Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

ACCURATE COMPUTATION OF THE RIGHT TAIL OF THE SUM OF DEPENDENT LOG-NORMAL VARIATES

Zdravko Botev

School of Mathematics and Statistics
The University of New South Wales
Sydney, NSW 2052, AUSTRALIA

Pierre L'Ecuyer

DIRO, Université de Montreal
C.P. 6128, Succ. Centre-Ville
Montréal (Québec), H3C 3J7, CANADA

ABSTRACT

We study the problem of the Monte Carlo estimation of the right tail of the distribution of the sum of correlated log-normal random variables. While a number of theoretically efficient estimators have been proposed for this setting, using a few numerical examples we illustrate that these published proposals may not always be useful in practical simulations. In other words, we show that the established theoretical efficiency of these estimators does not necessarily convert into Monte Carlo estimators with low variance. As a remedy to this defect, we propose a new estimator for this setting. We demonstrate that, not only is our novel estimator theoretically efficient, but, more importantly, its practical performance is significantly better than that of its competitors.

1 INTRODUCTION

Consider the following setting. Let $\Sigma \in \mathbb{R}^{d\times d}$ be a positive definite covariance matrix. Let $\boldsymbol{Y} \sim \mathsf{N}(\boldsymbol{\nu}, \Sigma)$ be a normally distributed[1] vector in $\mathbb{R}^d$ with mean $\boldsymbol{\nu}$ under probability measure $\mathbb{P}$. Then, we wish to estimate accurately

$$\ell(\gamma) = \mathbb{P}(\exp(Y_1) + \cdots + \exp(Y_d) \geq \gamma), \tag{1}$$

where γ may be arbitrarily large, making ℓ a small *rare-event* probability.

Such a setting arises, for example, in computing the probability of a large loss from a portfolio with asset prices, say $\boldsymbol{X} = \exp(\boldsymbol{Y})$, driven by the Black-Scholes geometric Brownian motion model (see, for example, (Kroese et al. 2011) and the references therein).

One of the challenges in estimating ℓ pertains to the heavy-tailed behavior of $\ell(\gamma)$ as $\gamma \uparrow \infty$, namely[2]:

$$\ell(\gamma) \simeq \ell_{\text{asymp}} \stackrel{\text{def}}{=} \sum_{k=1}^{d} \mathbb{P}(Y_i \geq \ln \gamma) = \sum_{k=1}^{d} \overline{\Phi}((\ln \gamma - \nu_k)/\sigma_k), \tag{2}$$

where $\sigma_k^2 = \Sigma_{k,k}$, and $\overline{\Phi}(\cdot) = 1 - \Phi(\cdot)$ denotes the tail of the standard normal distribution, see (Asmussen and Rojas-Nandayapa 2008). There is significant literature on efficient heavy-tailed probability estimation of sums of independent increments, see, for example, (Asmussen and Kroese 2006, Asmussen and Kortschak 2015, Nguyen and Robert 2014, Ortobelli et al. 2016), and a few proposals dealing with sums of dependent heavy-tailed increments (Asmussen et al. 2011).

The purpose of this article is to show that the existing proposals for estimating (1) can be unreliable in some simple examples of applied interest. More precisely, while the existing estimators work satisfactory

[1] We denote the multivariate normal density as $\phi_\Sigma(\cdot - \boldsymbol{\nu})$.

[2] The notation $f(x) \simeq g(x)$ as $x \to a$ stands for $\lim_{x\to a} f(x)/g(x) = 1$. Similarly, we define $f(x) = \mathcal{O}(g(x)) \Leftrightarrow \lim_{x\to a} |f(x)/g(x)| < \text{const.} < \infty$; $f(x) = o(g(x)) \Leftrightarrow \lim_{x\to a} f(x)/g(x) = 0$; also, $f(x) = \Theta(g(x)) \Leftrightarrow f(x) = \mathcal{O}(g(x))$ and $g(x) = \mathcal{O}(f(x))$.

when Σ in (1) is diagonal (that is, when the log-normal factors $\boldsymbol{X}$ are independent), these estimators exhibit very high variance in cases in which Σ induces positive correlation amongst the log-normal factors. Unfortunately, dependence structures which induce strong positive correlation are precisely the cases of practical interest — financial and insurance risk is typically underestimated with models that fail to account for the co-dependence of market disturbances, which in turn may cause a cascading mode of defaults or downside movements.

As a remedy to this defect of existing estimators, we propose a novel exponentially tilted estimator for (1), which is logarithmically efficient, but, more importantly, enjoys superior practical performance to alternative estimators. Interestingly, in the iid case, our estimator looks similar to estimator (1.3) proposed in (Asmussen and Kroese 2006), which was found by the authors to be second best to the celebrated *Asmussen-Kroese* conditional estimator.

The rest of the paper is organized as follows. In Section 2 we review the *importance sampling vanishing error* (ISVE) estimator proposed in (Asmussen et al. 2011), and show numerically how in some cases it may yield highly inaccurate and biased estimates. We give some intuitive explanations for the poor performance of the estimator. Next, in Section 3 we describe our novel estimator and its theoretical properties. This is followed by a numerical example, demonstrating the superior performance of the estimator, and, finally, we draw some conclusions.

2 Two Theoretically Efficient Estimators

The first estimator proposed in (Asmussen et al. 2011) is the *variance boosted* estimator defined as follows.

Let $\mathbb{P}_\theta$ be a probability measure under which $\boldsymbol{Y} \sim \mathsf{N}(\boldsymbol{\nu}, \Sigma/(1-\theta))$ for some $\theta \in [0,1)$. Thus, by taking a θ close to unity, the variance of $\boldsymbol{Y}$ can be boosted sufficiently to induce the event $\{S > \gamma\}$, where $S = \exp(Y_1) + \cdots + \exp(Y_d)$. Then, the variance boosted estimator:

$$\hat{\ell}_\theta(\gamma) = \frac{\exp(-\theta(\boldsymbol{Y}-\boldsymbol{\nu})^\top \Sigma^{-1}(\boldsymbol{Y}-\boldsymbol{\nu})/2)}{(1-\theta)^{d/2}} \mathbb{I}\{S > \gamma\}, \qquad \boldsymbol{Y} \sim \mathbb{P}_\theta \tag{3}$$

is simply the result of using $\mathbb{P}_\theta$ as an importance sampling measure. The asymptotically optimal choice of θ is

$$\theta = 1 - \frac{\sigma^2}{2(\ln(\gamma) - \nu)} + o(\ln^{-1}\gamma), \quad \gamma \uparrow \infty. \tag{4}$$

It is then not difficult to show, (Asmussen et al. 2011), that with this choice of θ

$$\frac{\mathbb{E}_\theta \hat{\ell}_\theta^2}{\ell^2} = \Theta([\ln\gamma]^{d/2+1}\gamma^{1/4})$$

In other words,

$$\liminf_\gamma \frac{\ln \mathbb{V}\mathrm{ar}_\theta(\hat{\ell}_\theta)}{\ln \ell(\gamma)} = 2,$$

which is the definition of a *logarithmically efficient* estimator. The efficiency label arises from the fact that the relative error, $\mathbb{V}\mathrm{ar}(\hat{\ell}_0)/\ell^2$, of the crude Monte Carlo estimator,

$$\hat{\ell}_0 = \mathbb{I}\{S > \gamma\}, \quad \boldsymbol{Y} \sim \mathbb{P}_0,$$

is of the order

$$\ell^{-1}(\gamma) = \Theta\left(\exp\left(\tfrac{(\ln(\gamma)-\nu)^2}{2\sigma^2}\right)\tfrac{\ln\gamma-\nu}{\sigma}\right),$$

where $\sigma = \max_k \sigma_k$ and $\nu = \max\{\nu_k : \sigma_k = \sigma\}$. Thus, while the error of $\hat{\ell}_0$ grows exponentially, the error of $\hat{\ell}_\theta$ grows only polynomially.

Consider an simple example in which all log-normals are iid with $\Sigma = \mathrm{I} \times 0.25^2, \boldsymbol{\nu} = \mathbf{0}$, and we have $d = 30$. Table 1 shows the estimated values for $\ell(\gamma)$ for different values of γ using three different estimators: the variance boosted $\hat{\ell}_\theta$; the Asmussen-Kroese estimator (Asmussen and Kroese 2006),

$$\hat{\ell}_{\mathrm{AK}} = d\overline{\Phi}\left(\tfrac{1}{\sigma}\ln\left[(\gamma - \textstyle\sum_{j<d} X_j) \vee \max_{j<d} X_j\right]\right), \quad \ln \boldsymbol{X} \sim \mathsf{N}(\mathbf{0}, \sigma^2\mathrm{I});$$

and our proposed estimator $\hat{\ell}$ in Section 3. The data was populated using $n = 10^7$ independent replications of each estimator. The difference in the CPU run times for all methods was negligible (all between 7 to 10 seconds), and hence not reported here. The conclusion from the results in the table is that the variance boosted estimator, $\hat{\ell}_\theta$, is not useful due to its high variability.

Table 1: Comparative performance of the variance-boosted and Asmussen-Kroese estimators. The proposed estimator $\hat{\ell}$ is given in column two and described in Section 3.

		relative error %		
γ	$\hat{\ell}$	$\mathrm{RE}(\hat{\ell})$	$\mathrm{RE}(\hat{\ell}_{\mathrm{AK}})$	$\mathrm{RE}(\hat{\ell}_\theta)$
30	0.742	0.199	0.0321	0.314
33	0.0797	0.26	0.0871	3.67
36	0.00052	0.403	0.684	39.8
39	2.94e-07	0.725	17.9	51.9
42	2.29e-11	1.45	54.6	99.9
45	3.92e-16	2.57	64.4	97.8
48	1.93e-21	4.44	31.7	97
51	3.98e-27	7.85	25.2	81.5
54	8.58e-33	3.22	15.3	100
57	3.44e-36	0.418	13.3	69.8
60	4.26e-39	0.203	5.21	99.7
63	1.06e-41	0.18	2.92	99
66	4.38e-44	0.162	1.58	64.8
69	2.75e-46	0.16	1.09	100
72	2.42e-48	0.155	0.686	98.3
75	2.83e-50	0.153	0.498	72.1
78	4.24e-52	0.151	0.414	95.7
81	7.87e-54	0.15	0.287	99.3
84	1.78e-55	0.15	0.26	100
87	4.74e-57	0.15	0.251	90.5
90	1.48e-58	0.15	0.189	100

It is important to note that while the authors of (Asmussen et al. 2011) select the variance-minimal value of θ heuristically using the so-called *cross-entropy* method, an exhaustive search shows that there is no value for θ that yields reasonably low variance. For example, Figure 1 shows the estimated relative error of $\hat{\ell}_\theta$ as a function of θ for $\gamma = 45$ and all other parameters being the same as in Table 1. The figure suggests that even if we knew the true variance-minimizing θ^* (obviating the need for approximating it), the estimator will still not be useful.

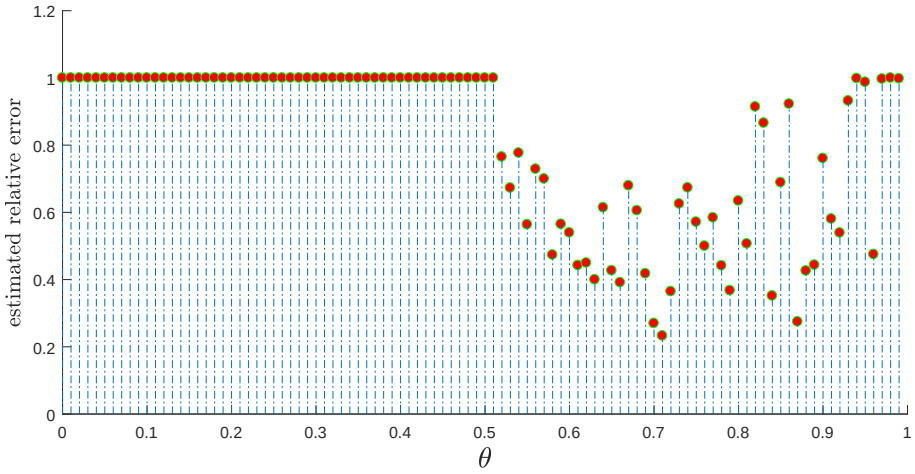

Figure 1: The estimated relative error of $\hat{\ell}_\theta$ as a function of θ using 10^7 replications. The smallest estimated relative error was 23%, corresponding to $\theta = 0.71$. Where the estimate of $\ell(45)$ is 0, the relative error is recorded as unity (100%).

2.1 A Vanishing Relative Error Estimator

Recognizing the deficiency of the variance boosted estimator (Asmussen et al. 2011) propose the superior ISVE estimator. Let $\boldsymbol{X} = \exp(\boldsymbol{Y})$ and denote $M = \max_i X_i$. The main idea of the ISVE estimator is to split (1) into two parts:

$$\ell = \mathbb{P}(M > \gamma) + \mathbb{P}(S > \gamma, M < \gamma),$$

and estimate $\ell_1 = \mathbb{P}(M > \gamma)$ and $\ell_2 = \mathbb{P}(S > \gamma, M < \gamma)$ separately using two different importance sampling schemes. In particular, ℓ_1 is estimated via

$$\hat{\ell}_1 = \frac{\ell_{\text{asymp}}(\gamma)}{\sum_{k=1}^d \mathbb{I}\{X_k > \gamma\}}, \quad \boldsymbol{X} \sim g(\boldsymbol{x}), \tag{5}$$

where g is the mixture density:

$$g(\boldsymbol{x}) \stackrel{\text{def}}{=} \frac{\phi_\Sigma(\boldsymbol{x} - \boldsymbol{\nu}) \sum_{k=1}^d \mathbb{I}\{x_k > \gamma\}}{\ell_{\text{asymp}}(\gamma)},$$

and the residual probability, ℓ_2, is estimated via a variance boosted estimator:

$$\hat{\ell}_{2,\theta}(\gamma) = \frac{\exp(-\theta(\boldsymbol{Y} - \boldsymbol{\nu})^\top \Sigma^{-1}(\boldsymbol{Y} - \boldsymbol{\nu})/2)}{(1-\theta)^{d/2}} \mathbb{I}\{S > \gamma, M < \gamma\}, \qquad \boldsymbol{Y} \sim \mathbb{P}_\theta, \tag{6}$$

where $\theta = 1 - \ln^{-2}(\gamma)$. With this setup the ISVE estimator is $\hat{\ell}_{\text{ISVE}} = \hat{\ell}_1 + \hat{\ell}_2$ and it enjoys the vanishing relative error property (Asmussen et al. 2011):

$$\frac{\mathbb{V}\text{ar}(\hat{\ell}_{\text{ISVE}})}{\ell^2(\gamma)} \downarrow 0, \quad \gamma \uparrow \infty.$$

Before we proceed to illustrate the practical performance of the ISVE estimator, we note that there are two issues that may indicate problematic performance.

First, using the sample variance of n independent replications of (5) is not a robust estimator of the true variance of $\hat{\ell}_1$ in the following sense.

Proposition 1 (Inefficiency of Sample Variance) Let S_n^2 be the sample variance based on n independent replications of (5) . Then,

$$\liminf_{\gamma \uparrow \infty} \frac{\mathbb{V}\mathrm{ar}(S_n^2)}{\mathbb{V}\mathrm{ar}^2(\hat{\ell}_1)} = \infty,$$

where the rate of growth to infinity can be as high as exponential: $c_1 \ln(\gamma) \exp(c_2 \ln^2(\gamma))$ for some positive constants c_1 and c_2.

Proof. The proof is a straightforward adaptation of the proof of Proposition 2 in (Botev et al. 2015), where the same statement holds for the estimator of the maximum of correlated Gaussian random variables, namely, $\mathbb{P}(\max_i Y_i > \gamma)$ with $\boldsymbol{Y} \sim \mathsf{N}(\boldsymbol{\nu}, \Sigma)$. Since the maximum of d log-normal random variables, M, can be linked to the maximum of d multivariate normal variables:

$$\mathbb{P}(M > \gamma) = \mathbb{P}(\max_i \exp(Y_i) > \gamma) = \mathbb{P}(\max_i Y_i \geq \ln \gamma),$$

the only necessary notational change in the proof of (Botev et al. 2015, Proposition 2) is to replace all instances of γ with $\ln(\gamma)$. □

The practical consequence of the result above is that the relative error of $\hat{\ell}_1$ is underestimated during simulation, and frequently reported as being zero.

Second, we already know that the variance boosted estimator (3) is unreliable for estimating ℓ, and that there is no value for θ that will render it a useful estimator. Upon examination of (6) we see that it only differs from (3) with the addition of the constraint $M < \gamma$, and in the different choice of θ. Thus, given that (3) is a bad estimator of ℓ for any θ, we should not be surprised to find that (6) is also a poor estimator of ℓ_2. Indeed, the same numerical example again demonstrates that there is no good value for θ that can make the relative error of (6) small enough. The behavior of the relative error of $\hat{\ell}_{2,\theta}$ as a function of θ is qualitatively the same as that on Figure 1.

2.2 Quality of Asymptotic Approximation

One of the arguments in favor of the ISVE estimator is that, while $\hat{\ell}_2$ may be a noisy estimator, it is a very small second order residual term, and will not affect noticeably the high accuracy of the leading order term $\hat{\ell}_1$.

Unfortunately, more often than not, the leading contribution term of $\ell = \ell_1 + \ell_2$ is not ℓ_1, but the residual ℓ_2. This may appear to contradict the fact that asymptotically $\ell \simeq \ell_1$, but it makes sense when one takes into account that, in the presence of a positive correlation, ℓ_1, or equivalently[3] the ℓ_{asymp}, can be an extremely poor approximation to ℓ.

For example, consider the case with $\boldsymbol{\nu} = \mathbf{0}$ and $\Sigma = \rho \mathbf{1}\mathbf{1}^\top + (1-\rho)\mathrm{I}$ for $\rho \in [0, 1)$, then (Kortschak and Hashorva 2014) show that

$$\begin{aligned} \ell(\gamma) - \ell_{\mathrm{asymp}}(\gamma) &= \ell(\gamma) - d\overline{\Phi}(\ln \gamma) \\ &\simeq d(d-1) \exp((1-\rho^2)/2) \frac{\ln(\gamma)}{\gamma^{1-\rho}} \overline{\Phi}(\ln \gamma) \end{aligned}$$

Figure 2.2 shows a logarithmic plot of $r(\rho)$ — the relative asymptotic deviation of ℓ from ℓ_{asymp},

$$\frac{\ell - \ell_{\mathrm{asymp}}}{\ell_{\mathrm{asymp}}} \simeq r(\rho) \stackrel{\mathrm{def}}{=} (d-1) \exp((1-\rho^2)/2) \frac{\ln(\gamma)}{\gamma^{1-\rho}}$$

[3] Note that $\frac{\ell_{\mathrm{asymp}}}{d} \leq \ell_1 \leq \ell_{\mathrm{asymp}}$

with $d = 100$, as function of ρ for three different values of $\ell_{\text{asymp}}(\gamma) \in \{10^{-5}, 10^{-10}, 10^{-40}\}$. What we can gather from the plot is that for large ρ the relative deviation of ℓ from the asymptotic approximation ℓ_{asymp} is of order 10^3 even for the most extreme of rare-events, for which we would have expected the asymptotic approximation to be accurate.

Monte Carlo simulations actually show that the situation can be even worse. Take, for example, the instances in Table 2.2, where $d = 10$ and $\boldsymbol{\nu} = \mathbf{0}, \Sigma = 0.25^2 \times (0.9 \times \mathbf{1}\mathbf{1}^\top + (1 - 0.9) \times \mathrm{I})$. The table shows the asymptotic value ℓ_{asymp} for different γ (second column), together with its relative deviation from the true ℓ (last column). The table also displays $\hat{\ell}$ with its approximate 95% confidence interval based on $n = 10^6$ independent replications of our method, described in the next section.

Figure 2: Asymptotic estimate of the relative deviation of ℓ from ℓ_1.

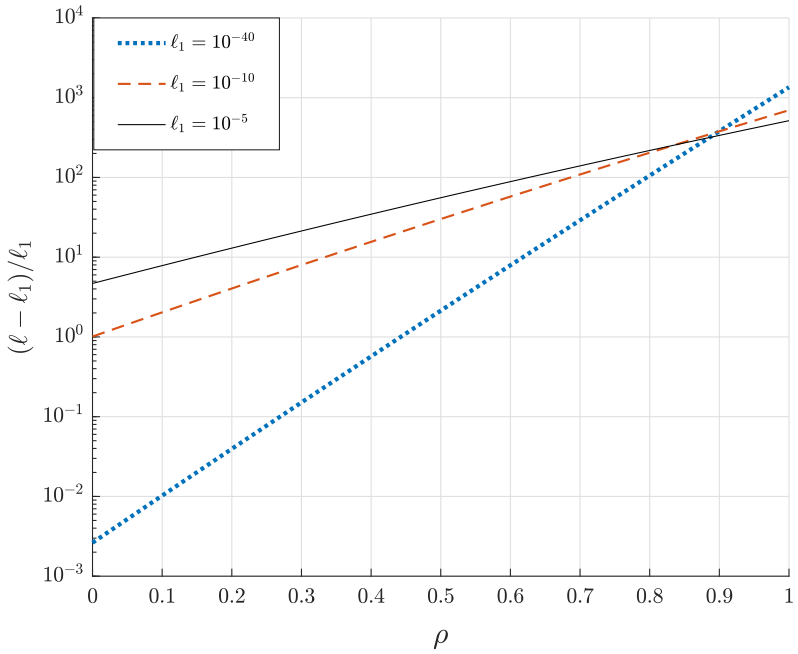

Table 2: Illustration of the inaccuracy of the asymptotic approximation for moderate values of $\ell(\gamma)$. The asymptotic approximation ℓ_{asymp} becomes useful only for probabilities smaller than about 10^{-233}.

γ	ℓ_{asymp}	95% CI for $\hat{\ell}$	$(\hat{\ell} - \ell_{\text{asymp}})/\ell_{\text{asymp}}$
15	$1.2113\ldots \times 10^{-26}$	0.012 ± 0.001	1.0×10^{24}
20	$2.1830\ldots \times 10^{-32}$	$(5.80 \pm 0.013) \times 10^{-5}$	2.66×10^{27}
40	$1.4175\ldots \times 10^{-48}$	$(6.33 \pm 0.016) \times 10^{-15}$	4.5×10^{33}
60	$1.3872\ldots \times 10^{-59}$	$(1.10 \pm 0.017) \times 10^{-23}$	8×10^{35}
100	$4.4834\ldots \times 10^{-75}$	$(8.04 \pm .018) \times 10^{-38}$	1.8×10^{37}
500	$1.0481\ldots \times 10^{-135}$	$(3.39 \pm .02) \times 10^{-105}$	3.2×10^{30}
1000	$2.3594\ldots \times 10^{-167}$	$(6.94 \pm .02) \times 10^{-145}$	3×10^{22}
1500	$2.0634\ldots \times 10^{-187}$	$(4.04 \pm .03) \times 10^{-171}$	2×10^{16}
2500	$2.6294\ldots \times 10^{-214}$	$(2.94 \pm .04) \times 10^{-207}$	1.1×10^{7}
3500	$5.1912\ldots \times 10^{-233}$	$(5.45 \pm .04) \times 10^{-233}$	0.05

The message from Table 2.2 is clear: the asymptotic approximation is useless for moderate values of γ (deviating from the true value of ℓ by as much as 10^{37}), and only becomes useful for extremely small probabilities (smaller than 10^{-233}). It is interesting that the relative deviation becomes much bigger before finally improving for extremely small probabilities.

3 Exponentially Tilted Estimator

In this section we introduce a novel estimator based on the method of exponential tilting. To this end, we first define $p_k(\gamma) \stackrel{\text{def}}{=} \mathbb{P}(Y_k \geq \ln(\gamma/d))$ and $p(\gamma) = \sum_k p_k(\gamma)$. Recall that $X_k = \exp(Y_k)$, then our proposed unbiased estimator is:

$$\hat{\ell}(\gamma) = \frac{\mathbb{I}\{K=k\}}{p_k/p} \exp(\mu_k^2/(2\sigma_k^2) - \mu_k(Y_k - \nu_k)/\sigma_k^2)\mathbb{I}\{S > \gamma, X_k = M\},$$

where μ_k is the solution to

$$\mu_k^* = \text{argmin}_\mu \left\{ \tfrac{\mu_k^2}{\sigma_k^2} + \ln \overline{\Phi}\left(\tfrac{\ln(\gamma/d) - \nu_k + \mu_k}{\sigma_k}\right)\right\} \tag{7}$$

and the vector $(K, Y_k, \boldsymbol{Y}_{-k})$ is simulated sequentially from the joint pdf:

$$(K, Y_k, \boldsymbol{Y}_{-k}) \sim g(k, \boldsymbol{y}) = p_k \phi_{\sigma_k^2}(y_k - \nu_k - \mu_k)\phi_\Sigma(\boldsymbol{y} - \boldsymbol{\nu}|y_k)$$

Implementation of one replication of the estimator is summarized in the following algorithm.

Algorithm 1 : Estimator of $\ell(\gamma)$

Require: $\boldsymbol{\nu}$, Σ, γ
- $\hat{\ell} \leftarrow 0$
- Simulate K from the distribution $\mathbb{P}(K=k) = p_k(\gamma)/p(\gamma)$ and set $k \leftarrow K$
- Solve (7) to deliver μ_k^*
- Simulate $Y_k \sim \mathsf{N}(\nu_k + \mu_k^*, \sigma_k^2)$ and set $y_k \leftarrow Y_k$
- Simulate the Gaussian vector $\boldsymbol{Y}_{-k} = (Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_d)^\top \sim \phi_\Sigma(\boldsymbol{y} - \boldsymbol{\nu}|y_k)$, given $Y_k = y_k$
- $\boldsymbol{X} \leftarrow (\exp(Y_1), \ldots, \exp(Y_d))^\top$ and $S \leftarrow X_1 + \cdots + X_d$
- **if** $S > \gamma$ and $X_k = \max_i X_i$ **then**
  - $\hat{\ell} \leftarrow p/p_k \exp((\mu_k^*)^2/(2\sigma_k^2) - \mu_k^*(Y_k - \nu_k)/\sigma_k^2)$
- **return** $\hat{\ell}$

The next result ensures that the estimator, just like its competitors, is robust with respect to γ.

Theorem 1 (Logarithmic Efficiency) With the choice (7), the estimator $\hat{\ell}$ is both unbiased and logarithmically efficient, that is,

$$\liminf_\gamma \frac{\ln \mathbb{V}\text{ar}_g(\hat{\ell}(\gamma))}{\ln \mathbb{E}_g \hat{\ell}(\gamma)} = 2,$$

where $\mathbb{E}_g$ and $\mathbb{V}\text{ar}_g$ denote expectation under the importance sampling density g.

Proof. First, recall that

$$\ell(\gamma) = \Theta\left(\overline{\Phi}((\ln\gamma - \nu)/\sigma)\right), \qquad \gamma \uparrow \infty,$$

where $\sigma = \max_k \sigma_k$ and $\nu = \max\{\nu_k : \sigma_k = \sigma\}$. Second, note that the estimator is unbiased, because

$$\begin{aligned}
\mathbb{E}_g \hat{\ell}(\gamma) &= \sum_k \mathbb{E}_{g(\cdot|k)} \exp(\mu_k^2/(2\sigma_k^2) - \mu_k(Y_k - \nu_k)/\sigma_k^2)\mathbb{I}\{S > \gamma, \exp(Y_k) = M\} \\
&= \sum_k \mathbb{P}(S > \gamma, X_k = M) = \mathbb{P}(S > \gamma)
\end{aligned}$$

The second moment of the importance sampling estimator is then bounded from above as follows:

$$\begin{aligned}\mathbb{E}_g\hat{\ell}^2(\gamma) &= \textstyle\sum_k \frac{p}{p_k}\mathbb{E}\exp(\mu_k^2/(2\sigma_k^2) - \mu_k(Y_k-\nu_k)/\sigma_k^2)\mathbb{I}\{S>\gamma, \exp(Y_k)=M\}\\ &= p(\gamma)\textstyle\sum_k \frac{\exp(\mu_k^2/\sigma_k^2)}{p_k(\gamma)}\mathbb{P}_{\mu_k}(S>\gamma, X_k = M),\end{aligned}$$

where under $\mathbb{P}_{\mu_k}$ we have $Y_k = \ln(X_k) \sim \mathsf{N}(\nu_k - \mu_k, \sigma_k^2)$. Hence, using the fact that

$$\mathbb{P}_{\mu_k}(S>\gamma, X_k = M) \le \mathbb{P}_{\mu_k}(X_k > \gamma/d) = \mathbb{P}(Y_k - \mu_k > \ln(\gamma/d))$$

we obtain with $\mu_k = \mu_k^*$ in (7):

$$\begin{aligned}\mathbb{E}_g\hat{\ell}^2(\gamma) &= p(\gamma)\textstyle\sum_k \frac{\exp((\mu_k^*/\sigma_k)^2)}{p_k(\gamma)}\mathbb{P}_{\mu_k^*}(S>\gamma, X_k = M)\\ &\le p(\gamma)\textstyle\sum_k \frac{\exp((\mu_k^*/\sigma_k)^2)}{p_k(\gamma)}\overline{\Phi}\left(\frac{\ln(\gamma/d)-\nu_k+\mu_k^*}{\sigma_k}\right)\\ (\text{for any } \mu_k \neq \mu_k^*) \quad &\le p(\gamma)\textstyle\sum_k \frac{\exp(\mu_k^2/\sigma_k^2)}{p_k(\gamma)}\overline{\Phi}\left(\frac{\ln(\gamma/d)-\nu_k+\mu_k}{\sigma_k}\right)\\ (\text{substitute } \mu_k = \ln(\gamma/d)-\nu_k \neq \mu_k^*) \quad &\le p(\gamma)\textstyle\sum_k \frac{\exp((\ln(\gamma/d)-\nu_k)^2/\sigma_k^2)}{p_k(\gamma)}\overline{\Phi}\left(2\frac{\ln(\gamma/d)-\nu_k}{\sigma_k}\right)\end{aligned}$$

Next, we set $t_k = (\ln(\gamma/d) - \nu_k)/\sigma_k$ as shorthand notation, and use the facts: first, $\frac{\phi(x)}{\overline{\Phi}(x)} = \Theta(x)$ for $x \uparrow \infty$; and second, with $t \stackrel{\text{def}}{=} \frac{\ln(\gamma/d)-\nu}{\sigma}$, $p(\gamma) = \Theta\left(\overline{\Phi}(t)\right)$. Using these, we have

$$\begin{aligned}\mathbb{E}_g\hat{\ell}^2(\gamma) &\le p(\gamma)\textstyle\sum_k \frac{\exp(t_k^2)}{p_k(\gamma)}\overline{\Phi}(2t_k)\\ &= \Theta\left(\overline{\Phi}(t)\right)\textstyle\sum_k \frac{\exp(t_k^2)}{\overline{\Phi}(t_k)}\overline{\Phi}(2t_k)\\ &= \Theta\left(\overline{\Phi}(t)\right)\textstyle\sum_k \exp(t_k^2)\Theta\left(\frac{\phi(2t_k)}{\phi(t_k)}\right)\\ &= \Theta\left(\overline{\Phi}(t)\right)\textstyle\sum_k \Theta\left(\phi(t_k)\right) = \Theta\left(\overline{\Phi}(t)\,\phi(t)\right)\end{aligned}$$

Therefore, using the fact that $\frac{\ln\overline{\Phi}(x+c)}{\ln\overline{\Phi}(x)} \to 1$ as $x \uparrow \infty$ for any constant c, standard calculus shows that

$$\lim_{\gamma\uparrow\infty}\frac{\ln\mathbb{E}_g\hat{\ell}^2(\gamma)}{\ln\ell} = \lim_{\gamma\uparrow\infty}\frac{\ln\overline{\Phi}(t(\gamma)) + \ln\phi(t(\gamma))}{\ln\overline{\Phi}((\ln\gamma-\nu)/\sigma)} = 2$$

This concludes the proof. □

4 A Numerical Example

All the estimators discussed in this article are asymptotically efficient. Yet, the numerical experiments suggest that some of them are not always useful in practice. Consider estimating $\ell(\gamma)$ with $\boldsymbol{\nu} = \mathbf{0}$ and $\Sigma = 0.25^2 \times (\rho \times \mathbf{1}\mathbf{1}^\top + (1-\rho)\times \mathrm{I})$ for $\rho = 0.9, d = 30$ and different values of γ. Table 3 gives the results using $n = 10^6$ replications. The *work normalized relative variance* of $\hat{\ell}$ is defined as $\tau \times \mathbb{V}\mathrm{ar}(\hat{\ell})/\ell^2$, where τ is the total CPU time needed to deliver $\hat{\ell}$. For the ISVE estimator we attempted to optimize the performance of the estimator by manually selecting the best possible θ. Our choice for this tuning parameter is thus given in brackets in the third column.

Table 3: Comparative performance of the ISVE and exponentially tilted estimators with $\rho = 0.9, d = 30$.

			relative error %		work normalized relative variance	
γ	$\hat{\ell}$	$\hat{\ell}_{\mathrm{ISVE}}$	$\mathrm{RE}(\hat{\ell})$	$\mathrm{RE}(\hat{\ell}_{\mathrm{ISVE}})$	$\mathrm{WNRV}(\hat{\ell})$	$\mathrm{WNRV}(\hat{\ell}_{\mathrm{ISVE}})$
40	0.116	0.114 ($\theta = 0.5$)	0.63	2.0	0.00032	0.00080
100	2.17×10^{-7}	1.18×10^{-7} ($\theta = 0.6$)	0.98	40	0.00061	0.31
150	6.83×10^{-12}	5.75×10^{-13} ($\theta = 0.75$)	1.1	84	0.00093	1.12
200	7.75×10^{-16}	2.09×10^{-17} ($\theta = 0.8$)	1.2	95	0.0010	1.22
400	6.57×10^{-28}	3.08×10^{-39} ($\theta = 0.9$)	1.4	80	0.0011	1.34
10^3	1.61×10^{-49}	1.21×10^{-80} ($\theta = 0.95$)	1.7	100	0.002	2.02
10^4	3.60×10^{-132}	1.80×10^{-294} ($\theta =?$)	2.1	-	0.0024	-

A number of conclusions can be drawn from the table.

First, the ISVE estimator does not have acceptably low variance for both small γ (when the event is not rare) and for large γ (when the event is rare).

Second, as with Figure 1, any attempt to optimize with respect to θ is fruitless, because there appears to be no value for $\theta \in [0, 1)$ that yields low variance.

Third, in the last row of the table, it was not possible to induce the event $\{S > \gamma, M < \gamma\}$ no matter what the value of θ. In other words, $\{S > \gamma, M < \gamma\}$ remains a rare-event for all values of $\theta \in [0, 1)$, and with very high probability $\hat{\ell}_{\mathrm{ISVE}} = \hat{\ell}_1 + \hat{\ell}_2 = \hat{\ell}_1$. Thus, despite the vanishing relative error property of the ISVE estimator, its performance deteriorates as γ becomes smaller and smaller to the point that it does not deliver meaningful estimates.

Of course, if γ is large enough, then we know from theory that $\hat{\ell}_{\mathrm{ISVE}}$ must ultimately yield a meaningful result. However, in our case γ has to be so large and the rare event so rare, that it is not possible to store the relevant numbers in computer memory due to numerical over- and under-flow issues. This is because the asymptotic approximation (2) is, at least in this particular case, not useful for small to moderately large γ.

5 Conclusions

We have presented a new exponentially tilted estimator, $\hat{\ell}$, for the estimation of the tail of the sum of dependent log-normal variables (1). The proposed estimator is shown to be, not only asymptotically optimal, but also useful in practical simulations. One of the observations we can draw from a number of numerical experiments is that sometimes an estimator with a vanishing relative error property may not necessarily exhibit low variance in practical simulations.

As future work, we propose to show that the growth of the relative error of $\hat{\ell}$ can be shown to be as small as $\mathcal{O}(\ln \gamma)$. Further, one would like to show that the estimator $\hat{\ell}$ enjoys the bounded normal approximation (Tuffin 1999). In other words, we would like to show that the sample standard deviation of an ensemble of $\hat{\ell}$ is a robust estimator of the true error of $\hat{\ell}$, and inefficiencies like the one defined in Proposition 1 do not hold for the estimator $\hat{\ell}$.

ACKNOWLEDGMENTS

Zdravko Botev has been supported by the Australian Research Council grant DE140100993.

REFERENCES

Asmussen, S., J. Blanchet, S. Juneja, and L. Rojas-Nandayapa. 2011. "Efficient simulation of tail probabilities of sums of correlated lognormals". *Annals of Operations Research* 189 (1): 5–23.

Asmussen, S., and D. Kortschak. 2015. "Error rates and improved algorithms for rare event simulation with heavy Weibull tails". *Methodology and Computing in Applied Probability* 17 (2): 441.

Asmussen, S., and D. P. Kroese. 2006. "Improved algorithms for rare event simulation with heavy tails". *Advances in Applied Probability* 38 (02): 545–558.

Asmussen, S., and L. Rojas-Nandayapa. 2008. "Asymptotics of sums of lognormal random variables with Gaussian copula". *Statistics & Probability Letters* 78 (16): 2709–2714.

Botev, Z. I., M. Mandjes, and A. Ridder. 2015. "Tail distribution of the maximum of correlated gaussian random variables". In *Winter Simulation Conference (WSC), 2015*, 633–642. IEEE.

Kortschak, D., and E. Hashorva. 2014. "Second order asymptotics of aggregated log-elliptical risk". *Methodology and Computing in Applied Probability* 16 (4): 969–985.

Kroese, D. P., T. Taimre, and Z. I. Botev. 2011. *Handbook of Monte Carlo methods*, Volume 706. John Wiley & Sons.

Nguyen, Q. H., and C. Y. Robert. 2014. "New efficient estimators in rare event simulation with heavy tails". *Journal of Computational and Applied Mathematics* 261:39–47.

Ortobelli, S., T. Lando, F. Petronio, and T. Tichỳ. 2016. "Asymptotic stochastic dominance rules for sums of iid random variables". *Journal of Computational and Applied Mathematics* 300:432–448.

Tuffin, B. 1999. "Bounded normal approximation in simulations of highly reliable Markovian systems". *Journal of Applied Probability* 36 (04): 974–986.

AUTHOR BIOGRAPHIES

ZDRAVKO BOTEV is Senior Lecturer at the School of Mathematics and Statistics at the University of New South Wales in Sydney, Australia. Previously he was an Australian Research Council fellow and a postdoctoral fellow at the University of Montreal. Currently, he is serving as an associate editor of the INFORMS Journal of Computing. His research interests include splitting and adaptive importance sampling methods for rare-event simulation. He is the author the most accurate nonparametric density estimation software, which has been widely used in the applied sciences. For more information, visit his webpage: http://web.maths.unsw.edu.au/~zdravkobotev/.

PIERRE L'ECUYER is Professor in the Département d'Informatique et de Recherche Opérationnelle, at the Université de Montréal, Canada. He holds the Canada Research Chair in Stochastic Simulation and Optimization and an Inria International Chair in Rennes, France. He is a member of the CIRRELT and GERAD research centers. His main research interests are random number generation, quasi-Monte Carlo methods, efficiency improvement via variance reduction, sensitivity analysis and optimization of discrete-event stochastic systems, and discrete-event simulation in general. He has served as Editor-in-Chief for *ACM Transactions on Modeling and Computer Simulation* from 2010 to 2013. He is currently Associate Editor for *ACM Transactions on Mathematical Software*, *Statistics and Computing*, and *International Transactions in Operational Research*. More information can be found on his web page: http://www.iro.umontreal.ca/~lecuyer.