Botev, Zdravko; Mandjes, Michel; Ridder, Ad

Working Paper

Tail Distribution of the Maximum of Correlated Gaussian Random Variables

Tinbergen Institute Discussion Paper, No. 15-132/III

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

TI 2015-132/III
Tinbergen Institute Discussion Paper

Tail Distribution of the Maximum of Correlated Gaussian Random Variables

Zdravko Botev[a]
Michel Mandjes[b]
Ad Ridder[c]

[a] *The University of New South Wales, Sydney, Australia;*
[b] *University of Amsterdam, the Netherlands;*
[c] *Faculty of Economics and Business Adminitration, VU University Amsterdam, and Tinbergen Institute, the Netherlands.*

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Tail Distribution of the Maximum of Correlated Gaussian Random Variables

Zdravko I. Botev[a], Michel Mandjes[b], Ad Ridder[c]

[a] The University of New South Wales,
Sydney, NSW 2052, Australia
botev@unsw.edu.au

[b] University of Amsterdam,
Amsterdam, The Netherlands
m.r.h.mandjes@uva.nl

[c] Vrije University & Tinbergen Institute,
Amsterdam, The Netherlands
ad.ridder@vu.nl

December 11, 2015

Abstract

In this article we consider the efficient estimation of the tail distribution of the maximum of correlated normal random variables. We show that the currently recommended Monte Carlo estimator has difficulties in quantifying its precision, because its sample variance estimator is an inefficient estimator of the true variance. We propose a simple remedy: to still use this estimator, but to rely on an alternative quantification of its precision. In addition to this we also consider a completely new sequential importance sampling estimator of the desired tail probability. Numerical experiments suggest that the sequential importance sampling estimator can be significantly more efficient than its competitor.

Keywords. Rare event simulation; Correlated Gaussian; Tail probabilities; Sequential importance sampling.

1 Introduction

Let $\boldsymbol{X}$ be a d-dimensional random variable with an $\mathsf{N}(\boldsymbol{a}, \Sigma)$ distribution and corresponding multivariate normal density $\phi(\boldsymbol{x}; \boldsymbol{a}, \Sigma)$. In this paper we focus on techniques to efficiently estimate the tail distribution

$$\ell(\gamma) \stackrel{\text{def}}{=} \mathbb{P}\left(\max_{i \in \{1,\ldots,d\}} X_i \geq \gamma\right),$$

in the asymptotic regime that γ grows large. This problem arises, for example, in dealing with Gaussian random fields in the physical sciences, but in various other application domains as well; see e.g. Adler and Taylor (2009), Adler et al. (2012), and Mandjes (2007), and the references therein.

Recently, Adler et al. (2008) and Blanchet and Li (2011) proposed an ingenious strongly efficient estimator for $\ell(\gamma)$. Their idea is to recognize that, under the condition that, for all pairs of distinct i, j in $\{1, \ldots, d\}$, $\mathbb{P}(X_i > \gamma \,|\, X_j > \gamma) = o(1)$ as $\gamma \uparrow \infty$ (which applies when $\mathbb{C}or(X_i, X_j) < 1$; we comment on this later), the inclusion-exclusion formula implies that[1]

$$\ell(\gamma) = \mathbb{P}\left(\bigcup_{i=1}^{d} \{X_i > \gamma\}\right) = \sum_{i=1}^{d} \mathbb{P}(X_i > \gamma) + \mathcal{O}\left(\sum_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma)\right), \quad \gamma \uparrow \infty.$$

The next step is to realize that, with $\overline{\Phi}$ denoting the complementary cdf of the $\mathsf{N}(0,1)$ estimator, the marginal probabilities

$$\alpha_i(\gamma) \stackrel{\text{def}}{=} \mathbb{P}(X_i > \gamma) = \overline{\Phi}\left(\frac{\gamma - a_i}{\sqrt{\Sigma_{ii}}}\right), \quad i = 1, \ldots, d,$$

are known (i.e., can be efficiently evaluated with arbitrary precision). As a consequence, the value

$$\alpha(\gamma) \stackrel{\text{def}}{=} \sum_{i=1}^{d} \mathbb{P}(X_i > \gamma)$$

is also easily computable. These observations suggest the following *mixture importance sampling* density, with $\phi(\boldsymbol{x}_{-i} \,|\, x_i; \boldsymbol{a}, \Sigma)$ defined in the obvious way:

$$\begin{aligned} m(\boldsymbol{x}) &= \frac{\phi(\boldsymbol{x}; \boldsymbol{a}, \Sigma) \sum_{i=1}^{d} \mathbb{I}\{x_i > \gamma\}}{\alpha(\gamma)} \\ &= \sum_{i=1}^{d} w_i \, \frac{\phi(x_i; a_i, \Sigma_{ii}) \mathbb{I}\{x_i > \gamma\}}{\alpha_i(\gamma)} \phi(\boldsymbol{x}_{-i} \,|\, x_i; \boldsymbol{a}, \Sigma), \quad \text{where } w_i \stackrel{\text{def}}{=} \frac{\alpha_i(\gamma)}{\alpha(\gamma)}. \end{aligned}$$

It is then straightforward to verify that the resulting importance sampling estimator, with $\boldsymbol{X} \sim m(\boldsymbol{x})$,

$$\breve{\ell} = \frac{\alpha(\gamma)}{\sum_{i=1}^{d} \mathbb{I}\{X_i > \gamma\}}$$

is a *vanishing relative error* one in the sense that, as an immediate consequence of the fact that the likelihood ratio is bounded from above by $\alpha(\gamma)$,

$$\frac{\mathbb{V}ar_m(\breve{\ell})}{\ell^2(\gamma)} = \frac{\mathbb{E}_m[\breve{\ell}^2]}{\ell^2(\gamma)} - 1 \leq \frac{\alpha^2(\gamma)}{\ell^2(\gamma)} - 1 = o(1), \quad \gamma \uparrow \infty;$$

here $\mathbb{E}_m[\cdot]$ and $\mathbb{V}ar_m(\cdot)$ denote the expectation and variance operators with respect to the density $m(\cdot)$.

In practical simulations Adler et al. (2008); Blanchet and Li (2011) estimate the precision by generating n independent realizations of $\breve{\ell}$, namely $\breve{\ell}_1, \ldots, \breve{\ell}_n$, and then computing the corresponding sample variance

$$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (\breve{\ell}_i - \bar{\ell})^2,$$

where $\bar{\ell} = (\breve{\ell}_1 + \cdots + \breve{\ell}_n)/n$. Ideally, $S_n/(\bar{\ell}\sqrt{n})$ would then yield a consistent estimator of the relative error of the sample mean $\bar{\ell}$ and in numerical experiments we would report either the pair $\bar{\ell}$ and $S_n/(\bar{\ell}\sqrt{n})$, or (say) the 95% approximate confidence interval $\bar{\ell} \pm 1.96 \times S_n/\sqrt{n}$.

[1]Throughout the paper $x = o(y)$ stands for $\lim_{y \uparrow \infty} x/y = 0$ and $x = \mathcal{O}(y)$ stands for $\limsup_{y \uparrow \infty} |x/y| < \infty$.

Despite the vanishing relative error property of $\bar{\ell}$, the practical performance of its error estimate $S_n/(\bar{\ell}\sqrt{n})$ is problematic, because S_n^2 is not a reliable and efficient estimator of the true variance of $\bar{\ell}$. More often than not, the pair $\bar{\ell}$ and S_n^2 does not provide any more useful information than the asymptotic approximation $\alpha(\gamma)$, because with very high probability $\breve{\ell}_i = \alpha(\gamma)$ for all $1 = 1, \ldots, n$, and hence $S_n^2 = 0$, resulting in severely underestimating the true variance of $\bar{\ell}$.

In this article we formally prove (Section 2) that the sample variance estimator S_n^2 is inefficient, in the sense that its relative error diverges. This has motivated us to advocate a simple remedy to this problem: we propose estimating or bounding the variance of $\bar{\ell}$ using an estimator different from the inefficient S_n^2. As a consequence, it is now always possible to quantify the accuracy of the estimator $\bar{\ell}$.

In addition to this simple remedy, we also investigate a new sequential importance sampling estimator (Section 3), whose likelihood ratio is a smooth function. The advantage of having a smooth likelihood ratio in the current context is that the corresponding sample variance estimator will deviate from zero.

On the down side, we do not provide a formal proof of the efficiency (or inefficiency, for that matter) of the sample variance of the sequential importance sampling estimator. Nevertheless, numerical experiments (Section 3.3) indicate that the sequential importance sampling estimator provides a reliable error estimate, and can achieve (in representative examples) a variance reduction over $\bar{\ell}$ of the order $\mathcal{O}(10^{12})$.

2 Quantifying the precision

The reason why it is difficult to quantify the precision of $\bar{\ell}$ using S_n^2 is a direct consequence of the following result. It entails that estimating the variance for large γ via S_n^2 is impractical.

Proposition 1 (Inefficiency of Sample Variance of $\breve{\ell}$)**.** *Let S_n^2 be the sample variance based on n independent replications of $\breve{\ell}$. Then,*

$$\liminf_{\gamma\uparrow\infty} \frac{\mathbb{V}ar_m(S_n^2)}{\mathbb{V}ar_m^2(\breve{\ell})} = \infty.$$

Proof: Define by N the number of entries of $\boldsymbol{X}$ larger than γ:

$$N \stackrel{\text{def}}{=} \sum_{i=1}^{d} \mathbb{I}\{X_i > \gamma\},$$

so that $\ell(\gamma) = \mathbb{P}(N \geq 1)$. Next, we define $\beta_{i,j}(\gamma) = \mathbb{P}(X_i > \gamma, X_j > \gamma)$ and $\beta(\gamma) = \sum_{i<j} \beta_{i,j}$ so that the residual $\alpha(\gamma) - \ell(\gamma)$ satisfies

$$r(\gamma) \stackrel{\text{def}}{=} \alpha(\gamma) - \ell(\gamma) = \beta(\gamma) + o\left(\sum_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma)\right).$$

Note that it holds that $\ell(\gamma) = \Theta(\alpha(\gamma))$, which means that both $\ell(\gamma) = \mathcal{O}(\alpha(\gamma))$ and $\alpha(\gamma) = \mathcal{O}(\ell(\gamma))$ are valid. In addition, $r(\gamma) = o(\alpha(\gamma))$; $\mathbb{P}(N > 1) = \Theta(r(\gamma))$; and $\mathbb{P}_m(N = 1) = \mathbb{P}(N = 1)/\alpha(\gamma) = \Theta(1)$.

Using these properties, we obtain by distinguishing between the possible values of $N \in \{1, \ldots, d\}$, that the k-th centered moment (for any $k \geq 1$) of $\breve{\ell}$ can be written as

$$\begin{aligned}
\mathbb{E}_m \left|\breve{\ell} - \ell(\gamma)\right|^k &= \sum_{j=1}^{d} \mathbb{E}_m \left[\left|\breve{\ell} - \ell(\gamma)\right|^k \mathbb{I}\{N = j\}\right] \\
&= |\alpha(\gamma) - \ell(\gamma)|^k \mathbb{P}_m(N = 1) + \sum_{j=2}^{d} \left|\frac{\alpha(\gamma)}{j} - \ell(\gamma)\right|^k \mathbb{P}_m(N = j) \\
&= r^k(\gamma)\mathbb{P}_m(N = 1) + \Theta(\alpha^k)\mathbb{P}_m(N > 1) \\
&= r^k(\gamma)\mathbb{P}_m(N = 1) + \Theta(\alpha^{k-1})\mathbb{P}(N > 1) \\
&= \Theta\left(r^k(\gamma)\right) + \Theta\left(\alpha^{k-1} r(\gamma)\right).
\end{aligned}$$

Therefore, for the relative error of S_n^2 given in equation 9 of L'Ecuyer et al. (2010) we obtain (abbreviating for the moment $r = r(\gamma)$)

$$\begin{aligned}
\frac{n\mathbb{V}ar_m(S_n^2)}{\mathbb{V}ar_m^2(\breve{\ell})} &= \frac{\mathbb{E}_m(\breve{\ell} - \ell(\gamma))^4}{[\mathbb{E}_m(\breve{\ell} - \ell(\gamma))^2]^2} - 1 + \frac{2}{n-1} \\
&= \frac{\Theta(r^4) + \Theta(\alpha^3 r)}{\Theta(r^4) + \Theta(\alpha r^3) + \Theta(\alpha^2 r^2)} + \mathcal{O}(1)
\end{aligned}$$

The first term diverges, because $r(\gamma) = o(\alpha(\gamma))$ implies that the denominator converges to zero faster than the numerator; in fact, the relative error grows at the rate α/r. $\square$

Even though the relative error diverges, one can still hope that the growth would be at a very slow rate. Unfortunately, this is not the case. Assume, for example, that all pairs (X_i, X_j) are jointly normal with zero mean, $\mathbb{V}ar(X_i) = \mathbb{V}ar(X_j) = 1$, and $\mathbb{C}or(X_i, X_j) = \rho_{i,j} < 1$, then using the results in Hashorva and Hüsler (2003) we can easily show that, for ease writing $\rho = \rho_{i,j}$,

$$\beta_{i,j} = \mathbb{P}(X_i > \gamma, X_j > \gamma) \simeq \frac{(1+\rho)^2}{2\pi\sqrt{1-\rho^2}\,\gamma^2} \exp\left(-\frac{\gamma^2}{2}\frac{2}{1+\rho}\right), \qquad \gamma \uparrow \infty .$$

In other words,

$$\frac{\beta_{i,j}}{\alpha_j} = \mathbb{P}(X_i > \gamma \,|\, X_j > \gamma) \simeq \frac{(1+\rho)^2}{\sqrt{2\pi(1-\rho^2)}\,\gamma} \exp\left(-\frac{\gamma^2}{2}\frac{1-\rho}{1+\rho}\right), \tag{1}$$

and from $r = \Theta(\sum_{i<j} \beta_{i,j})$ we conclude that the relative error of S_n^2 grows at the exponential rate $c_1\gamma\exp(c_2\gamma^2)$ for some strictly positive constants c_1, c_2.

Example 1 (Positive Correlation)**.** To illustrate the failure of S_n^2 as an estimator, consider the case with $\boldsymbol{a} = \mathbf{0}$ with covariance matrix with entries $\Sigma_{i,j} = \exp(-|i-j|)$ for $d = 10^3$. Table 1 below shows that S_n^2 is only nonzero when γ is small, and that it fails as an estimator for $\gamma \geq 5.5$. In fact, for the precision considered in this simulation, the estimator $\bar{\ell} = \alpha$ for $\gamma \geq 5.5$.

Table 1: Estimates of ℓ, using $n = 10^3$ and $d = 10^3$.

γ	$\alpha(\gamma)$	$\bar{\ell}$	$S_n/(\bar{\ell}\sqrt{n})$	κ
3	0.73	0.0165	0.036	0.036
4	0.0316	0.0311	0.0029	0.0046
5	0.0002866	0.0002865	0.00050	0.00078
5.5	1.899×10^{-5}	1.899×10^{-5}	0	0.00037
6	9.8659×10^{-7}	9.8659×10^{-7}	0	0.00018
6.5	4.0160×10^{-8}	4.0160×10^{-8}	0	8.2×10^{-5}
7	1.2798×10^{-9}	1.2798×10^{-9}	0	3.6×10^{-5}
7.5	3.1908×10^{-11}	3.1908×10^{-11}	0	1.5×10^{-5}
8	6.22096×10^{-13}	6.22096×10^{-13}	0	6.0×10^{-6}
9	1.128588×10^{-16}	1.128588×10^{-16}	0	8.0×10^{-7}
10	$7.6198530 \times 10^{-21}$	$7.6198530 \times 10^{-21}$	0	8.5×10^{-8}

Note that in this example the covariance matrix has positive correlation structure, which is beneficial for the performance of the estimator S_n^2, as seen from (1). For a negative correlation matrix, the performance of S_n^2 is even worse. The last column shows κ, which is an alternative estimator of the true relative error, which we introduce below. $\square$

Given the problem with quantifying the error of the estimator $\bar{\ell}$, a natural way to proceed is to attempt to modify $\breve{\ell}$ so that we *do* observe some variability during the course of the simulation. One such idea is the *undershooting adaptation* investigated by Blanchet and Li (2011) for completely different reasons (relating to the discretization of Gaussian random fields), but which may be helpful in our context as well. The idea is to sample from the modified importance sampling density which undershoots by ϵ, say $\epsilon = 1/\gamma$:

$$m_\epsilon(\boldsymbol{x}) = \frac{\phi(\boldsymbol{x}; \boldsymbol{a}, \Sigma) \sum_{i=1}^d \mathbb{I}\{x_i > \gamma - \epsilon\}}{\sum_{i=1}^d \mathbb{P}(X_i > \gamma - \epsilon)} = \frac{\phi(\boldsymbol{x}; \boldsymbol{a}, \Sigma) \sum_{i=1}^d \mathbb{I}\{x_i > \gamma - \epsilon\}}{\alpha_\epsilon(\gamma)}$$

and then use the importance sampling estimator

$$\breve{\ell}_\epsilon = \alpha_\epsilon(\gamma) \frac{\mathbb{I}\{\max_i X_i > \gamma\}}{\sum_{i=1}^d \mathbb{I}\{X_i > \gamma - \epsilon\}}.$$

With a suitable choice of ϵ, this modification introduces some variability in $\breve{\ell}_\epsilon$, but at a significant loss of efficiency. For example, for $\gamma = 10$ in Table 1 one can at most obtain an accuracy of one significant figure for the same simulation effort. Observe, however, that it does not seem logical to use an expensive Monte Carlo estimator to obtain a mediocre accuracy of one or two significant figures when the simple asymptotic approximation α is already accurate at least to seven significant figures (where we again refer to the κ column in Table 1). Summarizing, the estimator $\breve{\ell}$ should not be tampered with, because it already provides a nice transition from a Monte Carlo estimate (for small γ) to a very accurate deterministic estimate α (for large γ). The estimator is efficient, but the problem is that we simply do not know how efficient it is. The remedy we propose in this paper, is that we advise to use *an upper bound* on the relative error.

To find how efficient $\breve{\ell}$ is, one can consider the following upper bound on the relative error (writing for ease ℓ for $\ell(\gamma)$):

$$\begin{aligned}\frac{n\mathbb{V}ar(\bar{\ell})}{\ell^2} &= \frac{\alpha(\gamma)}{\ell^2}\mathbb{E}\left[\frac{\mathbb{I}\{N \geq 1\}}{N}\right] - 1 \\ &\leq \frac{\alpha(\gamma)}{\ell^2}\mathbb{P}(N \geq 1) - 1 \\ &\leq \frac{\alpha(\gamma)}{\ell} - 1 = \frac{\alpha(\gamma) - \ell}{\ell(\gamma)} \leq \frac{\beta(\gamma)}{\ell},\end{aligned}$$

where $\beta(\gamma) \stackrel{\text{def}}{=} \sum_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma)$. Thus, if we knew $\beta(\gamma)$, we can use $\kappa \stackrel{\text{def}}{=} \sqrt{\beta(\gamma)/(\bar{\ell}n)}$ as our estimate for the relative error of $\bar{\ell}$. This is the value displayed in the last column of Table 1.

Note that there are deterministic quadrature algorithms for the computation of each β_{ij}; see, for example, Drezner and Wesolowsky (1990). Thus, in principle, $\beta(\gamma)$ is computable in $\mathcal{O}(d^2)$ time, which is acceptable since the simulation of $\boldsymbol{X}$ requires $\mathcal{O}(d^3)$ time — the complexity of the Cholesky decomposition of the matrix Σ.

3 Sequential importance sampling estimator

As mentioned in the previous section, there is no point in considering an alternative to the already quite good $\breve{\ell}$, unless the new estimator is at least as efficient as $\breve{\ell}$ and provides an error estimate. In this section we provide one such alternative estimator. The new estimator also enjoys vanishing relative error, and in all numerical experiments that we performed it turned out to be more accurate than $\breve{\ell}$. The proposed estimator is based on the following splitting of the event into d components Kroese et al. (2011, Page 396):

$$\ell(\gamma) = \mathbb{P}\left(\max_{i\in\{1,\ldots,d\}} X_i > \gamma\right) = \sum_{i=1}^{d} \mathbb{P}\left(X_i > \gamma, X_i \geq \max_{k\neq i} X_k\right) = \sum_{i=1}^{d} \mathbb{P}\left(AP_i\, \boldsymbol{X} \geq \boldsymbol{l}\right),$$

where P_i is a permutation matrix that swaps the i-th entry with the first entry, and

$$A \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 & -1 \end{bmatrix}, \qquad \boldsymbol{l} \stackrel{\text{def}}{=} \begin{pmatrix} \gamma \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2}$$

Since $AP_i(\boldsymbol{X} - \boldsymbol{a}) \sim \mathsf{N}(\boldsymbol{0}, AP_i\Sigma P_i^\top A^\top)$ we can write

$$\mathbb{P}(AP_i\, \boldsymbol{X} \geq \boldsymbol{l}) = \mathbb{P}(L_i\boldsymbol{Z} \geq \boldsymbol{l}_i), \quad \boldsymbol{Z} \sim \mathsf{N}(\boldsymbol{0}, I),$$

where $\boldsymbol{l}_i = \boldsymbol{l} - AP_i\, \boldsymbol{a}$; the matrix L_i is the L factor in the LQ matrix decomposition (Golub and Van Loan, 2012) of the covariance

$$L_iL_i^\top = AP_i\Sigma P_i^\top A^\top$$

We insist that L_i has nonnegative entries down the main diagonal so that, whenever L_i is full rank, it coincides with the Cholesky factor of $L_i L_i^\top$.

We propose to estimate the probabilities $p_i = \mathbb{P}(L_i \boldsymbol{Z} \geq \boldsymbol{l}_i)$ separately. Suppose $\hat{p}_i$ is an unbiased Monte Carlo estimator of p_i constructed from a sample size of n_i, and with variance $\mathbb{V}ar(\hat{p}_i) = \sigma_i^2/n_i$. Our goal now is to estimate ℓ via $\hat{\ell} = \sum_i \hat{p}_i$, using a total computational budget of $n = \sum_i n_i$. We choose

$$n_i \stackrel{\text{def}}{=} \left\lceil n \times \frac{\alpha_i(\gamma)}{\alpha(\gamma)} \right\rceil \wedge 3,$$

that is, $n_i \propto \alpha_i$, for reasons that will be discussed later. All that remains to make $\hat{\ell}$ a viable estimator is to explain how we estimate a generic probability of the form $p = \mathbb{P}(L\,\boldsymbol{Z} \geq \boldsymbol{l})$, where L is lower triangular.

We use the sequential Monte Carlo method of Genz and Bretz (2009). The method relies on the observation that the set $\{\boldsymbol{z} : L\,\boldsymbol{z} \geq \boldsymbol{l}\}$ can be written as the intersection of

$$z_1 \geq \tilde{l}_1 \stackrel{\text{def}}{=} \frac{l_1}{L_{11}}, \quad z_2 \geq \tilde{l}_2 \stackrel{\text{def}}{=} \frac{l_2 - L_{21} z_1}{L_{22}}, \ \cdots, \ z_d \geq \tilde{l}_d \stackrel{\text{def}}{=} \frac{l_d - \sum_{j=1}^{d-1} L_{dj} z_j}{L_{dd}}$$

This decomposition suggests the sequential importance sampling pdf:

$$g(\boldsymbol{z}) = g(z_1) g(z_2 \,|\, z_1) \cdots g(z_{d-1} \,|\, z_d)$$

with

$$g(z_i \,|\, z_1, \ldots, z_{i-1}) = \frac{\phi(z_i; 0, 1)\mathbb{I}\{z_i > \tilde{l}_i\}}{\overline{\Phi}(\tilde{l}_i)},$$

where $\overline{\Phi}(\cdot)$ is, as before, the tail distribution of the standard normal. Thus, conditional on all preceding variables, each Z_i is drawn from a truncated normal density. The resulting importance sampling estimator of $p = \mathbb{P}(L\boldsymbol{Z} \geq \boldsymbol{l})$ based on a single simulation is

$$V = \frac{\phi(\boldsymbol{Z}; \boldsymbol{0}, I)}{g(\boldsymbol{Z})} = \exp\left(\sum_{i=1}^{d} \log \overline{\Phi}\left(\frac{l_i - \sum_{j=1}^{i-1} L_{ij} Z_j}{L_{ii}} \right) \right), \qquad \boldsymbol{Z} \sim g(\boldsymbol{z}) \tag{3}$$

We can now summarize the main ingredients of the proposed algorithm.

Algorithm 1 Estimating $\ell(\gamma) = \mathbb{P}(\max_{i \in \{1,\ldots,d\}} X_i \geq \gamma)$ with $\boldsymbol{X} \sim \mathsf{N}(\boldsymbol{a}, \Sigma)$.

Require: parameter γ; $d \times d$ covariance Σ; mean $\boldsymbol{a}$; sample sizes $\{n_i\}$ such that $\sum_{i=1}^{d} n_i = n$.

for $i = 1, \ldots, d$ **do**

  Let P_i be a permutation matrix swapping the i-th element with the first element.

  $\boldsymbol{l}_i \leftarrow \boldsymbol{l} - AP_i\,\boldsymbol{a}$, where $A, \boldsymbol{l}$ are defined in (2).

  Compute the lower triangular factor L_i such that $L_i L_i^\top = AP_i \Sigma P_i^\top A^\top$.

  Simulate n_i independent replications of (3), $V_1, \ldots, V_{n_i}$.

  $\hat{p}_i \leftarrow \frac{1}{n_i} \sum_j V_j$, so that $\mathbb{E}[\hat{p}_i] = \mathbb{P}(L_i\,\boldsymbol{Z} \geq \boldsymbol{l}_i)$

  $\hat{\sigma}_i^2 \leftarrow \frac{1}{n_i} \sum_j (V_j - \hat{p}_i)^2$

$\hat{\ell} \leftarrow \sum_i \hat{p}_i$

return $\hat{\ell}$ and its estimated relative error $\sqrt{\sum_i \hat{\sigma}_i^2/n_i} \Big/ \hat{\ell}$.

In the following two subsections we comment on two issues: (1) the efficiency of the newly constructed estimator $\hat{\ell}$ as $\gamma \uparrow \infty$, and the concomitant choice $n_i \propto \alpha_i/\alpha$, and (2) the efficient computation of the factors $L_1, \ldots, L_d$.

3.1 Efficiency and Strata Sample Sizes

First note that $p_i \simeq \alpha_i(\gamma)$, in the sense that

$$\begin{aligned}\alpha_i(\gamma) \geq p_i &= \alpha_i(\gamma) - \alpha_i(\gamma)\,\mathbb{P}\left(\max_{k\neq i} X_k > X_i \,|\, X_i > \gamma\right)\\ &\geq \alpha_i(\gamma) - \alpha_i(\gamma)\,\mathbb{P}\left(\max_{k\neq i} X_k > \gamma \,|\, X_i > \gamma\right) = \alpha_i(\gamma) \times (1 - o(1)).\end{aligned}$$

Next, observe that (3) can be rearranged as

$$\frac{V}{\mathbb{P}(L_{11}Z_1 > \gamma)} = \prod_{i=2}^{d} \overline{\Phi}\left(\frac{l_i - \sum_{j=1}^{i-1} L_{ij}Z_j}{L_{ii}}\right) \leq 1.$$

In other words, all the $\hat{p}_i$ s can be written as $\hat{p}_i = \alpha_i(\gamma)\,\hat{q}_i$ where $\hat{q}_i \leq 1$ is an unbiased estimator of the conditional probability $\mathbb{P}(\max_{k\neq i} X_k < X_i \,|\, X_i > \gamma)$. Therefore,

$$n_i \mathbb{V}ar(\hat{p}_i) = \sigma_i^2 \leq \alpha_i(\gamma)\,p_i - p_i^2$$

Then, $\sum_i p_i = \ell(\gamma) \simeq \alpha(\gamma)$ and $p_i \simeq \alpha_i(\gamma)$ in combination with $n_i \simeq n \times \alpha_i(\gamma)/\alpha(\gamma)$ imply a vanishing relative error estimator:

$$\begin{aligned}\frac{n\mathbb{V}ar(\hat{\ell})}{\ell^2} &= \frac{n\sum_i \sigma_i^2/n_i}{\ell^2}\\ &\simeq \frac{\alpha(\gamma)\,\sum_i \sigma_i^2/\alpha_i(\gamma)}{\ell^2}\\ &\leq \frac{\alpha(\gamma)\sum_i p_i}{\ell^2} - \frac{\alpha(\gamma)\sum_i p_i^2/\alpha_i(\gamma)}{\ell^2} = o(1).\end{aligned}$$

3.2 Computation of $L_1, L_2, \ldots, L_d$ in $\mathcal{O}(d^3)$ Time

At first examination, it would seem that in Step 4 of Algorithm 1 we need to compute each lower triangular factor L_i from scratch at a cost of $\mathcal{O}(d^3)$, and since this step is inside a loop of length d, the running time of the algorithm would be $\mathcal{O}(d^4)$. But in fact, the computation of all the d lower triangular factors can still be accomplished in $\mathcal{O}(d^3)$ time, as we point out now. Let

$$C_1\,C_1^\top = P_1\,\Sigma\,P_1 = \Sigma$$

be the standard Cholesky decomposition of the unpermuted covariance Σ. Set $Q_1 = I$, so that C_1Q_1 is the LQ decomposition of C_1 (or equivalently $Q_1^\top C_1^\top$ is the QR decomposition of $C_1^\top$). The computation of C_1 takes $\mathcal{O}(d^3)$ operations.

Given C_1 we wish to compute the Cholesky decomposition $C_iC_i^\top$ of the permuted covariance matrix

$$P_i\,\Sigma\,P_i^\top = (P_i\,C_1)(P_i\,C_1)^\top$$

in $\mathcal{O}(d^2)$ operations. To this end, let $\boldsymbol{e}_i$ denote the unit column vector with one in the i-th position and $\boldsymbol{c}_i$ denote the i-th row of C_1. Then, we can express $P_i\, C_1$ as a rank one perturbation of matrix C_1:

$$P_i\, C_1 = C_1 + (\boldsymbol{e}_1 - \boldsymbol{e}_i)(\boldsymbol{c}_i^\top - \boldsymbol{c}_1^\top)$$

It is well known, see for example (Golub and Van Loan, 2012, Page 593), that given the LQ factors of C_1, we can obtain the LQ factors, C_i and Q_i, of the rank-1 perturbed matrix

$$C_1 + (\boldsymbol{e}_1 - \boldsymbol{e}_i)(\boldsymbol{c}_i^\top - \boldsymbol{c}_1^\top)$$

in $\mathcal{O}(d^2)$ time. Hence, the entire list $C_1, \ldots, C_d$ is computable in $\mathcal{O}(d^3)$ time. From this list we can easily obtain the factors L_i as follows.

Observe that $A\, C_i$ is lower triangular, because A is lower triangular. Set $\boldsymbol{u}_i = \mathrm{diag}(A\, C_i)$ to be the vector containing the diagonal elements of matrix $A\, C_i$ and let $D_i = \mathrm{diag}(\mathrm{sign}(\boldsymbol{u}_i))$ be a diagonal matrix with diagonal corresponding to the signs of the elements of $\boldsymbol{u}_i$. Then,

$$L_i = A\, C_i\, D_i$$

is the desired factor of $P_i \Sigma P_i^\top$, because L_i is lower triangular by construction, and has nonnegative diagonal entries. Note that, whenever L_i is full rank, it coincides with the Cholesky factor of $P_i\, \Sigma\, P_i^\top$.

3.3 A Numerical Study

In this subsection we discuss two numerical examples.

Example 2 (Negative Correlation)**.** We consider estimating ℓ with $\boldsymbol{a} = 2 \times \mathbf{1}$, $d = 100$, and precision matrix (Fernández et al., 2007):

$$\Sigma^{-1} = \frac{1}{2} I + \frac{1}{2} \mathbf{1}\mathbf{1}^\top$$

Table 2 below shows the estimate $\hat{\ell}$ and the relative error estimator for both $\hat{\ell}$ and $\bar{\ell}$ (including the new estimate κ). The last column shows the factor by which the variance is reduced using the sequential importance sampling estimator, namely, the ratio $\widehat{\mathbb{V}ar}(\bar{\ell})/\widehat{\mathbb{V}ar}(\hat{\ell})$. $\square$

Table 2: Estimates of ℓ using $n = 10^5$ for both methods.

γ	$\hat{\ell}$	rel. err. $\hat{\ell}$	$S_n/(\bar{\ell}\sqrt{n})$	κ	variance gain
6	0.202	2.2×10^{-4}	7×10^{-4}	1.0×10^{-3}	3^2
7	1.887×10^{-2}	2.0×10^{-5}	2.0×10^{-4}	2.8×10^{-4}	10^2
8	1.0044×10^{-3}	1.2×10^{-6}	4.2×10^{-5}	6.3×10^{-5}	35^2
8.5	1.92672×10^{-4}	2.4×10^{-7}	1.8×10^{-5}	2.7×10^{-5}	80^2
8.8	6.74778×10^{-5}	1.2×10^{-7}	1.2×10^{-5}	1.6×10^{-5}	100^2
9	3.271876×10^{-5}	5.4×10^{-8}	5×10^{-6}	1.1×10^{-5}	92^2
9.5	4.917356×10^{-6}	9.0×10^{-9}	7×10^{-6}	4.2×10^{-6}	460^2
10	6.5380233×10^{-7}	1.0×10^{-9}	0	1.5×10^{-6}	1500^2
11	$7.990587315 \times 10^{-9}$	1.3×10^{-11}	0	1.6×10^{-7}	10^8
12	$5.95823969666 \times 10^{-11}$	1.5×10^{-13}	0	1.3×10^{-8}	10^9
13	$2.70552710714302 \times 10^{-13}$	6.1×10^{-16}	0	8.4×10^{-10}	10^{12}

Example 3 (Random Correlation Matrices)**.** We next compare the two estimators by using a large scale simulation with randomly generated test correlation matrices. The random correlation matrices are simulated via the method of Davies and Higham (2000), whereby the eigenvalues $\{\lambda_i\}$ of each correlation matrix are uniformly distributed over the simplex $\{\boldsymbol{\lambda} : \sum_i \lambda_i = d,\ \lambda_i > 0\}$.

Thus, the experiments consist of the following. For a given value of γ and $d = 100$, we simulate 100 independent realizations of $\boldsymbol{a} \sim \mathsf{U}(0,1)^{100}$ (uniformly distributed in the unit hypercube) and Σ (simulated according to the Davies and Higham (2000) generator). In each of these 100 experiments we compute $\hat{\ell}$ and $\bar{\ell}$ (estimators of $\ell(\gamma),\ \boldsymbol{X} \sim \mathsf{N}(\boldsymbol{a}, \Sigma)$) and their respective relative errors using a sample size of $n = 10^4$.

The boxplots on Figure 1 show the empirical distribution of the relative errors of $\hat{\ell}$ and $\bar{\ell}$ (using $n = 10^4$) for different values of γ. Each boxplot is built based on the 100 independent replications. The boxplots corresponding to $\bar{\ell}$ are all in the upper left part of the graph (in black). The boxplots corresponding to $\hat{\ell}$ are always in lower position (in blue).

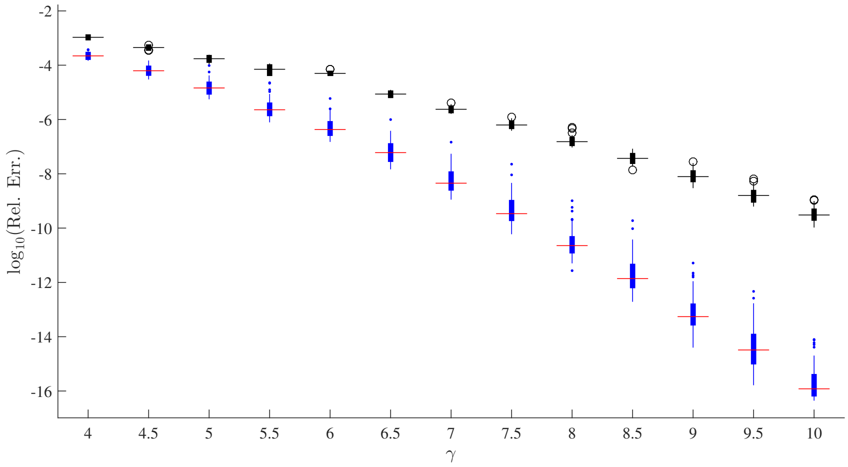

Figure 1: Distribution of relative errors of $\hat{\ell}$ and $\bar{\ell}$ for different values of γ.

We observe that the relative error of $\hat{\ell}$ is always lower compared to that of $\bar{\ell}$, and that the difference in computational cost between the algorithms is negligible. In addition, the sample variance S_n^2 works only up until $\gamma \leq 6$, after which it yields the meaningless estimate of 0. In fact, the boxplot for $\gamma = 6$ is not fully formed, because most of the 100 independent simulations did not yield a meaningful variance estimate. The rest of the boxplots for $\gamma > 6$ are computed from the distribution of our new error estimate κ. $\square$

4 Concluding remarks

We considered the estimation of the tail distribution of the maximum of correlated normal random variables. We showed that the sample variance of an vanishing relative error estimator is not suitable for assessing its Monte Carlo variance. As a simple remedy, we propose an alternative estimate of the error. In addition, we consider a new sequential Monte Carlo estimator of the tail probability, which turns out to be more accurate in all of the examples we considered.

In future research we hope to achieve even better accuracy by exploiting the representation

$$\ell(\gamma) = \alpha(\gamma) - \sum_i \alpha_i(\gamma)\mathbb{P}\left(\max_{k>i} X_k > \gamma \,|\, X_i > \gamma\right),$$

and applying the mixture importance sampling idea of Adler and Taylor (2009) to the estimation of the terms $\mathbb{P}(\max_{k>i} X_k > \gamma \,|\, X_i)$, conditional on having simulated an $X_i > \gamma$. This approach seems promising in that one may be able to prove a bounded relative error property for the sample variance estimator.

Acknowledgment

Zdravko Botev has been supported by the *Australian Research Council* grant DE140100993. M. Mandjes' research is partly funded by the NWO Gravitation project NETWORKS, grant number 024.002.003.

References

Adler, R. J., J. H. Blanchet, and J. Liu (2008). Efficient simulation for tail probabilities of gaussian random fields. In *Proceedings of the 40th Conference on Winter Simulation*, pp. 328–336. Winter Simulation Conference.

Adler, R. J., J. H. Blanchet, and J. Liu (2012). Efficient monte carlo for high excursions of gaussian random fields. *The Annals of Applied Probability 22*(3), 1167–1214.

Adler, R. J. and J. E. Taylor (2009). *Random fields and geometry*, Volume 115. Springer.

Blanchet, J. H. and C. Li (2011). Efficient simulation for the maximum of infinite horizon discrete-time gaussian processes. *Journal of Applied Probability 48*(2), 467–489.

Davies, P. I. and N. J. Higham (2000). Numerically stable generation of correlation matrices and their factors. *BIT Numerical Mathematics 40*(4), 640–651.

Drezner, Z. and G. O. Wesolowsky (1990). On the computation of the bivariate normal integral. *Journal of Statistical Computation and Simulation 35*(1-2), 101–107.

Fernández, P. J., P. A. Ferrari, and S. P. Grynberg (2007). Perfectly random sampling of truncated multinormal distributions. *Advances in Applied Probability 39*, 973–990.

Genz, A. and F. Bretz (2009). *Computation of multivariate normal and t probabilities*, Volume 195. Springer.

Golub, G. H. and C. F. Van Loan (2012). *Matrix computations*, Volume 3. JHU Press.

Hashorva, E. and J. Hüsler (2003). On multivariate gaussian tails. *Annals of the Institute of Statistical Mathematics 55*(3), 507–522.

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

L'Ecuyer, P., J. H. Blanchet, B. Tuffin, and P. W. Glynn (2010). Asymptotic robustness of estimators in rare-event simulation. *ACM Transactions on Modeling and Computer Simulation (TOMACS) 20*(1), 6.

Mandjes, M. R. H. (2007). *Gaussian Queues*. Wiley.