

Proposition 2.

$$
\mathbb{E}_\theta Z_1^2(b) = (1-\theta^2)^{-d/2} Q_\theta(S_d > b). \tag{2}
$$

Proof.

$$
\begin{aligned}
\mathbb{E}_\theta Z_1^2(b) &= \int \frac{\mathbb{I}(e^{y_1} + ... + e^{y_d} > b) \exp\left(-2\theta(\mathbf{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2\right)}{(1-\theta)^d} \\
&\quad \times \frac{\exp\left(-(1-\theta)(\mathbf{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2\right)}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma}/(1-\theta))^{1/2}} dy_1 ... dy_d \\
&= \int \mathbb{I}(e^{y_1} + ... + e^{y_d} > b) \\
&\quad \times \frac{\exp\left(-(1+\theta)(\mathbf{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2\right)}{(1-\theta)^{d/2}(1+\theta)^{d/2}(2\pi)^{d/2} \det(\boldsymbol{\Sigma}/(1+\theta))^{1/2}} dy_1 ... dy_d \\
&= (1-\theta^2)^{-d/2} Q_\theta(S_d > b).
\end{aligned}
$$

□

As an immediate consequence of the previous result we obtain that the estimator $Z_1(\theta)$ is logarithmic efficient if one chooses $\theta(b) \longrightarrow 1$ at an appropriate speed. We wish to select θ close to unity because under $Q_\theta(\cdot)$ the variances are multiplied by the factor $1/(1+\theta)$ and, to obtain logarithmic efficiency, we wish to match the rate of decay of $\alpha(b)^2$ which, again in logarithmic terms as seen by (1), is determined by the factor one half times the largest variance parameter.

Theorem 2. *Suppose that $\psi(b) := 1 - \theta(b) = o(1)$, then for $\epsilon \geq 0$*

$$
\frac{\mathbb{E}_{\theta(b)} Z_1^2(b)}{\alpha(b)^{2-\epsilon}} = \Theta\left((\log b)^{1-\epsilon} \psi(b)^{-d/2} \exp\left(-\frac{(\epsilon - \psi(b))(\log b - \mu)^2}{2\sigma^2}\right)\right). \tag{3}
$$

In particular, if $1/\psi(b) = o(e^{p(\log b)^2})$ for all $p > 0$, then $Z_1(b)$ is logarithmically efficient.

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

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

Since we have that

$$\alpha(b)^{2-\epsilon} = \Theta\left(\frac{1}{(\log b)^{2-\epsilon}}\exp\left(-\frac{(\log b - \mu)^2(2-\epsilon)}{2\sigma^2}\right)\right),$$

the result follows by noting that

$$\left(1-\theta(b)^2\right)^{-d/2} = \Theta\left((\psi(b))^{-d/2}\right),$$

and plugging in this estimate together with that of $Q_{\theta(b)}\left(S_d > b\right)$ into expression (2). $\square$

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

Proposition 3. The function $\theta(b)$ given as the unique root of the equation

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

is such that

$$\frac{1}{\log b - \mu} \le \frac{2(1-\theta(b))}{\sigma^2} \le \frac{1}{\log b - \mu - \log(d)},$$

for all b sufficiently large.

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

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

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

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

We must have that $1-\theta\left(b\right)\leq 1-\theta_{+}\left(b\right)$. However,

$$\frac{\sigma^2}{2\left(1-\theta_{+}\left(b\right)\right)}=\log b-\mu-\log\left(d\right).$$

Moreover, we also have that

$$1-\theta\left(b\right)\geq 1-\theta_{+}\left(b\right)=\frac{\sigma^2}{2\left(\log b-\mu\right)}.$$

These observations imply the statement of the proposition. □

The precise form of the algorithm for $Z_1\left(b\right)$ that we implement in Section 6 is given next.

Algorithm 1

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

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

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

3. Return

$$Z_1(b)=\mathbb{I}\left(S_d>b\right)\frac{\exp\left(-\theta\left(\mathbf{Y}-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})/2\right)}{(1-\theta)^{d/2}}.$$

The following result follows from the analysis in Theorem 2 and Proposition 3.

Proposition 4. The estimator $Z_1(b)$ given by Algorithm 1 satisfies

$$\frac{\mathbb{V}\mathrm{ar}_{\,\theta}Z_1(b)}{\alpha\left(b\right)^2}=O\left(b^{1/4}\log b^{d/2+1}\right).$$

Proof. Note that

$$\exp\left(-\frac{\left(1-\theta\left(b\right)\right)\left(\log b-\mu\right)^2}{2\sigma^2}\right)=O\left(b^{1/4}\right),$$

since $1-\theta(b)=\sigma^2\left(\log\left(b\right)-\mu\right)^{-1}/2+O\left(1\right)$ by Proposition 1. The result follows by inserting this in (3). □

Although, the estimator $Z_1\left(b\right)$ possesses two very convenient features, namely, it is very easy to implement and it is asymptotically optimal, it also has the disadvantage that the premultiplying factor in the asymptotic variance expression (3) might grow substantially involving a factor such as $O\left(\log b^{d/2+1}\right)$. So, for moderate values of b and d, the variance performance of the estimator might degrade in a significant way. To cope with this problem one can introduce additional variance reduction techniques, such as stratified sampling or conditional Monte Carlo. Preliminary work on this direction is reported in Blanchet, Juneja and Rojas-Nandayapa (2008). Another alternative that takes advantage of the intuitive interpretation given by Prposition 1 and that achieves bounded relative error will be studied later, but first, we provide another refinement to the estimator $Z_1\left(b\right)$ using Cross-Entropy based ideas.

3.1 Cross-Entropy Implementation of IS via Variance Scaling

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

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

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

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

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

Here we provide directly the expression for the parameters omitting the details of the calculation. For more details on the Cross-Entropy method we refer the reader to Rubinstein and Kroese (2004) or (pp. 198, Asmussen and Glynn, 2007).

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

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

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

$$\widetilde{\boldsymbol{\mu}}_k := \frac{\sum_{i=1}^r w_{i,k}\, \mathbf{y}_{i,k}}{\sum_{i=1}^r w_{i,k}}, \qquad \frac{1}{1-\theta_k} := \frac{\sum_{i=1}^r w_{i,k}\, (\mathbf{y}_{i,k} - \widetilde{\boldsymbol{\mu}}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_{i,k} - \widetilde{\boldsymbol{\mu}}_k)}{r \sum_{i=1}^r w_{i,k}}, \tag{4}$$

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

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

It is an easy calculus exercise to verify that these expressions satisfy the conditions of the Cross-Entropy method. One could try to choose a larger family of importance sampling distributions to provide better estimates, however, the expressions can quickly become complicated and more difficult to implement. We perform numerical experiments (the output is given in Section 6) and note that the algorithm converges in a few iterations suggesting that our initial distribution is not that far from the optimal distribution within the new family. The precise description of the algorithm is given below.

Cross-Entropy Sampling Algorithm.

1. Let $k = 1$ and $\widetilde{\boldsymbol{\mu}}_0 := \boldsymbol{\mu}$. Define $\theta_0 := \theta(b)$ as the solution of
$$e^{\mu_1+\sigma_1^2\theta(b)/2} + \ldots + e^{\mu_d+\sigma_d^2\theta(b)/2} = b.$$

2. Simulate a sequence of random vectors r i.i.d. r.v.'s $(\mathbf{Y}_{i,k} : 1 \le i \le r)$, $\mathbf{Y}_{i,k} \sim N(\widetilde{\boldsymbol{\mu}}_k, \boldsymbol{\Sigma}/(1-\theta_k))$ and calculate $\widetilde{\boldsymbol{\mu}}_{k+1}$ and θ_{k+1} as given in (4). If the new parameters satisfy a convergence criteria go to 3 (see our comments below for a convergence criteria that we used in our numerical examples). Else make $k := k + 1$ and repeat 2.

3. Simulate $\mathbf{Y} \sim N(\widetilde{\boldsymbol{\mu}}_{k+1}, \boldsymbol{\Sigma}/(1-\theta_{k+1}))$, let S_d the sum of the elements of $\mathbf{Y}$ and return
$$\widetilde{Z}_1(b) := (1-\theta_{k+1})^{-d/2} \frac{\exp\left(-(\mathbf{Y}-\widetilde{\boldsymbol{\mu}}_{k+1})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\widetilde{\boldsymbol{\mu}}_{k+1})\right)}{\exp\left(-(\mathbf{Y}-\widetilde{\boldsymbol{\mu}}_{k})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\widetilde{\boldsymbol{\mu}}_{k})\right)} \mathbb{I}(S_d > b).$$

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

4 Vanishing Relative Error IS

In our discussion leading to Proposition 1, we heuristically argued that large values of S_d happen due to the contribution of a single large jump (the maximum). On the other hand, in the previous section, we constructed a weakly efficient estimator using an importance sampler based on the fact that, roughly speaking (i.e., in the logarithmic sense) and according to (1) and Theorem 1, the variances dictate the tail behavior S_d. The idea in this section is to combine these two intuitive observations to produce a strongly efficient importance sampling estimator. First, note that
$$\alpha(b) = \alpha_1(b) + \alpha_2(b),$$
where
$$\alpha_1(b) = \mathbb{P}\left(\max_{1\le i\le d} X_i > b\right),$$
$$\alpha_2(b) = \mathbb{P}\left(S_d > b, \max_{1\le i\le d} X_i \le b\right).$$

In view of Theorem 1 we must have that $\alpha_2(b) = o(\alpha_1(b))$ as $b \nearrow \infty$, so the most important contribution comes from the term $\alpha_1(b)$. We shall refer to $\alpha_2(b)$ as the "residual probability".

The strategy is to design independent and unbiased estimators, say $Z_{2,1}(b)$ and $Z_{2,2}(b)$, for the terms $\alpha_1(b)$ and $\alpha_2(b)$ respectively. This idea has been exploited previously in the literature, see Juneja (2007), in the context of i.i.d. increment distributions. The gain comes if $Z_{2,1}(b)$ is strongly efficient for $\alpha_1(b)$ even if $Z_{2,2}(b)$ has a coefficient of variation of order

$O\left(\alpha\left(b\right)/\alpha_2\left(b\right)\right)$ as $b\nearrow\infty$. In other words, $Z_{2,2}\left(b\right)$ may not be strongly efficient for $\alpha_2\left(b\right)$, but its coefficient of variation could grow slowly enough so that the combined estimator $Z_2\left(b\right)=Z_{2,1}\left(b\right)+Z_{2,2}\left(b\right)$ for $\alpha\left(b\right)$ is strongly efficient.

For $Z_{2,2}\left(b\right)$ we propose to use (recall the notation introduced in Section 3) $\mathbb{P}_\theta$ as our importance sampling distribution (i.e. $\mathbf{Y}$ has distribution $N\left(\boldsymbol{\mu},,\boldsymbol{\Sigma}/(1-\theta)\right))$ with $\theta=\theta\left(b\right)=1-\log\left(b\right)^{-2}$. The corresponding estimator takes the form

$$Z_{2,2}\left(b\right)=I\left(S_d>b,\max_{1\le i\le d}X_i\le b\right)\frac{\exp\left(-\theta\left(b\right)(\mathbf{Y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})/2\right)}{(1-\theta\left(b\right))^{d/2}}.$$

The reason for using $\mathbb{P}_\theta$ as importance sampler is that for estimating $\alpha_2\left(b\right)$ one must induce the underlying rare event $\{S_d>b,\max_{1\le i\le d}X_i\le b\}$ by means of more than one large component (which might be achieved by inflating the variances), as opposed to inducing a single large jump as suggested by Corollary 1.

Just as in Section 3, we conclude that

$$E_{\theta(b)}Z_{2,2}\left(b\right)^2=\left(1-\theta\left(b\right)^2\right)^{-d/2}Q_{\theta(b)}\left(S_d>b,\max_{1\le i\le d}X_i\le b\right). \tag{5}$$

The following result provides the necessarily elements to analyze the $E_{\theta(b)}Z_{2,2}\left(b\right)^2$.

Proposition 5. Under Assumption A, and for $1-\theta\left(b\right)=\Theta\left(\log\left(b\right)^{-p}\right)$ for some $p>0$,

$$\frac{E_{\theta(b)}Z_{2,2}\left(b\right)^2}{\alpha_1\left(b\right)^2}\to 0$$

as $b\nearrow\infty$.

Proof. Recall (5) and note that

$$Q_{\theta(b)}\left(S_d>b,\max_{1\le i\le d}X_i\le b\right)=\sum_{k=1}^{d}Q_{\theta(b)}\left(S_d>b,X_k=\max_{1\le i\le d}X_i\le b\right).$$

Moreover, define

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

and consider the following decomposition (which is valid for every $\beta\in(0,1)$)

$$\begin{aligned}
&\sum_{k=1}^{d}Q_{\theta(b)}\left(S_d>b,\ X_k=\max_{1\le i\le d}X_i<b\right)\\
&\quad=\sum_{k=1}^{d}Q_{\theta(b)}\left(S_d>b,\ X_k=\max_{1\le i\le d}X_i<b,\ S_{d,-k}>b^\beta\right)\\
&\qquad\qquad+Q_{\theta(b)}\left(S_d>b,\ X_k=\max_{1\le i\le d}X_i<b,\ S_{d,-k}<b^\beta\right)\\
&\quad\le\sum_{k=1}^{d}Q_{\theta(b)}\left(S_{d,-k}(b)>b^\beta,X_k(b)>b^\beta/d\right)+Q_{\theta(b)}\left(b-b^\beta<X_k(b)<b\right).
\end{aligned} \tag{6}$$

The proof of Proposition 5 follows as an immediate consequence of Lemmas 1 and 2 (on pages 12 and 14, respectively) combined with (5) and the fact that

$$\left(1-\theta\left(b\right)^2\right)^{-d/2}=\Theta\left(\log\left(b\right)^{pd}\right).$$

$\square$

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

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

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

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

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

The assertion follows by using Mill's ratio to prove that the last term is dominated by the tail of $\alpha^2(b)$ even after pre-multiplying by a power term $b^{\gamma}=\mathrm{e}^{\gamma\log b}$.

Case 2. If $\sigma_k=\sigma$ define $\eta=\max\{\sigma_{\ell,k}/\sigma^2:\ell\neq k\}$. By Assumption A, we have that $\rho_{\ell,k}<1$ whenever $\sigma_\ell^2=\sigma_k^2=\sigma^2$. Therefore,

$$1>\max_{\ell\neq k}\{\rho_{\ell,k}\}=\max_{\ell\neq k}\left\{\frac{\sigma_{\ell,k}}{\sigma_\ell\sigma_k}\right\}.$$

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

$$\begin{aligned}&Q_{\theta(b)}\left(S_{d,-k}>b^{\beta_k},X_k>b^{\beta_k}/d\right)\\&\quad\le Q_{\theta(b)}\left(S_{d,-k}(b)>b^{\beta_k},b^{\beta_k}/d<X_k<b^{1/\beta_k}\right)+Q_{\theta(b)}\left(S_{d,-k}>b^{\beta_k},b^{1/\beta_k}<X_k\right)\\&\quad\le Q_{\theta(b)}\left(S_{d,-k}(b)>b^{\beta_k},b^{\beta_k}/d<X_k<b^{1/\beta_k}\right)+Q_{\theta(b)}\left(b^{1/\beta_k}<X_k\right).\end{aligned}\tag{7}$$

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

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

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

$$\begin{aligned}
Q_{\theta(b)}&(S_{d,-k} > b^{\beta_k}, b^{1/\beta_k} > X_k > b^{\beta_k}/d) \\
&= E^{Q_{\theta(b)}}\left(Q'_{\theta(b),Y_k}(S_{d,-k} > b^{\beta_k})\,;\, \beta_k \log b < Y_k < \log b/\beta_k\right) \\
&\leq E^{Q_{\theta(b)}}\left(Q'_{\theta(b),0}(S_{d,-k}\, e^{\eta Y_k} > b^{\beta_k})\,;\, \beta_k \log b < Y_k < \log b/\beta_k\right). \qquad (8)
\end{aligned}$$

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

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

The previous two factors have lognormal tails due to Theorem 1. In fact, since the covariances of the Gaussian conditional random variables are never larger than the unconditional ones we obtain the following relation

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

and in turn we have

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

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

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

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

In the case where $\eta > 0$, the expression (8) can be bounded by

$$Q'_{\theta(b),0}\left(S_{d,-k} > b^{\beta_k - \eta/\beta_k}\right) Q_{\theta(b)}\left(X_k > b^{\beta_k}/d\right).$$

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

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

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

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

□

Lemma 2.
$$\frac{Q_{\theta(b)}(b-b^{\beta}<X_k<b)}{\alpha(b)^2}=O\left(\frac{(\log b)^2}{b^{1-\beta}}\right),\qquad k=1,\ldots,d,$$
for any $0<\beta<1$.

Proof. Take

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

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

$$\Theta\left(\frac{\log b}{b^{1-\beta}\sigma^2}\right).$$

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

Finally, we turn our attention to $Z_{2,1}(b)$, which involves computing

$$\alpha_1(b)=\mathbb{P}(\max_{1\le j\le d}Y_j>\log(b)).$$

We shall use $f_j(y_j)$ to denote the marginal density of Y_j evaluated at $y_j\in\mathbb{R}$ and y_{-j} to denote the vector $(y_1,..,y_{j-1},y_{j+1},...,y_d)$. The expression $f(y_{-j}|\,y_j)$ is used to denote the conditional density of $\mathbf{Y}_{-j}=(Y_1,..,Y_{j-1},Y_{j+1},...,Y_d)^T$ evaluated at y_{-j} given $Y_j=y_j$. The density of the vector $\mathbf{Y}$ evaluated at y is denoted by $f(y)$. Note that for all j we have that $f(y)=f_j(y_j)f(y_{-j}|\,y_j)$. We consider as importance sampling density $g(\cdot)$ defined via

$$g(y)=\sum_{j=1}^{d}p_j(b)f_j(y_j)f(y_{-j}|\,y_j)\frac{I(y_j>\log(b))}{\mathbb{P}(Y_j>\log(b))},\tag{9}$$

where

$$p_j(b)=\mathbb{P}(Y_j>\log(b))/(\sum_{i=1}^{d}\mathbb{P}(Y_i>\log(b))).$$

We shall use the notation $\mathbb{V}\mathrm{ar}_g(\cdot)$ to denote the variance operator under the probability measure induced by $g(\cdot)$.

In other words, we first select the j-th index with probability proportional to $\mathbb{P}(Y_j>\log(b))$. Then, given that j^* has been selected we sample Y_{j^*} given that $Y_{j^*}>\log(b)$. Finally, we sample the rest of the components under the nominal distribution given that $Y_{j^*}=y_{j^*}$ (i.e. we use the law $f(\cdot|\,y_{j^*})$). The corresponding estimator is

$$Z_{2,1}(b)=\frac{f(\mathbf{Y})}{g(\mathbf{Y})}=\frac{\sum_{i=1}^{d}\mathbb{P}(Y_i>\log(b))}{\sum_{j=1}^{d}I(Y_j>\log(b))}\le\sum_{i=1}^{d}\mathbb{P}(Y_i>\log(b)).$$

This sampler is proposed and studied in Adler, Blanchet and Liu (2008). It follows immediately from Theorem 1 and Proposition 1 that under Assumption A, the coefficient of variation of estimator $Z_{2,1}(b)$ converges to zero as $b \nearrow \infty$. We record this property in the following Lemma.

Lemma 3. *Under Assumption A, the estimator $Z_{2,1}(b)$ generated under the density $g(\cdot)$ in (9) possesses asymptotically negligible coefficient of variation.*

Combining Propositions 5 and 3 we arrive at the following result summarizing the performance of the estimator $Z_2(b) = Z_{2,1}(b) + Z_{2,2}(b)$.

Theorem 1. Under Assumption A, for $\psi(b) := 1 - \theta(b) = \Theta\left(\log(b)^{-p}\right)$ for some $p > 0$, the unbiased estimator $Z_2(b)$ has *bounded relative error* in the sense that

$$\sup_{b\geq 0} \frac{\mathbb{V}\text{ar}\, Z_2(b)}{\alpha(b)^2} = \sup_{b\geq 0} \left(\frac{\mathbb{V}\text{ar}\, Z_{2,1}(b)}{\alpha(b)^2} + \frac{\mathbb{V}\text{ar}\, Z_{2,2}(b)}{\alpha(b)^2} \right) < \infty.$$

Moreover, $Z_2(b)$ has *vanishing relative error* in the sense that

$$\frac{\mathbb{V}\text{ar}\, Z_2(b)}{\alpha(b)^2} \longrightarrow 0$$

as $b \nearrow \infty$.

4.1 Cross-Entropy Implementation of the Residual Probability

The Cross-Entropy method can be implemented to improve the estimator $Z_{2,2(b)}$ of the residual probability $\alpha_2(b)$ in an entirely similar fashion as in Section 3.1. The proposal is to keep $\mathbf{\Sigma}$ fixed, and consider the parametric multivariate distribution family

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

The parameters for the k-th iteration of the algorithm are identical to those given in equation (4) but the weights are given instead by

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

5 Numerical Examples

We implemented the proposed estimators in three examples corresponding to low, medium and high correlations. In particular, we use 10 lognormal random variables with parameters $\mu_i = i - 10$ and $\sigma_i^2 = i$ with common correlation parameter as indicated: Example 1 assumes that the Gaussian factors involved are i.i.d.; in Example 2 we use a common correlation parameter equal to 0.4 and Example 3 involves a common correlation parameter equal to 0.9. The number of replications was $r = 100,000$.

In the construction of the tables we use the following abbreviations: **IS** denotes the importance sampling strategy based on variance scaling discussed in Section 3, **CE** corresponds to the Cross-Entropy method also discussed in 3.1, **ISVE** and **CEVE** relates to the importance sampling and Cross-Entropy strategies with asymptotically vanishing error described in Section 4. On the other hand, in the columns we considered the value of the *Estimation* using $r = 500,000$, the empirical *Standard Error* of the estimator (to obtain the standard error of the estimation we simply divide it by $\sqrt{r}$), the *Variation Coeff.* which denotes the empirical coefficient of variation of the estimator (i.e. the *Standard Error* divided by the empirical mean), the *CPU Time* which is the time consumed in our system to generate the *Estimation* and a measure of *Efficiency* of the estimator, say $\widehat{z}$, given as

$$\mathrm{Eff}(\widehat{z}) = \frac{s_{\widehat{z}}^2 \cdot \mathrm{Time}_{\widehat{z}}}{s_{CMC}^2 \cdot \mathrm{Time}_{CMC}}.$$

as defined in Hammersley and Handscomb (1964). The smaller the efficiency measure the better the proposed estimator.

We compare our results against Crude Monte Carlo **CMC** and a multivariate version of the algorithm proposed in Asmussen and Kroese (2006) which was empirically studied in Asmussen and Rojas-Nandayapa (2006); Rojas-Nandayapa (2008) and will be referred as AK estimator. In particular, it is proved that the **AK** estimator has *asymptotically vanishing relative error* in the i.i.d. case. However, it is not the case when any two random variables are positively correlated.

Based on the numerical experiments below and supported by the theoretical observation indicated in the previous paragraph, in a regime with low enough correlations the estimator **AK** maybe favored over the proposed algorithms in this paper. However, under medium and high correlations the proposed algorithms outperform the AK estimator. The **CEVE** and **ISVE** errors are favored over the **CE** and **IS** estimators (which supports our theoretical findings given the vanishing relative error property). In all cases, the Cross-Entropy implementation produces a significative empirical improvement over the importance sampling estimators. This occurs because there is some parameter tuning for the variance scaling contribution. The order of the scaling as a function of the tail parameter is dictated by our theoretical development and Cross-Entropy is an effective way of fine tuning such variance scaling in the prelimit.

It is interesting to note that **ISVE** algorithm tends to exhibit large relative variance in the case of high correlations. We note the same for **IS**. This suggests that the variance scaling component for **ISVE** (which is common to both **IS** and **ISVE**) contributes to the variance of the estimator substantially. Consequently, the estimators might be more sensitive to fine tuning of the variance scaling parameter in high correlation environments, which is why Cross-Entropy is particularly useful for parameter tuning in these cases.

Example 1. *Independent Case.*

Method	Estimation	Standard Error	Variation Coeff.	Time	Efficiency
CMC	1.064000e-03	3.260168e-02	3.064068e 01	1.5070	1.000000e 00
AK	1.024016e-03	1.903071e-05	1.858438e-02	23.1470	5.233734e-06
IS	1.046254e-03	1.704238e-02	1.628894e 01	3.6140	6.553232e-01
CE	9.955858e-04	8.486089e-03	8.523714e 00	3.6720	1.650914e-01
ISVE	1.023144e-03	2.628214e-04	2.568762e-01	4.8450	2.089402e-04
CEVE	1.023760e-03	1.311825e-04	1.281379e-01	4.9330	5.299930e-05

Table 1: $\mathbb{P}(S_{10} > 20,000)$.

Method	Estimation	Standard Error	Variation Coeff.	Time	Efficiency
CMC	4.200000e-04	2.048961e-02	4.878480e 01	1.4920	1.000000e 00
AK	4.631055e-04	6.510130e-06	1.405755e-02	23.3420	1.579359e-06
IS	4.442258e-04	9.230495e-03	2.077883e 01	3.6140	4.915883e-01
CE	4.649729e-04	5.285332e-03	1.136696e 01	3.6720	1.637610e-01
ISVE	4.632881e-04	2.973776e-04	6.418848e-01	4.8440	6.838867e-04
CEVE	4.630269e-04	3.982836e-05	8.601737e-02	4.9160	1.244973e-05

Table 2: $\mathbb{P}(S_{10} > 40,000)$.

Method	Estimation	Standard Error	Variation Coeff.	Time	Efficiency
CMC	1.600000e-05	3.999971e-03	2.499982e 02	1.6530	1.000000e 00
AK	1.795086e-05	4.506600e-08	2.510519e-03	23.9490	1.839071e-09
IS	1.829972e-05	8.855762e-04	4.839287e 01	3.8230	1.133625e-01
CE	1.756953e-05	2.458142e-04	1.399093e 01	3.8540	8.805197e-03
ISVE	1.794993e-05	1.062404e-06	5.918708e-02	5.0710	2.164146e-07
CEVE	1.795066e-05	8.006665e-07	4.460373e-02	5.2110	1.263099e-07

Table 3: $\mathbb{P}(S_{10} > 500,000)$.

Example 2. *Medium Correlations.*

Method	Estimation	Standard Error	Variation Coeff.	Time	Efficiency
CMC	1.044000e-03	3.229415e-02	3.093309e 01	1.5430	1.000000e 00
AK	1.054200e-03	4.374403e-03	4.149497e 00	23.6110	2.807620e-01
IS	1.042750e-03	1.698920e-02	1.629268e 01	3.6780	6.596967e-01
CE	1.049872e-03	6.003377e-03	5.718194e 00	3.6790	8.239623e-02
ISVE	1.050490e-03	2.518548e-03	2.397498e 00	27.4470	1.081886e-01
CEVE	1.048263e-03	8.882581e-04	8.473617e-01	27.4470	1.345735e-02

Table 4: $\mathbb{P}(S_{10} > 25,000)$.

Method	Estimation	Standard Error	Variation Coeff.	Time	Efficiency
CMC	5.180000e-04	2.275374e-02	4.392614e 01	1.5580	1.000000e 00
AK	4.752473e-04	2.438196e-03	5.130373e 00	23.7070	1.747193e-01
IS	4.640689e-04	9.538512e-03	2.055408e 01	3.6940	4.166631e-01
CE	4.783542e-04	3.488472e-03	7.292655e 00	3.6940	5.573070e-02
ISVE	4.723230e-04	1.223158e-03	2.589665e 00	27.1670	5.038884e-02
CEVE	4.725120e-04	3.596767e-04	7.612012e-01	27.1980	4.362030e-03

Table 5: $\mathbb{P}(S_{10} > 50,000)$.

Method	Estimation	Standard Error	Variation Coeff.	Time	Efficiency
CMC	2.000000e-05	4.472095e-03	2.236047e 02	1.6210	1.000000e 00
AK	1.782992e-05	2.230840e-04	1.251177e 01	24.0350	3.689571e-02
IS	1.810141e-05	8.342138e-04	4.608556e 01	3.8030	8.163492e-02
CE	1.841197e-05	2.016085e-04	1.094986e 01	3.8350	4.808152e-03
ISVE	1.806308e-05	2.984878e-05	1.652474e 00	27.2590	7.491312e-04
CEVE	1.812194e-05	2.191502e-05	1.209308e 00	27.1510	4.022209e-04

Table 6: $\mathbb{P}(S_{10} > 500,000)$.

Example 3. *High Correlations.*

Method	Estimation	Standard Error	Variation Coeff.	Time	Efficiency
CMC	1.126000e-03	3.353705e-02	2.978424e 01	1.5430	1.000000e 00
AK	1.128070e-03	2.224404e-02	1.971866e 01	23.7830	6.780760e 00
IS	1.110579e-03	1.858937e-02	1.673845e 01	3.8030	7.572512e-01
CE	1.133637e-03	5.246348e-03	4.627888e 00	3.7410	5.933159e-02
ISVE	1.136932e-03	9.707067e-03	8.537945e 00	27.6914	1.515592e 01
CEVE	1.127411e-03	2.666377e-03	2.365043e 00	27.8200	1.139683e-01

Table 7: $\mathbb{P}(S_{10} > 20,000)$.

Method	Estimation	Standard Error	Variation Coeff.	Time	Efficiency
CMC	5.460000e-04	2.336028e-02	4.278440e 01	1.5590	1.000000e 00
AK	5.259898e-04	1.452258e-02	2.761001e 01	23.7390	5.885021e 00
IS	4.949535e-04	1.014978e-02	2.050653e 01	3.8810	4.699530e-01
CE	5.173048e-04	2.776677e-03	5.367584e 00	3.7250	3.375783e-02
ISVE	5.193674e-04	5.258282e-03	1.012439e 01	28.7710	9.350618e-01
CEVE	5.187105e-04	1.428647e-03	2.754228e 00	28.5850	6.857815e-02

Table 8: $\mathbb{P}(S_{10} > 40,000)$.

Method	Estimation	Standard Error	Variation Coeff.	Time	Efficiency
CMC	1.800000e-05	4.242606e-03	2.357003e 02	1.5590	1.000000e 00
AK	1.911332e-05	2.074465e-03	1.085350e 02	23.6430	3.625795e 00
IS	2.020919e-05	9.307957e-04	4.605803e 01	3.7560	1.159638e-01
CE	2.087080e-05	2.012609e-04	9.643181e 00	3.8180	5.511162e-03
ISVE	2.110596e-05	4.190455e-04	1.985436e 01	28.2720	1.769161e-01
CEVE	2.074119e-05	9.332348e-05	4.499427e 00	28.1800	8.746035e-03

Table 9: $\mathbb{P}(S_{10} > 500,000)$.

References

Adler, R., Blanchet, J., and Liu, J. C. (2008) Fast simulation of Gaussian random fields. *Proceedings of the Winter Simulation Conference. Mason, S.J., Hill, R.R., Mönch, L., Rose, O., Jefferson, T. and Fowler, J.W. eds.*, 328-336.

Adler, R.J., Feldman, R. & Taqqu, M.S. eds. (1998) *A User's Guide to Heavy Tails.* (Birkhäuser).

Albrecher, H., Asmussen, S. & Kortschak, D. (2005) Tail asymptotics for the sum of dependent heavy-tailed risks. *Extremes* **9**, 107–130.

Alink, S., Löwe, M. and Wüthrich, M. V. (2005) Diversification of aggregate dependent risks. *Insurance: Mathematics and Economics* **35**, 77-95.

Asmussen, S. (2000) *Ruin Probabilities.* World Scientific.

Asmussen, S. (2003) *Applied Probability and Queues* (2nd ed.). Springer-Verlag.

Asmussen, S. & Glynn, P.W. (2007) *Stochastic Simulation. Algorithms and Analysis.* Springer-Verlag.

Asmussen, S. & Kroese, D.P. (2006) Improved algorithms for rare event simulation with heavy tails. *Advances in Applied Probability* **38**, 545-558.

Asmussen, S. & Rojas-Nandayapa, L. (2006) Sums of dependent lognormal random variables: Asymptotics and simulation. Unpublished manuscript; available from `www.thiele.au.dk`

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

Blanchet, J., Glynn, P., and Liu, J. (2007) Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue. *Queueing Systems* **57**, 99-113.

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

Blanchet, J., Juneja, S., & Rojas-Nandayapa, L. (2008) Efficient Tail Estimation for Sums of Correlated Lognormals. *Proceedings of the Winter Simulation Conference. Mason, S.J., Hill, R.R., Mönch, L., Rose, O., Jefferson, T. and Fowler, J.W. eds.*, 607-614.

Bucklew, J. (2004) *Introduction to Rare-event Simulation.* Springer-Verlag, New York.

Cossette, H., Denuit, M. and Marceau, É. (2002) Distributional bounds for functions of dependent risks. *Schweiz. Aktuarver. Mitt.*, 45–65.

Cruz, M.G. (2002) *Modeling, Measuring and Hedging Operational Risk.* Wiley.

Denuit, M., Genest, C. and Marceau, É. (1999). Stochastic bounds on sums of dependent risks. *Insurance Math. Econom.* **25**, 85–104.

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

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997) *Modeling Extremal Events for Finance and Insurance*. Springer–Verlag.

Embrechts, P. & Puccetti, G. (2005) Bounds for functions of dependent risks. *Finance and Stochastics* **10**, 341–352.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications* **II** (2nd ed.). Wiley.

Frachot, A., Moudoulaud, O. & Roncalli, T. (2004) Loss distribution approach in practice. *The Basel Handbook: A Guide for Financial Practitioners. M. Ong ed.* Risk Books.

Gentz, A. (1994) Comparison of Methods for the Computation of Multivariate Normal Probabilities. *Computing Science and Statistics* **25**, 400–405.

Glasserman, P. (2000) *Monte Carlo Methods in Finance.* Springer, New York.

Glasserman, P., Heidelberger, P., and Shahabuddin, P. (1998) Gaussian Importance Sampling and Stratification: Computational Issues. *Proceedings of the 1998 Winter Simulation Conference. D. J. Meideros, E. F. Watson, J. S. Carson and M. S. Manivannan, eds.*, 685–693.

Hammersley, J. M., and Handscomb, D. C. (1964) *Monte Carlo Methods*. Chapman & Hall.

Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Chapman & Hall.

Juneja, S. (2007) Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error. *Queueing Systems* **57**, 115–127.

Juneja, S., Karandikar, R. L., and Shahabuddin, P. (2007) Asymptotics and fast simulation for tail probabilities of maximum of sums of few random variables. *ACM TOMACS* **17**, 2, Article 7.

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

Juneja, S. and Shahabuddin, P. (2006) Rare event simulation techniques: An introduction and recent advances. *In Simulation: Handbooks in Operations Research and Management Science. S. G. Henderson and B. L. Nelson eds.*, Elsevier. Amsterdam, The Netherlands, 291-350.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994) *Continuous Univariate Distributions* **I** (2nd ed.). Wiley.

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

Resnick, S. (1992) *Adventures in Stochastic Processes.* Birkhaauser, Boston.

Rojas Nandayapa, L. (2008) *Risk Probabilities: Asymptotics and Simulation*. PhD Thesis. Aarhus University, Denmark.

Rubinstein, R.Y., and Kroese. D.P. (2004) *The Cross-Entropy Method*. Springer, New York.

Somerville, P.N. (1998) Numerical Computation of Multivariate Normal and Multivariate-t Probabilities Over Convex Regions. *Journal of Computational and Graphical Statistics* **7**, 529–544.