Communications for Statistical Applications and Methods
2024, Vol. 31, No. 1, 65–85

https://doi.org/10.29220/CSAM.2024.31.1.065
Print ISSN 2287-7843 / Online ISSN 2383-4757

An importance sampling for a function of a multivariate random variable

Jae-Yeol Park[a], Hee-Geon Kang[a], Sunggon Kim[1,a]

[a]Department of Statistics, University of Seoul, Korea

Abstract

The tail probability of a function of a multivariate random variable is not easy to estimate by the crude Monte Carlo simulation. When the occurrence of the function value over a threshold is rare, the accurate estimation of the corresponding probability requires a huge number of samples. When the explicit form of the cumulative distribution function of each component of the variable is known, the inverse transform likelihood ratio method is directly applicable scheme to estimate the tail probability efficiently. The method is a type of the importance sampling and its efficiency depends on the selection of the importance sampling distribution. When the cumulative distribution of the multivariate random variable is represented by a copula and its marginal distributions, we develop an iterative algorithm to find the optimal importance sampling distribution, and show the convergence of the algorithm. The performance of the proposed scheme is compared with the crude Monte Carlo simulation numerically.

Keywords: importance sampling, copula, inverse transform likelihood ratio method, Monte Carlo simulation, rare event simulation

1. Introduction

The estimation of rare-event probabilities has received great attention due to its relevance in various fields such as financial and insurance risk management, inventory management, structural systems reliability, computer networks, and telecommunications networks. The estimation of such probabilities often poses challenges when the events of interest are located in the tails of probability distributions, where the crude Monte Carlo simulation exhibits inefficiency. When the occurrence of the events of interest is rare, the accurate estimation of the tail probability requires a huge number of samples and it takes a very long time to achieve satisfactory results through the crude Monte Carlo simulation. To address this issue, importance sampling methods have emerged as a powerful tool, enabling accurate estimation of the tail probability (Glynn and Iglehart, 1989; Juneja and Shahabuddin, 2006). By sampling more frequently on the regions of interest, the importance sampling can significantly reduce the variance in the estimation of the tail probability with a smaller number of samples.

Kroese and Rubinstein (2004) proposed an importance sampling method called the transform likelihood ratio method for rare event simulation. In the method, the multivariate random variable (or the random vector) of interest is transformed into another multivariate random variable in order to find a more efficient importance sampling distribution. If the random variable of interest is one dimensional and the transformed random variable is uniformly distributed on $(0, 1)$, then it is called the inverse

[1]Corresponding author: Department of Statistics, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Korea. E-mail: sgkim@uos.ac.kr

Published 31 January 2024 / journal homepage: http://csam.or.kr

transform likelihood ratio method. When the form of the cumulative distribution function of the random variable is known, then the inverse transform likelihood ratio method is a simple and unifying way of generating the importance samples efficiently in the region of interest. In the tail probability estimation of a monotone increasing function of a one dimensional random variable, Kroese and Rubinstein (2004) showed that the inverse transform likelihood ratio method gives an estimator with bounded relative error.

The efficiency of an importance sampling method depends on the selection of the importance sampling distribution. In the case of light-tailed distributions, the exponential twisting is the common approach to find the efficient importance sampling distribution. It gives the optimal importance sampling estimator in some cases (Sadowsky and Bucklew, 1990; Sadowsky, 1993). In the case of subexponential distributions such as the Weibull or the lognormal, the exponential twisting is not applicable due to the non-existence of the moment generating function. For the tail probability estimation of the random or deterministic sums of independent and identically distributed (i.i.d.) subexponential random variables, Asmussen *et al.* (2000) proposed a number of estimators, which are asymptotically optimal. For the same problem, Juneja and Shahabuddin (2002) proposed a method called the hazard rate twisting, and showed the asymptotic optimality of the method. The performance of the estimators by Asmussen *et al.* (2000) and Juneja and Shahabuddin (2002) has been improved by the conditional Monte Carlo estimator proposed by Asmussen and Kroese (2006). For the case of sums of i.i.d. random variables with regularly varying tails, Dupuis *et al.* (2007) proposed an algorithm for creating importance sampling estimator with bounded relative error. When the random variables are independent but not identically distributed, Rached *et al.* (2015) and Rached *et al.* (2018) proposed a generalized hazard rate twisting for the tail probability estimation.

For the case of the sums of correlated lognormal random variables, Asmussen *et al.* (2011) proposed two importance sampling estimators, one of which is asymptotically efficient, and the other of which shows bounded relative error. This method has been extended to apply to more general cases by Blanchet and Rojas-Nandayapa (2011). They proposed an asymptotically optimal importance sampling method for the estimation of the tail probability of $e^{X_1} + \cdots + e^{X_d}$, where $(X_1, \ldots, X_d)$ is a d-variate random variable. In the case that $(X_1, \ldots, X_d)$ follows the d-dimensional multivariate normal distribution, $e^{X_1} + \cdots + e^{X_d}$ is the sum of the correlated lognormal random variables.

By confining the importance sampling distribution to a parametric family of distributions, we can apply analytical methods to find the optimal importance sampling distribution from the family. The cross entropy method and the variance minimization method are generally applied to find the optimal parameter. In the latter method the parameter minimizing the variance of the importance sampling estimator is found to be the optimal (Rubinstein and Shapiro, 1993). In the former method which was proposed by Rubinstein (2002) and Homem-de-Mello and Rubinstein (2002), the optimal parameter to be found is the one minimizing the Kullback-Leibler divergence of the importance sampling distribution from the zero variance distribution, which is defined in Section 2.3. In both methods, the optimal parameter is usually found by solving a series of stochastic optimization problems iteratively; see De Boer *et al.* (2005) for more details.

In this paper, we propose an inverse transform likelihood ratio method for the tail probability estimation of an increasing function of a multivariate random variable. This method was proposed originally for the rare event simulations with one-dimensional random variables (Kroese and Rubinstein, 2004). Due to the Sklar's theorem (Nelsen, 2006), the cdf (cumulative distribution function) of a multivariate random variable is represented by its marginal distributions and a copula. Note that a copula is the multivariate uniform distribution with a certain dependence structure. The cdf of the multivariate random variable of interest in this form is assumed to be known. We transform the distribution

of the multivariate random variable into a copula, and confine importance sampling distributions to a family of multivariate distributions in which the marginal distribution of each component follows the beta distribution and the dependence of a multivariate distribution is modeled by the Gaussian copula. By decomposing the Kullback-Leibler divergence of an importance sampling pdf from the zero variance pdf into two parts and minimizing each part in order, we find the pseudo-optimal parameter in an iterative manner. We also show the convergence of the proposed scheme.

The rest of the paper is organized as follows. The problem statement is given in Section 2. In the same section, we introduce the inverse transform likelihood method and the cross entropy method briefly, and describe how to apply these methods to our problem. We propose an algorithm for finding the optimal parameter of the importance sampling distribution. In Section 3, we show the convergence of the proposed algorithm. The performance of the proposed scheme is compared numerically with the crude Monte Carlo simulation in Section 4. Finally, we conclude the paper in Section 5.

2. Method

2.1. Problem statement

We consider a d-dimensional continuous random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with support $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d)$, where $-\infty \leq a_i < b_i \leq \infty$, $i = 1, \ldots, d$. We denote by F the joint cdf of $\boldsymbol{X}$, and by F_i the marginal cdf of X_i, $i = 1, \ldots d$. We also let f be the pdf (probability density function) of $\boldsymbol{X}$. For a real valued function L defined on Ω, we want to estimate the tail probability of $L(\boldsymbol{X})$ over threshold γ, i.e.

$$l = \Pr\{L(\boldsymbol{X}) > \gamma\}. \tag{2.1}$$

We call $L(\boldsymbol{X})$ the loss, and $L(\cdot)$ the loss function. We assume that $L(\boldsymbol{x})$ is strictly increasing and continuous on Ω. For $\boldsymbol{x} = (x_1, \ldots, x_d)$, $\boldsymbol{x}' = (x'_1, \ldots, x'_d) \in \mathbb{R}^d$, we define that $\boldsymbol{x} < \boldsymbol{x}'$ if and only if $x_i \leq x'_i$, $\forall i$ and $\boldsymbol{x} \neq \boldsymbol{x}'$. Then, we have that

$$L(\boldsymbol{x}) < L(\boldsymbol{x}'), \quad \boldsymbol{x} < \boldsymbol{x}'.$$

We also assume that the support of $L(\boldsymbol{X})$ is (s_L, ∞), where $-\infty \leq s_L < \infty$, and that for a sequence $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \in \Omega$ such that $\max\{F_1(x_1^{(i)}), \ldots, F_d(x_d^{(i)})\} \to 1$ as $i \to \infty$,

$$\lim_{i \to \infty} L\left(\boldsymbol{x}^{(i)}\right) = \infty. \tag{2.2}$$

We consider the case that the value of γ is so large that the tail loss probability l is near 0.

Sklar's theorem (Nelsen, 2006) says that there is a copula C_0 satisfying

$$F(x_1, \ldots, x_d) = C_0(F_1(x_1), \ldots, F_d(x_d)), \quad (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{2.3}$$

We denote by c_0 the pdf of C_0, i.e.

$$c_0(\boldsymbol{u}) = \frac{\partial^n C_0(\boldsymbol{u})}{\partial u_1 \cdots \partial u_d}, \quad \boldsymbol{u} = (u_1, \ldots, u_d) \in (0, 1)^d.$$

We call F the nominal cdf of $\boldsymbol{X}$, and C_0 the nominal copula, respectively. Suppose that $\boldsymbol{U}$ is a d-dimensional random vector on $(0, 1)^d$, and that the joint cdf of $\boldsymbol{U}$ is C_0. Let U_i, $i = 1, \ldots, d$, be the i^{th} element of $\boldsymbol{U}$. Then, U_i is uniformly distributed on $(0, 1)$. We denote by $\tilde{\boldsymbol{X}} = (F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d))$. Then the cdf of $\tilde{\boldsymbol{X}}$ is F, i.e.

$$\tilde{\boldsymbol{X}} =_d \boldsymbol{X}, \tag{2.4}$$

where $=_d$ means the same in distribution.

We define a function $\tilde{L}(\boldsymbol{u})$ on $(0,1)^d$ as follows:

$$\tilde{L}(\boldsymbol{u}) = L\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right), \quad \boldsymbol{u} \in (0,1)^d. \tag{2.5}$$

Since $L(\boldsymbol{x})$ is continuous on Ω and $F_i^{-1}(u)$, $i = 1, \ldots, d$, is continuous on $(0,1)$, $\tilde{L}(\boldsymbol{u})$ is also continuous on $(0,1)^d$. Note that $F_i^{-1}(u)$, $i = 1, \ldots, d$, is also strictly increasing on $(0,1)$ due to the continuity of X_i. Then, Equation (2.5) says that $\tilde{L}(\boldsymbol{u})$ is strictly increasing on $(0,1)^d$, i.e. for $\boldsymbol{u}, \boldsymbol{u}' \in (0,1)^d$,

$$\tilde{L}(\boldsymbol{u}) < \tilde{L}(\boldsymbol{u}'), \quad \boldsymbol{u} < \boldsymbol{u}'. \tag{2.6}$$

The assumption (2.2) on $L(\boldsymbol{x})$ and the definition of $\tilde{L}(\boldsymbol{u})$ given in Equation (2.5) imply that $\tilde{L}(\boldsymbol{u})$ also diverges as $\max_{1 \le i \le d} u_i$ goes to 1, i.e.

$$\lim_{\max\{u_1, \ldots, u_d\} \to 1} \tilde{L}(\boldsymbol{u}) = \infty. \tag{2.7}$$

When $\boldsymbol{U}$ follows the copula C_0, it follows from Equations (2.4) and (2.5) that $\tilde{L}(\boldsymbol{U}) =_d L(\boldsymbol{X})$. The tail probability l in Equation (2.1) is represented by

$$l = \Pr\left\{\tilde{L}(\boldsymbol{U}) > \gamma\right\}. \tag{2.8}$$

Let $I(A)$ be the indicator function of an event A and $H(\boldsymbol{u}) = I(\tilde{L}(\boldsymbol{u}) > \gamma)$ for $\boldsymbol{u} \in (0,1)^d$. Then, Equation (2.8) is rewritten as

$$l = E_{c_0}\left[H(\boldsymbol{U})\right], \tag{2.9}$$

where $E_c[h(\boldsymbol{U})]$ is the expected value of $h(\boldsymbol{U})$ for a real valued function h and a random vector $\boldsymbol{U}$ following a pdf c.

Usually, the distribution of $\tilde{L}(\boldsymbol{U})$ is not tractable, and it is not easy to compute the value of l analytically. Instead, we can get an estimate of it by the Monte Carlo simulation. Suppose that sampling from copula C_0 is not hard. Then, we generate N samples of $\boldsymbol{U}$ independently from C_0, and denote them by $\boldsymbol{U}^{(1)}, \ldots, \boldsymbol{U}^{(N)}$, respectively. Equation (2.9) gives the CMC (crude Monte Carlo) estimator of l as follows:

$$\hat{l}_{\mathrm{CMC}} = \frac{1}{N} \sum_{j=1}^{N} H\left(\boldsymbol{U}^{(j)}\right). \tag{2.10}$$

2.2. Inverse transform likelihood ratio method

For the efficient estimation of l, a sampling distribution different from $C_0(\boldsymbol{u})$ can be applied to generate random copies of $\boldsymbol{U}$. If we choose the sampling density of $\boldsymbol{U}$ appropriately, then the variance of the importance sampling estimator may be much less than that of the CMC estimator in Equation (2.10). We propose an importance sampling by applying the inverse transform likelihood method (Kroese and Rubinstein, 2004). We denote by $G(\boldsymbol{u})$ the importance sampling cdf of $\boldsymbol{U}$. We consider the following form of distribution as a candidate of $G(\boldsymbol{u})$:

$$G(\boldsymbol{u}) = C\left(G_1(u_1), \ldots, G_d(u_d)\right), \quad \boldsymbol{u} \in (0,1)^d, \tag{2.11}$$

where $C(\boldsymbol{u})$ is a copula defined on $(0,1)^d$ and $G_i(u)$, $i = 1,\ldots,d$, is a cdf with support $(0,1)$. Note that $G_i(u)$ is the marginal cdf of the i^{th} component of $G(u)$. We assume that $G_i(u)$ is differentiable and $G_i'(x) = g_i(x)$, $x \in (0,1)$. If we denote by $g(\boldsymbol{u})$ the pdf of $G(\boldsymbol{u})$, then $g(\boldsymbol{u})$ is represented as

$$g(\boldsymbol{u}) = c\left(G_1(u_1),\ldots,G_d(u_d)\right)\prod_{i=1}^{d} g_i(u_i), \quad \boldsymbol{u} \in (0,1)^d, \tag{2.12}$$

where c is the pdf of C.

Suppose that $\boldsymbol{U}$ is a sample generated from G. Then, the likelihood ratio of $\boldsymbol{U}$ with respect to c_0 is given by

$$W(\boldsymbol{U}) = \frac{c_0(\boldsymbol{u})}{g(\boldsymbol{u})}, \quad \boldsymbol{u} \in (0,1)^d,$$

and Equation (2.9) is rewritten as follows:

$$l = E_g\left[H(\boldsymbol{U})W(\boldsymbol{U})\right]. \tag{2.13}$$

Suppose that we have N random copies $\boldsymbol{U}^{(1)},\ldots,\boldsymbol{U}^{(N)}$ of $\boldsymbol{U}$. Then, it follows from Equation (2.13) that an importance sampling estimator of l is given by

$$\hat{l}_{\mathrm{IS}} = \frac{1}{N}\sum_{j=1}^{N} H\left(\boldsymbol{U}^{(j)}\right) W\left(\boldsymbol{U}^{(j)}\right). \tag{2.14}$$

To get a sample of random vector following G in Equation (2.11), we first need to generate a random vector $\boldsymbol{V} = (V_1,\ldots,V_d)$ from copula $C(\boldsymbol{u})$. Since $G_i(u)$ is strictly increasing on $(0,1)$, it has the inverse function G_i^{-1}. If we consider a random vector defined as

$$\boldsymbol{U}' = \left(G_1^{-1}(V_1),\ldots,G_d^{-1}(V_d)\right),$$

then it follows that for $\boldsymbol{u} = (u_1,\ldots,u_d) \in (0,1)^d$,

$$\begin{aligned}\Pr\left\{\boldsymbol{U}' \le \boldsymbol{u}\right\} &= \Pr\left\{G_1^{-1}(V_1) \le u_1,\ldots,G_d^{-1}(V_d) \le u_d\right\}\\ &= \Pr\left\{V_1 \le G_1(u_1),\ldots,V_d \le G_d(u_d)\right\}.\end{aligned}$$

Since the random vector $\boldsymbol{V}$ follows the cdf $C(\boldsymbol{u})$, the cdf of $\boldsymbol{U}'$ is $G(\boldsymbol{u})$ in Equation (2.11).

Algorithm 1 shows the procedure to get an estimate of l as described above. For efficient estimation of l, sampling from $C(\boldsymbol{u})$ should not be hard. We also should choose $G(\boldsymbol{u})$ such that $G_i^{-1}(u)$ and $g_i(u)$, $i = 1,\ldots,d$, have the explicit form or the numerical computation of their values for given u is not difficult.

2.3. The cross entropy method

Kroese and Rubinstein (2004) adopted the cross entropy method to find the optimal importance distribution. We extend their method to our case. Let $g^*(\cdot)$ be the zero-variance pdf given by

$$g^*(\boldsymbol{u}) = \frac{H(\boldsymbol{u})c_0(\boldsymbol{u})}{l}, \quad \boldsymbol{u} \in (0,1)^d.$$

Algorithm 1 : Inverse transform likelihood ratio method for a function of a multivariate random variable

Require: loss function $\tilde{L}$, threshold γ, nominal copula C_0, marginal importance sampling cdfs $G_1, \ldots, G_d$, marginal importance sampling pdfs $g_1, \ldots, g_d$, importance sampling copula C

Ensure: $\hat{\mathbb{P}}(\tilde{L}(\boldsymbol{U}) > \gamma)$

1: Set N as the total number of samples.

2: Generate $\boldsymbol{V}^{(1)}, \ldots, \boldsymbol{V}^{(N)}$ independently from copula C.

3: Get $\boldsymbol{U}^{(1)}, \ldots, \boldsymbol{U}^{(N)}$ using the transform: For $j = 1, \ldots, N$,

$$\boldsymbol{U}^{(j)} = \left(G_1^{-1}\left(V_1^{(j)}\right), \ldots, G_d^{-1}\left(V_d^{(j)}\right)\right).$$

4: Compute

$$H\left(\boldsymbol{U}^{(j)}\right) = I\left(\tilde{L}\left(\boldsymbol{U}^{(j)}\right) > \gamma\right), \quad j = 1, \ldots, N.$$

5: Compute the likelihood ratio

$$W\left(\boldsymbol{U}^{(j)}\right) = \frac{c_0\left(\boldsymbol{U}^{(j)}\right)}{g\left(\boldsymbol{U}^{(j)}\right)}, \quad j = 1, \ldots, N.$$

6: Return

$$\hat{l}_{\mathrm{IS}} = \frac{1}{N}\sum_{j=1}^{N} H\left(\boldsymbol{U}^{(j)}\right) W\left(\boldsymbol{U}^{(j)}\right).$$

Then, $g^*(\cdot)$ is the optimal importance pdf in the sense that generation of random copies of $\boldsymbol{U}$ from $g^*(\cdot)$ gives the minimum variance of $\hat{l}_{\mathrm{IS}}$ in Equation (2.14) (Rubinstein and Kroese, 2016). However, we do not know the value of l so that the sampling from $g^*(\cdot)$ is not easy, or inefficient. Instead, we try to find the importance sampling pdf with the minimum cross-entropy from $g^*(\cdot)$.

We confine the importance sampling pdf to a parametric family of pdfs $\mathcal{G}$. For a multivariate distribution belonging to $\mathcal{G}$, the marginal distribution of each component is the beta distribution and the dependence is modeled by the Gaussian copula. Let $\boldsymbol{U}$ be an importance sample following a distribution belonging to $\mathcal{G}$. Then, U_i, $i = 1, \ldots, d$, follows the beta distribution with parameters 1 and β_i. We reparametrize β_i as $\psi_i = 1/\beta_i$. Then, the pdf of U_i is

$$g_i(u; \psi_i) = \frac{1}{\psi_i}(1-u)^{-1+\frac{1}{\psi_i}}, \quad u \in (0,1), \tag{2.15}$$

and the cdf is

$$G_i(u; \psi_i) = 1 - (1-u)^{\frac{1}{\psi_i}}, \quad u \in (0,1). \tag{2.16}$$

When U_i follows the above cdf, we denote that U_i follows $\mathrm{Beta}(1, \psi_i)$ in what follows. The d-dimensional Gaussian copula has a d-dimensional correlation matrix as its parameter. The Gaussian copula with correlation matrix R has the following form:

$$C^G(\boldsymbol{u}; R) = \Phi_d\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d); R\right), \quad \boldsymbol{u} \in (0,1)^d,$$

where $\Phi_d(\cdot; R)$ is the cdf of the d-dimensional multivariate normal distribution with mean $(0, \ldots, 0)$ and variance-covariance matrix R, and Φ is the cdf of the standard normal distribution. We define $c^G(\cdot; R)$ as the pdf of $C^G(\cdot; R)$, and ϕ as the pdf of the standard normal distribution, respectively. Then, the explicit form of $c^G(\cdot; R)$ is as follows:

$$c^G(\boldsymbol{u}; R) = \frac{1}{\sqrt{|\det R|}} \exp\left\{-\frac{1}{2} z^T \left(R^{-1} - I_d\right) z\right\}, \tag{2.17}$$

where $\det R$ is the determinant of R, $z = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))^T$, and I_d is the d-dimensional identity matrix. We define $\boldsymbol{\theta} = (\psi_1, \ldots, \psi_d, R)$ and $\boldsymbol{\Theta} = (0, \infty)^d \times \mathcal{M}_d$, where $\mathcal{M}_d$ is the set of d-dimensional correlation matrices. It follows from Equation (2.12) that for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$g(\boldsymbol{u}; \boldsymbol{\theta}) = c^G\left(G_1(u_1; \psi_1), \ldots, G_d(u_d; \psi_d); R\right) \prod_{i=1}^{d} g_i(u_i; \psi_i), \quad \boldsymbol{u} \in (0, 1)^d. \tag{2.18}$$

Then, $\mathcal{G}$ is represented as $\{g(\boldsymbol{u}; \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$.

The optimal importance sampling distribution with minimum cross entropy from $g^*(\cdot)$ is obtained by solving the following problem (Rubinstein and Kroese, 2016) :

$$\underset{\boldsymbol{\theta} \in \Theta}{\text{maximize}}\, E_{c_0}\left[H(\boldsymbol{U}) \log g(\boldsymbol{U}; \boldsymbol{\theta})\right]. \tag{2.19}$$

Let $\boldsymbol{\theta}^*$ be the solution of Equation (2.19). Generally, it is not easy to find value of $\boldsymbol{\theta}^*$ analytically. Instead, we can find the approximate solution of Equation (2.19) by solving the following stochastic programming:

$$\underset{\boldsymbol{\theta} \in \Theta}{\text{maximize}} \sum_{j=1}^{M} H\left(\boldsymbol{U}^{(j)}\right) \log g\left(\boldsymbol{U}^{(j)}; \boldsymbol{\theta}\right), \tag{2.20}$$

where $\boldsymbol{U}^{(1)}, \ldots, \boldsymbol{U}^{(M)}$ are independent samples from $c_0(\cdot)$.

When the value of γ is large, only a small number of samples in $\boldsymbol{U}^{(1)}, \ldots, \boldsymbol{U}^{(M)}$ make $H(\boldsymbol{U}^{(j)})$ non-zero, and the solution of Equation (2.20) does not approximate $\boldsymbol{\theta}^*$ well. In this case, an iterative method is appropriate (Rubinstein and Korese, 2016). In the method, the sampling distribution of $\boldsymbol{U}$ is changed iteratively and the solution of Equation (2.19) with increasing thresholds instead of γ is found in each iteration. In the t^{th} step of the iterative method, we consider an importance sampling pdf $g(\cdot; \boldsymbol{\omega}^{(t-1)})$ of $\boldsymbol{U}$ and a threshold s_t so that $L(\boldsymbol{U})$ over s_t is not rare. In this end, for a positive real number ρ small but not close to 0, s_t is usually set to be the the $(1-\rho)$ quantile of the distribution with pdf $g(\cdot; \boldsymbol{\omega}^{(t-1)})$. The solution of the following maximization problem is found approximately in the t^{th} iteration:

$$\underset{\boldsymbol{\theta} \in \Theta}{\text{maximize}}\, E_{\boldsymbol{\omega}^{(t-1)}}\left[I\left(\tilde{L}(\boldsymbol{U}) \geq s_t\right) \log g(\boldsymbol{U}; \boldsymbol{\theta})\, W\left(\boldsymbol{U}; \boldsymbol{\omega}^{(t-1)}\right)\right], \tag{2.21}$$

where $E_{\omega}[\cdot]$ means the expectation with respect to the pdf $g(\cdot; \boldsymbol{\omega})$ and

$$W(\boldsymbol{u}; \boldsymbol{\omega}) = \frac{c_0(\boldsymbol{u})}{g(\boldsymbol{u}; \boldsymbol{\omega})}. \tag{2.22}$$

Since the occurrence of $L(\boldsymbol{U})$ over s_t is not rare, s_t is estimated as $\hat{s}_t$, the $(1-\rho)$ quantile of the generated samples of $\boldsymbol{U}$ following $g(\cdot;\boldsymbol{\omega}^{(t-1)})$. The solution of Equation (2.21) can be found approximately by solving the corresponding stochastic programming given by

$$\underset{\boldsymbol{\theta}\in\Theta}{\text{maximize}} \sum_{j=1}^{M} I\left(\tilde{L}\left(\boldsymbol{U}^{(j)}\right) \geq \hat{s}_t\right) \log g\left(\boldsymbol{U}^{(j)};\boldsymbol{\theta}\right) W\left(\boldsymbol{U}^{(j)},\boldsymbol{\omega}^{(t-1)}\right). \tag{2.23}$$

The solution of the above problem will be used as $\boldsymbol{\omega}^{(t)}$, the parameter of the importance sampling pdf in the $(t+1)$-st iteration. The iteration continuous until $\hat{s}_t \geq \gamma$.

We define $\boldsymbol{\theta}(s)$ as follows:

$$\boldsymbol{\theta}(s) = \underset{\boldsymbol{\theta}\in\Theta}{\text{argmax}}\, E_{c_0}\left[H\left(\boldsymbol{U};s\right)\log g\left(\boldsymbol{U};\boldsymbol{\theta}\right)\right],$$

where $H(\boldsymbol{u};s) = I(\tilde{L}(\boldsymbol{u}) \geq s)$. The continuity of $\tilde{L}(\cdot)$ implies that $\boldsymbol{\theta}(s)$ is continuous with respect to $s\,(s>0)$. Note that $\boldsymbol{\theta}(s)$ can be rewritten as for $\boldsymbol{\omega}\in\boldsymbol{\Theta}$,

$$\boldsymbol{\theta}(s) = \underset{\boldsymbol{\theta}\in\Theta}{\text{argmax}}\, E_{\boldsymbol{\omega}}\left[H(\boldsymbol{U};s)\log g(\boldsymbol{U};\boldsymbol{\theta})W(\boldsymbol{U};\boldsymbol{\omega})\right]. \tag{2.24}$$

Equation (2.24) shows that $\boldsymbol{\theta}(s)$ does not depend on the value of $\boldsymbol{\omega}$, and that $\boldsymbol{\theta}(s_t)$, $t=1,2,\ldots,$ is the solution of the problem (2.21). Note that $\boldsymbol{\theta}^* = \boldsymbol{\theta}(\gamma)$. If the sequence $\{\hat{s}_1,\hat{s}_2,\ldots\}$ becomes larger than γ and the iteration ends at the step t, then we obtain an approximate value of $\boldsymbol{\theta}^*$ by substituting γ for $\hat{s}_t$ in Equation (2.23).

2.4. The proposed scheme

When we apply the procedure described in the previous subsection, we need to solve Equation (2.21). It follows from Equation (2.18) that $E_{\boldsymbol{\omega}}[H(\boldsymbol{U};s)\log g(\boldsymbol{U};\boldsymbol{\theta})W(\boldsymbol{U};\boldsymbol{\omega})]$ is rewritten as

$$\begin{aligned}\underset{\boldsymbol{\theta}\in\Theta}{\text{maximize}}\Big(&\sum_{i=1}^{d} E_{\boldsymbol{\omega}}\left[H(\boldsymbol{U};s)\log g_i(U_i;\psi_i)W(\boldsymbol{U};\boldsymbol{\omega})\right]\\ &+ E_{\boldsymbol{\omega}}\left[H(\boldsymbol{U};s)\log c^{G}\left(G_1(U_1;\psi_1),\ldots,G_d(U_d;\psi_d);R\right)W(\boldsymbol{U};\boldsymbol{\omega})\right]\Big).\end{aligned} \tag{2.25}$$

Due to the complex form of $c^G(\boldsymbol{u};R)$, solving the above problem analytically is nearly impossible, and it is also not easy to solve the corresponding stochastic programming obtained by generating samples from $g(\cdot;\boldsymbol{\omega})$. Instead, we decompose the above optimization problem into two sub-problems. Let $\boldsymbol{\psi}(s) = (\psi_1(s),\ldots,\psi_d(s))$ be the solution of the first term of problem (2.25), i.e.

$$\psi_i(s) = \underset{\psi_i\in(0,\infty)}{\text{argmax}}\, E_{\boldsymbol{\omega}}\left[H(\boldsymbol{U};s)\log g_i(U_i;\psi_i)W(\boldsymbol{U};\boldsymbol{\omega})\right], \quad i=1,\ldots,d. \tag{2.26}$$

After substituting $\psi_i(s)$ for ψ_i in the second term of problem (2.25), the correlation matrix maximizing the second term is obtained as

$$R(s) = \underset{R\in\mathcal{M}_d}{\text{argmax}}\, E_{\boldsymbol{\omega}}\left[H(\boldsymbol{U};s)\log c^{G}\left(G_1(U_1;\psi_1(s)),\ldots,G_d(U_d;\psi_d(s));R\right)W(\boldsymbol{U};\boldsymbol{\omega})\right]. \tag{2.27}$$

Then $(\boldsymbol{\psi}(s),R(s))$ is an approximate value of $\boldsymbol{\theta}(s)$. We may apply this procedure to solve Equation (2.21) stochastically. However, it is not guaranteed that the sequence $\{\hat{s}_1,\hat{s}_2,\ldots\}$ becomes larger than γ.

In our proposed scheme, we choose an increasing sequence $\{\gamma_1, \gamma_2, \ldots, \gamma_K\}$ with γ_K being equal to γ, and find approximate values of $\boldsymbol{\psi}(\gamma_k)$ and $R(\gamma_k)$, $k = 1, 2, \ldots, K$. We denote them by $\hat{\boldsymbol{\psi}}(\gamma_k)$ and $\hat{R}(\gamma_k)$, respectively. Given $\hat{R}(\gamma_{k-1})$, we find $\hat{\boldsymbol{\psi}}(\gamma_k)$ by solving iteratively a series of stochastic programmings corresponding to Equation (2.26) in the same manner as described in the previous section. In each iteration, the correlation matrix R, which is an element of the importance sampling parameter $\boldsymbol{\omega}$, is fixed to $\hat{R}(\gamma_{k-1})$. Theorem 5 in Section 3 shows that the value of $\hat{s}_t$ is eventually larger than γ_k for a $t > 0$. In this case, we set $\hat{s}_t = \gamma$ and determine the value of $\hat{R}(\gamma_k)$ by solving Equation (2.27) stochastically. Note that there are sub-iterations in the iteration of finding $(\hat{\boldsymbol{\psi}}(\gamma_k), \hat{R}(\gamma_k))$. To avoid confusion, we call the iteration of finding $(\boldsymbol{\psi}(\gamma_k), R(\gamma_k))$ the stage k for $k = 1, 2, \ldots, K$, and call the t^{th} sub-iteration in a stage the t^{th} step for $t = 1, 2, \ldots$.

At the start of the stage k, we assume that $\hat{R}(\gamma_{k-1})$ is available for $k = 2, 3, \ldots, K$. In stage 1, we set $\hat{R}(\gamma_0)$ as the d-dimensional identity matrix, where we use the notation $\hat{R}(\gamma_0)$ for the notational consistency. At the first step of the stage k, we set $\boldsymbol{\psi}^{(0)} = (1, \ldots, 1) \in \mathbb{R}^d$. We generate M independent samples following pdf $g(\cdot; \boldsymbol{\psi}^{(0)}, \hat{R}(\gamma_{k-1}))$. At the t^{th} step ($t \geq 2$), we generate M independent samples following pdf $g(\cdot; \hat{\boldsymbol{\psi}}^{(t-1)}, \hat{R}(\gamma_{k-1}))$, where $\hat{\boldsymbol{\psi}}^{(t-1)}$ is determined in the previous step. In order to get random samples following $g(\cdot; \hat{\boldsymbol{\psi}}^{(t-1)}, \hat{R}(\gamma_{k-1}))$ for $t \geq 1$, we first generate independently the multivariate normal vectors $\boldsymbol{Z}^{(1)}, \ldots, \boldsymbol{Z}^{(M)}$ with mean $(0, \ldots, 0)$ and variance-covariance matrix $\hat{R}(k-1)$, and transform it into

$$V_i^{(j)} = \Phi\left(Z_i^{(j)}\right), \quad i = 1, \ldots, d,$$

where $Z_i^{(j)}$ is the i^{th} element of $\boldsymbol{Z}^{(j)}$. Then, $(V_1^{(j)}, \ldots, V_d^{(j)})$ has copula $C^G(\cdot; \hat{R}(\gamma_{k-1}))$ as its cdf. Using the following transform, we get

$$\boldsymbol{U}^{(j)} = \left(G_1^{-1}\left(V_1^{(j)}; \hat{\psi}_1^{(t-1)}\right), \ldots, G_d^{-1}\left(V_d^{(j)}; \hat{\psi}_d^{(t-1)}\right)\right), \quad j = 1, \ldots, M.$$

As we explained in Section 2.2, $\boldsymbol{U}^{(j)}$ has pdf $g(\cdot; \hat{\boldsymbol{\psi}}^{(t-1)}, \hat{R}(\gamma_{k-1}))$. Moreover, $\boldsymbol{U}^{(1)}, \ldots, \boldsymbol{U}^{(M)}$ are mutually independent due to the independence of $\boldsymbol{Z}^{(1)}, \ldots, \boldsymbol{Z}^{(M)}$. By substituting $\boldsymbol{u} = \boldsymbol{U}^{(j)}$ and $\boldsymbol{\omega} = (\hat{\boldsymbol{\psi}}^{(t-1)}, \hat{R}(\gamma_{k-1}))$ into Equation (2.22), we obtain the likelihood ratio $W_j^{(t)}$ of the sample $\boldsymbol{U}^{(j)}$, i.e.

$$W_j^{(t)} = \frac{c_0\left(\boldsymbol{U}^{(j)}\right)}{g\left(\boldsymbol{U}^{(j)}; \hat{\boldsymbol{\psi}}^{(t-1)}, \hat{R}(\gamma_{k-1})\right)}, \quad j = 1, \ldots, M.$$

Suppose that we have generated M samples of $\boldsymbol{U}$ at the t^{th} step. Then, we denote by $\hat{s}_t$ the $(1-\rho)$ quantile of the generated losses $\{\tilde{L}(\boldsymbol{U}^{(j)}), j = 1, \ldots, M\}$, and let $\mathcal{E}(\hat{s}_t) = \{j : \tilde{L}(\boldsymbol{U}^{(j)}) > \hat{s}_t, j = 1, \ldots, M\}$. The stochastic programming corresponding to Equation (2.26) is as follows:

$$\underset{\psi_i \in (0,\infty)}{\text{maximize}} \sum_{j \in \mathcal{E}(\hat{s}_t)} W_j^{(t)} \log g_i\left(U_i^{(j)}; \psi_i\right), \quad i = 1, \ldots, d. \tag{2.28}$$

It follow from Equation (2.15) that the solution of Equation (2.28) is denoted by

$$\hat{\psi}_i^{(t)} = \underset{\psi_i \in (0,1)}{\text{argmax}} \sum_{j \in \mathcal{E}(\hat{s}_t)} W_j^{(t)} \left(-\log \psi_i + \left(\frac{1}{\psi_i} - 1\right) \log\left(1 - U_i^{(j)}\right)\right).$$

The explicit form of $\hat{\psi}_i^{(t)}$ is obtained as

$$\hat{\psi}_i^{(t)} = -\frac{\sum_{j \in \mathcal{E}(\hat{s}_t)} W_j^{(t)} \log\left(1 - U_i^{(j)}\right)}{\sum_{j \in \mathcal{E}(\hat{s}_t)} W_j^{(t)}}. \tag{2.29}$$

Algorithm 2 : The proposed scheme

Require: nominal pdf $c_0(\boldsymbol{u})$, loss function $\tilde{L}(\boldsymbol{u})$, threshold γ, the number of iterations K, thresholds in each iteration, $\gamma_1, \gamma_2, \ldots, \gamma_{K-1}$, the importance sampling pdf $g(\boldsymbol{u};\boldsymbol{\theta})$ given in Equation (2.18), the value of ρ, the likelihood ratio function $W(\boldsymbol{u};\boldsymbol{\theta})$.

Ensure: $\hat{\mathbb{P}}(\tilde{L}(\boldsymbol{U}) > \gamma)$

$\gamma_K \leftarrow \gamma$
$\hat{R}(\gamma_0) \leftarrow$ the d-dimensional identity matrix
for $k = 1$ to K **do**
  $\hat{s} \leftarrow 0$
  $\hat{\boldsymbol{\psi}} \leftarrow (1, \ldots, 1)$
  while $\hat{s} < \gamma_k$ **do**
    Generate M random samples $\boldsymbol{U}^{(1)}, \ldots, \boldsymbol{U}^{(M)}$ from $g(\boldsymbol{u}; \hat{\boldsymbol{\psi}}, \hat{R}(\gamma_{k-1}))$
    $\hat{s} \leftarrow \min\{\gamma_k,$ the $(1-\rho)$ quantile of $\tilde{L}(\boldsymbol{U}^{(1)}), \ldots, \tilde{L}(\boldsymbol{U}^{(M)})\}$
    $\mathcal{E}(\hat{s}) \leftarrow \{j : \tilde{L}(\boldsymbol{U}^{(j)}) > \hat{s}, j = 1, \ldots, M\}$
    for $j \in \mathcal{E}(\hat{s})$ **do**
      $W_j \leftarrow W(\boldsymbol{U}^{(j)}; \hat{\boldsymbol{\psi}}, \hat{R}(\gamma_{k-1}))$
    end for
    for $i = 1$ to d **do**

$$\hat{\psi}_i \leftarrow -\frac{\sum_{j \in \mathcal{E}(\hat{s})} W_j \log\left(1 - U_i^{(j)}\right)}{\sum_{j \in \mathcal{E}(\hat{s})} W_j}$$

    end for
    $\hat{\boldsymbol{\psi}} \leftarrow (\hat{\psi}_1, \ldots, \hat{\psi}_d)$
  end while
  $\hat{\boldsymbol{\psi}}(\gamma_k) \leftarrow \hat{\boldsymbol{\psi}}$
  for $j \in \mathcal{E}(\gamma^{(k)})$ **do**
    $\boldsymbol{Z}^{(j)} \leftarrow (\Phi^{-1}(G_1(U_1^{(j)}; \hat{\psi}_1(\gamma_k))), \ldots, \Phi^{-1}(G_d(U_d^{(j)}, \hat{\psi}_d(\gamma_k))))$
  end for

$$\hat{R}(\gamma_k) \leftarrow \frac{\sum_{j \in \mathcal{E}(\gamma^{(k)})} W_j \left(\boldsymbol{Z}^{(j)}\right)^T \boldsymbol{Z}^{(j)}}{\sum_{j \in \mathcal{E}(\gamma^{(k)})} W_j}$$

end for
$\hat{\boldsymbol{\theta}} \leftarrow (\hat{\boldsymbol{\psi}}(\gamma_K), \hat{R}(\gamma_K))$
Generate N random samples $\boldsymbol{U}^{(1)}, \ldots, \boldsymbol{U}^{(N)}$ from $g(\boldsymbol{u}; \hat{\boldsymbol{\theta}})$
Return

$$\hat{l}_{\mathrm{IS}} = \frac{1}{N} \sum_{j=1}^{N} H\left(\boldsymbol{U}^{(j)}\right) W\left(\boldsymbol{U}^{(j)}; \hat{\boldsymbol{\theta}}\right).$$

We iterate the above procedure for $t = 1, 2, \ldots$ until $\hat{s}_t > \gamma_k$. Theorem 5 in Section 3 shows that the iteration terminates with probability 1. Suppose that $\hat{s}_t$ is larger than γ_k at step $t = \tau$, and that $\{\boldsymbol{U}^{(j)}, j = 1, \ldots, M\}$ are the generated samples at this step. Then, each element of $\hat{\boldsymbol{\psi}}(\gamma_k)$ is obtained by

substituting γ_k for $\hat{s}_t$ in Equation (2.29), i.e.

$$\hat{\psi}_i(\gamma_k) = -\frac{\sum_{j\in\mathcal{E}(\gamma_k)} W_j^{(\tau)} \log\left(1 - U_i^{(j)}\right)}{\sum_{j\in\mathcal{E}(\gamma_k)} W_j^{(\tau)}}, \quad i = 1, \ldots, d.$$

We derive the stochastic programming corresponding to Equation (2.27) with $s = \gamma_k$ is as follows:

$$\underset{R\in\mathcal{M}_d}{\text{maximize}} \sum_{j\in\mathcal{E}(\gamma_k)} W_j^{(\tau)} \log c^G\left(G_1\left(U_1^{(j)}; \hat{\psi}_1(\gamma_k)\right), \ldots, G_d\left(U_d^{(j)}; \hat{\psi}_d(\gamma_k)\right); R\right). \tag{2.30}$$

Let $Z_i^{(j)} = \Phi^{-1}(G_i(U_i^{(j)}; \hat{\psi}_i(\gamma_k)))$, $i = 1, \ldots, d$, and let $\boldsymbol{Z}^{(j)} = (Z_1^{(j)}, \ldots, Z_d^{(j)})$, $j = 1, \ldots, M$. By applying the form of c^G in Equation (2.17) to Equation (2.30), we can see that $R(\gamma_k)$ is estimated as

$$\hat{R}(\gamma_k) = \underset{R\in\mathcal{M}_d}{\text{argmin}} \sum_{j\in\mathcal{E}\left(\gamma^{(k)}\right)} W_j^{(\tau)} \left(\left(\boldsymbol{Z}^{(j)}\right)^T R^{-1} \boldsymbol{Z}^{(j)} + \log|\det R|\right). \tag{2.31}$$

It can be easily shown that the solution of Equation (2.31) is given by

$$\hat{R}(\gamma_k) = \frac{\sum_{j\in\mathcal{E}(\gamma^{(k)})} W_j^{(\tau)} \left(\boldsymbol{Z}^{(j)}\right)^T \boldsymbol{Z}^{(j)}}{\sum_{j\in\mathcal{E}(\gamma^{(k)})} W_j^{(\tau)}}. \tag{2.32}$$

Some diagonals of $\hat{R}(\gamma_k)$ in Equation (2.32) may not be 1. We normalize it so that it is a correlation matrix.

If we let $\hat{\boldsymbol{\theta}}(\gamma_k) = (\hat{\boldsymbol{\psi}}(\gamma_k), \hat{R}(\gamma_k))$, then $\hat{\boldsymbol{\theta}}(\gamma_k)$ is an approximate value of $\boldsymbol{\theta}(\gamma_k)$ for $k = 1, 2, \ldots, K$. Since γ_K is equal to γ, $\hat{\boldsymbol{\theta}}(\gamma_K)$ is the estimate of $\boldsymbol{\theta}^*$. Algorithm 2 shows the procedure to get the pseudo-optimal parameter $\hat{\boldsymbol{\theta}}(\gamma_K)$.

3. Convergence of the iterative scheme

In this section, we will show that if the number of importance samples generated in each step is sufficiently large, then the algorithm terminates in a finite number of steps with a probability close to 1 We define $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_d)$. Suppose that a random multivariate variable $\boldsymbol{U}$ follows the joint pdf $g(\cdot; \boldsymbol{\psi}, R)$. Since $\tilde{L}(\boldsymbol{u})$ is continuous and strictly increasing on $(0, 1)^d$ and the support of $g(\boldsymbol{u}; \boldsymbol{\psi}, R)$ is $(0, 1)^d$, $\tilde{L}(\boldsymbol{U})$ is a continuous random variable and the support of $\tilde{L}(\boldsymbol{U})$ is the same as that of $L(\boldsymbol{X})$, where $\boldsymbol{X}$ is a multivariate random variable with cdf F. Thus, the support of $\tilde{L}(\boldsymbol{U})$ is (s_L, ∞).

The cdf of $\tilde{L}(\boldsymbol{U})$ is as follows:

$$F_{\tilde{L}}(s; \boldsymbol{\psi}, R) = \int_{(0,1)^d} 1\left(\tilde{L}(\boldsymbol{u}) \le s\right) g(\boldsymbol{u}; \boldsymbol{\psi}, R)\, d\boldsymbol{u}, \quad s > 0. \tag{3.1}$$

The representation of $g(\cdot; \boldsymbol{\psi}, R)$ in Equation (2.18) says that $g(\cdot; \boldsymbol{\psi}, R)$ is continuous with respect to $\boldsymbol{\psi}$ on $(0, \infty)^d$, which means that $F_{\tilde{L}}(s; \boldsymbol{\psi}, R)$ is also continuous with respect to $\boldsymbol{\psi}$ on $(0, \infty)^d$. We denote by $s(\boldsymbol{\psi}; R)$ the $(1 - \rho)$ quantile of $\tilde{L}(\boldsymbol{U})$. Then, $s(\boldsymbol{\psi}; R)$ is the unique solution of the following equation:

$$F_{\tilde{L}}(s; \boldsymbol{\psi}, R) = 1 - \rho, \quad 0 < \rho < 1. \tag{3.2}$$

Theorem 1. *$s(\boldsymbol{\psi}; R)$ is continuous with respect to $\boldsymbol{\psi}$ on $(0, \infty)^d$.*

Proof: Since $F_{\tilde{L}}(s;\boldsymbol{\psi},R)$ is continuous with respect to $\boldsymbol{\psi}$ on $(0,\infty)^d$, we have that

$$\lim_{\boldsymbol{\psi}\to\boldsymbol{\psi}_0} F_{\tilde{L}}\left(s(\boldsymbol{\psi}_0;R);\boldsymbol{\psi},R\right) = F_{\tilde{L}}\left(s(\boldsymbol{\psi}_0;R);\boldsymbol{\psi}_0,R\right).$$

It follows from Equations (3.1) and (3.2) that

$$\lim_{\boldsymbol{\psi}\to\boldsymbol{\psi}_0} \int_{(0,1)^d} 1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi}_0;R)\right) g(\boldsymbol{u};\boldsymbol{\psi},R)\, d\boldsymbol{u} - (1-\rho) = 0. \tag{3.3}$$

Since $s(\boldsymbol{\psi};R)$ is the $(1-\rho)$ quantile of $\tilde{L}(\boldsymbol{U})$, we have that for $\boldsymbol{\psi}\in(0,\infty)^d$,

$$\int_{(0,1)^d} 1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi};R)\right) g(\boldsymbol{u};\boldsymbol{\psi},R)\, d\boldsymbol{u} = 1-\rho.$$

By applying the above equation to Equation (3.3), we have that

$$\lim_{\boldsymbol{\psi}\to\boldsymbol{\psi}_0} \int_{(0,1)^d} \left\{1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi}_0;R)\right) - 1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi};R)\right)\right\} g(\boldsymbol{u};\boldsymbol{\psi},R)\, d\boldsymbol{u} = 0. \tag{3.4}$$

If $s(\boldsymbol{\psi}_0;R) < s(\boldsymbol{\psi};R)$, then

$$1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi}_0;R)\right) - 1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi};R)\right) = \begin{cases} -1, & s(\boldsymbol{\psi}_0;R) < \tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi};R), \\ 0, & \text{otherwise.} \end{cases}$$

If $s(\boldsymbol{\psi};R) \le s(\boldsymbol{\psi}_0;R)$, then

$$1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi}_0;R)\right) - 1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi};R)\right) = \begin{cases} 1, & s(\boldsymbol{\psi};R) < \tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi}_0;R), \\ 0, & \text{otherwise.} \end{cases}$$

The above two equations give that

$$1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi}_0;R)\right) - 1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi};R)\right) = (-1)^{1(s(\boldsymbol{\psi}_0,R)\le s(\boldsymbol{\psi},R))} \left|1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi}_0;R)\right) - 1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi};R)\right)\right|.$$

Then, Equation (3.4) implies that

$$\lim_{\boldsymbol{\psi}\to\boldsymbol{\psi}_0} \int_{(0,1)^d} \left|1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi}_0;R)\right) - 1\left(\tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi};R)\right)\right| g\left(\boldsymbol{u};\boldsymbol{\psi},R\right)\, d\boldsymbol{u} = 0.$$

The integrand of the above equation is nonnegative for each $\boldsymbol{\psi}$ on $(0,\infty)^d$. Thus, it converges to 0 almost everywhere. Since $g(\boldsymbol{u};\boldsymbol{\psi},R) > 0$ for $\boldsymbol{u}\in(0,1)^d$, the above equation implies that the set $\{\boldsymbol{u} : s(\boldsymbol{\psi}_0;R) < \tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi};R)\} \cup \{\boldsymbol{u} : s(\boldsymbol{\psi};R) < \tilde{L}(\boldsymbol{u}) \le s(\boldsymbol{\psi}_0;R)\}$ converges to a set whose measure is 0. The continuity of $\tilde{L}(\boldsymbol{u})$ on $(0,1)^d$ implies that $s(\boldsymbol{\psi};R) \to s(\boldsymbol{\psi}_0;R)$ as $\boldsymbol{\psi}\to\boldsymbol{\psi}_0$. □

When $\boldsymbol{U}$ follows the pdf $g(\cdot;\boldsymbol{\psi},R)$, $E[U_i]$ is an increasing function of ψ_i, $i = 1,\ldots,d$. Thus, we can guess that $\Pr\{\tilde{L}(\boldsymbol{U}) > \gamma\}$ may be increasing with respect to each of ψ_i, or equivalently, the $(1-\rho)$ quantile of $\tilde{L}(\boldsymbol{U})$ may be increasing with respect to each of ψ_i. The following theorem shows that this is true.

Theorem 2. *$s(\boldsymbol{\psi};R)$ is a strictly increasing function of $\boldsymbol{\psi}$.*

Proof: For $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_d) \in (0, \infty)^d$ and $\boldsymbol{\psi}' = (\psi'_1, \ldots, \psi'_d) \in (0, \infty)^d$, we assume that $\boldsymbol{\psi} < \boldsymbol{\psi}'$. Then, it follows that for given $\boldsymbol{v} = (v_1, \ldots, v_d) \in (0, 1)^d$,

$$\left(1 - (1 - v_1)^{\psi_1}, \ldots, 1 - (1 - v_d)^{\psi_d}\right) < \left(1 - (1 - v_1)^{\psi'_1}, \ldots, 1 - (1 - v_d)^{\psi'_d}\right).$$

Since $\tilde{L}(\boldsymbol{u})$ is a strictly increasing function of $\boldsymbol{u}$, we have that for given $(v_1, \ldots, v_d) \in (0, 1)^d$,

$$\tilde{L}\left(1 - (1 - v_1)^{\psi_1}, \ldots, 1 - (1 - v_d)^{\psi_d}\right) < \tilde{L}\left(1 - (1 - v_1)^{\psi'_1}, \ldots, 1 - (1 - v_d)^{\psi'_d}\right). \tag{3.5}$$

Suppose that $\boldsymbol{V} = (V_1, \ldots, V_d)$ is a random vector following the Gaussian copula with correlation matrix R. We define two random vectors

$$\boldsymbol{U} = \left(1 - (1 - V_1)^{\psi_1}, \ldots, 1 - (1 - V_d)^{\psi_d}\right),$$
$$\boldsymbol{U}' = \left(1 - (1 - V_1)^{\psi'_1}, \ldots, 1 - (1 - V_d)^{\psi'_d}\right).$$

Then, $\boldsymbol{U}$ and $\boldsymbol{U}'$ follow the pdfs $g(\cdot; \boldsymbol{\psi}, R)$ and $g(\cdot; \boldsymbol{\psi}', R)$, respectively. Since $s(\boldsymbol{\psi}; R)$ is the $(1 - \rho)$ quantile of $\tilde{L}(\boldsymbol{U})$, we have that

$$\Pr\left\{\tilde{L}(\boldsymbol{U}) \le s(\boldsymbol{\psi}; R)\right\} = 1 - \rho.$$

Equation (3.5) says that $\Pr\{\tilde{L}(\boldsymbol{U}') \le s(\boldsymbol{\psi}; R)\} < \Pr\{\tilde{L}(\boldsymbol{U}) \le s(\boldsymbol{\psi}; R)\}$. It implies that

$$\Pr\left\{\tilde{L}(\boldsymbol{U}') \le s(\boldsymbol{\psi}; R)\right\} < 1 - \rho.$$

Note that $s(\boldsymbol{\psi}'; R)$ is the $(1 - \rho)$ quantile of $\tilde{L}(\boldsymbol{U}')$. Then, we have that $s(\boldsymbol{\psi}; R) < s(\boldsymbol{\psi}'; R)$. □

In this subsection, we assume that the converse of Equation (2.6) is also valid in a stochastic sense. Specifically, we assume that for $\boldsymbol{U}$ with pdf c_0, the conditional random variable $U_i|(\tilde{L}(\boldsymbol{U}) = s)$, $i = 1, \ldots, d$, is stochastically increasing with respect to s in the strict sense, i.e. for $s < s'$,

$$\Pr\left\{U_i > u \mid \tilde{L}(\boldsymbol{U}) = s\right\} < \Pr\left\{U_i > u \mid \tilde{L}(\boldsymbol{U}) = s'\right\}, \quad u \in (0, 1). \tag{3.6}$$

The solution of Equation (2.26) is rewritten as

$$\psi_i(s) = \underset{\psi_i \in (0,\infty)}{\operatorname{argmax}} E_{c_0}\left[I\left(\tilde{L}(\boldsymbol{U}) > s\right) \log g_i(U_i; \psi_i)\right].$$

It follows from Equation (2.15) that $\psi_i(s)$ is represented as

$$\psi_i(s) = E_{c_0}\left[-\log(1 - U_i) \mid \tilde{L}(\boldsymbol{U}) > s\right], \quad s > 0. \tag{3.7}$$

Theorem 3. *Suppose that for $\boldsymbol{U}$ with pdf c_0, the conditional random variable $U_i|(\tilde{L}(\boldsymbol{U}) = s)$, $i = 1, \ldots, d$, is stochastically increasing with respect to s. Then, $\psi_i(s)$, $i = 1, \ldots, d$, is a strictly increasing and continuous function of s, $s > 0$.*

Proof: Let $f_{\tilde{L}}(\cdot)$ is the pdf of $\tilde{L}(\boldsymbol{U})$ when $\boldsymbol{U}$ follows pdf c_0, and let $f_i(u, s)$ be the joint pdf of U_i and $\tilde{L}(\boldsymbol{U})$. We also denote by $f_{U_i|\tilde{L}}(u|s)$ the conditional pdf of U_i given that $\tilde{L}(\boldsymbol{U}) = s$, i.e.

$$f_{U_i|\tilde{L}}(u \mid s) = \frac{f_i(u, s)}{f_{\tilde{L}}(s)}.$$

Since $d\Pr\{U_i > u|\tilde{L}(\boldsymbol{U}) = s\}/du = -f_{U_i|\tilde{L}}(u|s)$, we have that

$$\begin{aligned} E_{c_0}\left[-\log(1-U_i)\,|\,\tilde{L}(\boldsymbol{U}) = s\right] &= \int_0^1 -\log(1-u) f_{U_i|\tilde{L}}(u\,|\,s)\,du \\ &= \int_0^1 \frac{\Pr\left\{U_i > u\,|\,\tilde{L}(\boldsymbol{U}) = s\right\}}{1-u}\,du. \end{aligned}$$

By applying Equation (3.6) to the above equation, we have that for $s < s'$,

$$E_{c_0}\left[-\log(1-U_i)\,|\,\tilde{L}(\boldsymbol{U}) = s\right] < E_{c_0}\left[-\log(1-U_i)\,|\,\tilde{L}(\boldsymbol{U}) = s'\right].$$

Then, Equation (3.7) implies that $\psi_i(s)$ is strictly increasing with respect to s for $i = 1, \ldots, d$.

It follows from Equation (3.7) that

$$\psi_i(s) = \frac{\int_s^\infty \int_0^1 -\log(1-u) f_i(u,y)\,dudy}{\Pr\left\{\tilde{L}(\boldsymbol{U}) > s\right\}}. \tag{3.8}$$

Both of the numerator and the denominator of $\psi_i(s)$ in Equation (3.8) are continuous with respect to $s \in (0,\infty)$. Thus, $\psi_i(s)$ is a continuous function of s on $(0,\infty)$. □

Suppose that a correlation matrix $R \in \mathcal{M}_d$ are given. Let $\boldsymbol{\psi}^{(0)} = (1, \ldots, 1)$. We define $\boldsymbol{\psi}^{(t)}$ and s_t, $t = 1, 2, \ldots$, iteratively, as follows:

$$\begin{aligned} s_t &= s\left(\boldsymbol{\psi}^{(t-1)}; R\right), \\ \psi_i^{(t)} &= \psi_i(s_t), \quad i = 1, \ldots, d, \end{aligned} \tag{3.9}$$

where $\boldsymbol{\psi}^{(t)} = (\psi_1^{(t)}, \ldots, \psi_d^{(t)})$. It can be easily shown that

$$E_{c_0}\left[-\log(1-U_i)\right] = 1, \quad i = 1, \ldots, d,$$

which means that

$$\psi_i^{(0)} = E_{c_0}\left[-\log(1-U_i)\,|\,\tilde{L}(\boldsymbol{U}) > 0\right].$$

Since $s_1 > 0$, we have from Theorem 3 that

$$\boldsymbol{\psi}^{(1)} > \boldsymbol{\psi}^{(0)}.$$

and from Theorem 2 that $s(\boldsymbol{\psi}^{(1)}; R) > s(\boldsymbol{\psi}^{(0)}; R)$, equivalently,

$$s_2 > s_1.$$

Then, Equation (3.9) gives that for $t = 1, 2, \ldots$,

$$\begin{aligned} s_{t+1} &> s_t, \\ \boldsymbol{\psi}^{(t+1)} &> \boldsymbol{\psi}^{(t)}. \end{aligned} \tag{3.10}$$

Moreover, Equation (3.9) says that the mapping from $\boldsymbol{\psi}^{(t)}$ to $\boldsymbol{\psi}^{(t+1)}$ is represented as

$$h(\boldsymbol{\psi}) = (\psi_1(s(\boldsymbol{\psi}; R)), \ldots, \psi_d(s(\boldsymbol{\psi}; R))) . \tag{3.11}$$

In stage k of the proposed scheme described in Section 2.4, $\hat{s}_t$ and $\hat{\boldsymbol{\psi}}^{(t)}$, $t = 1, 2, \ldots$, are the approximate values of s_t and $\boldsymbol{\psi}^{(t)}$ with R being equal to $\hat{R}(\gamma_{k-1})$. Note that $\hat{s}_t$ is the $(1-\rho)$ quantile of the samples following $g(\cdot; \hat{\boldsymbol{\psi}}^{(t-1)}, \hat{R}(\gamma_{k-1}))$, and that $\hat{\boldsymbol{\psi}}^{(t)}$, $t = 1, 2, \ldots$, is the cross entropy estimator of $\boldsymbol{\psi}^{(t)}$. The following theorem due to Lieber (1998) describes the conditions under which the sequence $\{\hat{s}_1, \hat{s}_2, \ldots\}$ becomes larger than γ_k in a stochastic sense (Rrubinstein, 1999).

Theorem 4. (Lieber, 1998) *If the following conditions hold:*

(i) The sequence $\{s_t, t = 1, 2, \ldots\}$ is strictly monotonic increasing.

(ii) $s(\boldsymbol{\psi}; R)$ is continuous with respect to $\boldsymbol{\psi}$.

(iii) $s(\boldsymbol{\psi}; R)$ is a proper function of $\boldsymbol{\psi}$, i.e. for a closed interval A, $\{\boldsymbol{\psi} : s(\boldsymbol{\psi}; R) \in A\}$ is compact.

(iv) $\Delta(\boldsymbol{\psi}) = s(h(\boldsymbol{\psi}); R) - s(\boldsymbol{\psi}; R)$ is lower semi-continuous,

then there exists an integer t such that for $x > 0$,

$$\lim_{M \to \infty} \Pr\{\hat{s}_t < \gamma_k\} = 0,$$

where M is the number of samples in each step.

By exploiting the above theorem, we can show that if the number of samples is sufficiently large, the sequence $\{\hat{s}_t, t = 1, 2, \ldots\}$ becomes eventually larger than γ_k.

Theorem 5. *In stage k for $k = 1, 2, \ldots, K$, there exists an integer t such that*

$$\lim_{M \to \infty} \Pr\{\hat{s}_t < \gamma_k\} = 0.$$

Proof: It is sufficient to show that the four conditions of Theorem 4 are satisfied. We can see from Equation (3.10) that the condition (i) is satisfied, and have shown that the condition (ii) is satisfied in Theorem 1. Since $s(\boldsymbol{\psi}; R)$ is continuous with respect to $\boldsymbol{\psi} \in (0, \infty)^d$ and $\psi_i(s)$ is continuous with respect to $s \in (0, \infty)$, $\psi_i(s(\boldsymbol{\psi}; R))$ is continuous with respect to $\boldsymbol{\psi} \in (0, \infty)^d$ for $i = 1, \ldots, d$. In other words, each component of $h(\boldsymbol{\psi})$ in (3.11) is continuous with respect to $\boldsymbol{\psi}$. Thus, $\Delta(\boldsymbol{\psi})$ is a continuous function of $\boldsymbol{\psi}$, which proves the condition (iv).

It remains to prove the condition (iii). For a closed interval $A = [a_1, a_2]$, we let $s^{-1}(A; R) = \{\boldsymbol{\psi} : s(\boldsymbol{\psi}; R) \in A\}$. Since $s(\boldsymbol{\psi}; R)$ is continuous with respect to $\boldsymbol{\psi}$, $s^{-1}(A; R)$ is closed. Suppose that $s^{-1}(A; R)$ is not bounded. In this case, $s^{-1}(A; R)$ contains a sequence $\{\boldsymbol{\psi}^{(1)}, \boldsymbol{\psi}^{(2)}, \ldots\}$ such that for constant $K_0 > 0$,

$$\left|\boldsymbol{\psi}^{(n)}\right| > \sqrt{d} K_0 n, \quad n = 1, 2, \ldots .$$

Then, there is an $i \in \{1, \ldots, d\}$ such that $\psi_i^{(n)} > K_0 n$, where $\psi_i^{(n)}$ is the i^{th} element of $\boldsymbol{\psi}^{(n)}$. Let $\boldsymbol{U}^{(n)}$ be the random vector following the pdf $g(\cdot; \boldsymbol{\psi}^{(n)}, R)$. Then,

$$\Pr\left\{\tilde{L}\left(\boldsymbol{U}^{(n)}\right) \leq a_2\right\} \geq 1 - \rho, \quad n = 1, 2, \ldots, \tag{3.12}$$

and it follows from Equation (2.16) that for $0 < u_0 < 1$,

$$\Pr\left\{U_i^{(n)} > u_0\right\} = (1-u_0)^{\frac{1}{\psi_i^{(n)}}},$$

where $U_i^{(n)}$ is the i^{th} element of $\boldsymbol{U}^{(n)}$. The above equation gives that

$$\Pr\left\{U_i^{(n)} > u_0\right\} > (1-u_0)^{\frac{1}{(K_0 n)}}.$$

Since $\Pr\{\max_{1\le i\le d} U_i^{(n)} > u_0\} \ge \Pr\{U_i^{(n)} > u_0\}$, we have that

$$\Pr\left\{\max_{1\le i\le d} U_i^{(n)} > u_0\right\} > (1-u_0)^{\frac{1}{(K_0 n)}},$$

which implies that

$$\lim_{n\to\infty} \max_{1\le i\le d} U_i^{(n)} = 1, \quad \text{in probability.}$$

Then, Equation (2.7) says that

$$\lim_{n\to\infty} \tilde{L}\left(\boldsymbol{U}^{(n)}\right) = \infty, \quad \text{in probability,}$$

which implies that $\lim_{n\to\infty} \Pr\{\tilde{L}(\boldsymbol{U}^{(n)}) \le a_2\} = 0$. It contradicts to Equation (3.12). Thus, $s^{-1}(A;R)$ is bounded. Then it is a compact set. □

Suppose that the current stage is the k^{th} stage with temporal threshold γ_k. If $\hat{s}_t < \gamma_k$ for $t = 1,\dots,\tau-1$, and $\hat{s}_\tau \ge \gamma_k$, then stage k ends at step τ. Let ϵ be a sufficiently small positive real number. According to Theorem 5, there are positive integers τ_k and M_k such that if the number of importance samples at each step is larger than M_k, then

$$\Pr\{\hat{s}_{\tau_k} \ge \gamma_k\} > 1 - \frac{\epsilon}{K}. \tag{3.13}$$

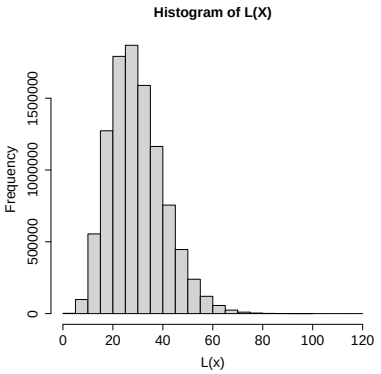
In the case that $\hat{s}_{\tau_k} \ge \gamma_k$, the stage k ends at step τ_k or before. Suppose that M in Algorithm 2 is greater than $\max_{1\le k\le K} M_k$. Then the number of steps required to terminate the stage k is less than or equal to τ_k with a probability larger than $1-\epsilon/K$. To get an estimate of $\boldsymbol{\theta}^*$, K stages must be completed. Let T be the total number of steps required to complete all stages. Then it follows that

$$\begin{aligned}\Pr\left\{T \le \sum_{k=1}^{K} \tau_k\right\} &\ge \Pr\{\hat{s}_{\tau_1} \ge \gamma_1, \dots, \hat{s}_{\tau_K} \ge \gamma_K\} \\ &= 1 - \Pr\left\{\cup_{k=1}^{K} \{\hat{s}_{\tau_k} < \gamma_k\}\right\} \\ &\ge 1 - \sum_{k=1}^{K} \Pr\{\hat{s}_{\tau_k} < \gamma_k\}.\end{aligned}$$

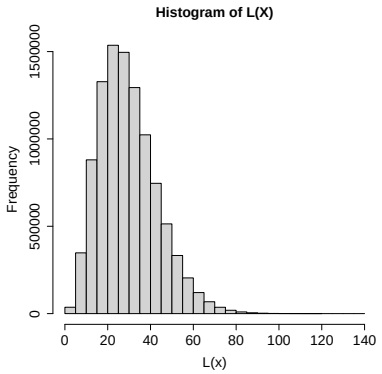
By applying Equation (3.13) to the above inequality, we have that

$$\Pr\left\{T \le \sum_{k=1}^{K} \tau_k\right\} \ge 1-\epsilon,$$

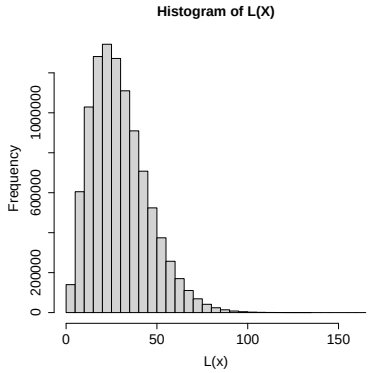
i.e. we can get an estimate of $\boldsymbol{\theta}^*$ within $\sum_{k=1}^{K} \tau_k$ steps with a probability larger than $1-\epsilon$. In a practical scenario, we do not know the value of τ_k and M_k. However, if we set the value M to be sufficiently large, then the algorithm terminates in a finite number of steps with a probability close to 1.



(a) *Samples of* $L(\boldsymbol{X})$ *with* $\eta = 0.25$.



(b) *Samples of* $L(\boldsymbol{X})$ *with* $\eta = 0.5$.



(c) *Samples of* $L(\boldsymbol{X})$ *with* $\eta = 0.75$.

Figure 1: *Histrogram of the* 10^7 *random samples of* $L(\boldsymbol{X})$ *with varying values of* η.

4. Numerical results

In this section, we compare the performance of the proposed scheme with that of the CMC estimator given in Equation (2.10). We estimate numerically the tail probabilities given in Equation (2.1) over various thresholds. The CMC estimator and the proposed scheme given in Algorithm 2 are applied to estimate the tail loss probabilities. We call the methods CMC and ITLR, respectively. We consider the loss function defined as follows:

$$L(\boldsymbol{x}) = \sum_{i=1}^{5} x_i + 2 \sum_{i=6}^{10} x_i, \quad \boldsymbol{x} \in (0, \infty)^{10}. \tag{4.1}$$

The joint cdf of $\boldsymbol{X} = (X_1, \ldots, X_{10})$ has the form in Equation (2.3). In this section, we assume that C_0 is the Gaussian copula with exchangeable correlation matrix R_0, and that X_i, $i = 1, \ldots, 10$, follows the Weibull distribution. The shape parameter of X_i, $i = 1, \ldots, 10$, is set to be

$$a_1 = a_2 = a_3 = a_6 = a_7 = a_8 = 1.5,$$
$$a_4 = a_5 = a_9 = a_{10} = 2.5,$$

Table 1: Summary statistics of the generated losses

Case	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
$\eta = 0.25$	1.51	21.70	28.39	29.61	36.21	116.83
$\eta = 0.5$	0.31	19.69	27.86	29.62	37.67	136.18
$\eta = 0.75$	0.08	17.90	27.30	29.62	38.85	161.86

and the scale parameter of X_i, $i = 1, \ldots, 10$, is set to be

$$\sigma_1 = \sigma_2 = \sigma_9 = \sigma_{10} = 1,$$
$$\sigma_3 = \sigma_4 = \sigma_7 = \sigma_8 = 2,$$
$$\sigma_5 = \sigma_6 = 5.$$

Let η be the off-diagonal element of R_0. We consider three cases of $\eta = 0.25, 0.5$, and 0.75. In order to infer the distributional properties of the loss $L(\boldsymbol{X})$, we have generated 10^7 random samples of the loss for each case of η. Figure 1 shows the histograms of the generated losses with varying values of η. We can see from the figure that the distributions of the losses don't look very different from each other except it seems to be more centered around the mean as the value of η becomes small.

Table 1 shows summary statistics of the generated losses. It follows from the table that the sample means of the generated losses have very similar values with varying values of η. This is the same for the case of sample medians, while the dispersion of the samples becomes large with increasing value of η. Note that the maximum loss may fluctuate greatly for each set of generated samples due to the large variability of the Weibull distribution.

When $\boldsymbol{X}$ follows the joint distribution described above and the loss function L is as defined in Equation (4.1), we have estimated the tail loss probabilities over various values of threshold by CMC and ITLR methods, respectively. We have chosen the values of threshold so that the tail loss probabilities are in the range $(10^{-6}, 10^{-3})$. Then, the occurrence of $L(\boldsymbol{X})$ larger than the threshold is rare. The adopted values of threshold γ are shown in Table 2. In the CMC estimation, we generated 10^7 samples of $\boldsymbol{X}$ independently from cdf $C_0(F_1(x_1), \ldots, F_{10}(x_{10}))$ for each combination of γ and η. The scheme described in Section 2.1 was applied for the generation of samples of $\boldsymbol{X}$. In the ITLR estimation, we also generated 10^7 importance samples of $\boldsymbol{X}$ for each combination of γ and η. To find the optimal sampling distribution of $\boldsymbol{X}$ and compute the likelihood ratio of a sample, we applied the scheme in Algorithm 2 with $\gamma_i = (i/3)\gamma$, $i = 1, 2, 3$. Through some pilot simulations, we found that such setting of γ_i's is appropriate.

The estimated tail loss probability $\hat{l}$ for each combination of η and γ is given in Table 2. Since ITLR estimates the tail loss probabilities with lower standard error than CMC in all cases, we have chosen the estimated value by ITLR as $\hat{l}$. We can see the tendency that the standard errors of CMC and ITLR estimates become less as γ increases. The tendency is obviously due to the fact that the tail loss probability decreases as γ increases.

When we estimate the tail loss probability using a Monte Carlo simulation scheme, the efficiency of the scheme can be measured by the product of the sample variance of the estimate to the tail loss probability and the simulation time to get the estimate (Glynn and Whitt, 1992; Sak and Hörmann, 2012). We denote the product by time*variance. The smaller value of time*variance implies the better performance. For each combination of η and γ, the time*variance of each estimate by CMC and ITLR, respectively, is shown in Table 2. The time*variance of ITLR estimate to the tail loss probability is larger than that of CMC estimate in the case of $\eta = 0.5$ and $\gamma = 75$. However, in the other cases, ITLR shows the time*variance lower than CMC.

Table 2: The estimated tail loss probabilities and some performance measures for various values of thresholds

$\eta = 0.25$							
γ	$\hat{l}$	simulation time		s.e.		time*variance	
		CMC	ITLR	CMC	ITLR	CMC	ITLR
70	$1.66{\cdot}10^{-3}$	63.5	250.0	$1.29{\cdot}10^{-5}$	$3.49{\cdot}10^{-6}$	$1.06{\cdot}10^{-8}$	$3.04{\cdot}10^{-9}$
80	$2.23{\cdot}10^{-4}$	61.6	240.5	$4.82{\cdot}10^{-6}$	$6.65{\cdot}10^{-7}$	$1.43{\cdot}10^{-9}$	$1.06{\cdot}10^{-10}$
90	$2.66{\cdot}10^{-5}$	63.1	238.8	$1.54{\cdot}10^{-6}$	$9.55{\cdot}10^{-8}$	$1.49{\cdot}10^{-10}$	$2.18{\cdot}10^{-12}$
100	$2.62{\cdot}10^{-6}$	63.2	233.5	$5.20{\cdot}10^{-7}$	$1.16{\cdot}10^{-8}$	$1.71{\cdot}10^{-11}$	$3.16{\cdot}10^{-14}$
$\eta = 0.5$							
γ	$\hat{l}$	simulation time		s.e.		time*variance	
		CMC	ITLR	CMC	ITLR	CMC	ITLR
75	$3.73{\cdot}10^{-3}$	63.1	249.6	$1.93{\cdot}10^{-5}$	$1.41{\cdot}10^{-5}$	$2.38{\cdot}10^{-8}$	$4.99{\cdot}10^{-8}$
90	$3.97{\cdot}10^{-4}$	60.1	240.4	$6.29{\cdot}10^{-6}$	$2.53{\cdot}10^{-6}$	$2.37{\cdot}10^{-9}$	$1.54{\cdot}10^{-9}$
105	$3.26{\cdot}10^{-5}$	63.4	228.3	$1.83{\cdot}10^{-6}$	$2.78{\cdot}10^{-7}$	$2.12{\cdot}10^{-10}$	$1.76{\cdot}10^{-11}$
120	$2.08{\cdot}10^{-6}$	62.3	229.2	$3.87{\cdot}10^{-7}$	$1.83{\cdot}10^{-8}$	$9.35{\cdot}10^{-12}$	$7.68{\cdot}10^{-14}$
$\eta = 0.75$							
γ	$\hat{l}$	simulation time		s.e.		time*variance	
		CMC	ITLR	CMC	ITLR	CMC	ITLR
90	$1.69{\cdot}10^{-3}$	63.5	245.0	$1.30{\cdot}10^{-5}$	$6.13{\cdot}10^{-6}$	$1.08{\cdot}10^{-8}$	$9.21{\cdot}10^{-9}$
110	$1.21{\cdot}10^{-4}$	64.6	235.3	$3.46{\cdot}10^{-6}$	$1.18{\cdot}10^{-6}$	$7.74{\cdot}10^{-10}$	$3.25{\cdot}10^{-10}$
125	$1.39{\cdot}10^{-5}$	64.3	242.5	$1.11{\cdot}10^{-6}$	$1.70{\cdot}10^{-7}$	$7.97{\cdot}10^{-11}$	$7.03{\cdot}10^{-12}$
135	$2.97{\cdot}10^{-6}$	60.8	234.9	$4.58{\cdot}10^{-7}$	$4.52{\cdot}10^{-8}$	$1.27{\cdot}10^{-11}$	$4.81{\cdot}10^{-13}$

Table 3: Time*variance ratio of CMC to ITLR

$\eta = 0.25$		$\eta = 0.5$		$\eta = 0.75$	
γ	time*variance ratio	γ	time*variance ratio	γ	time*variance ratio
70	3.5	75	0.48	90	1.2
80	13.5	90	1.5	110	2.4
90	68.3	105	12.0	125	11.3
100	541.1	120	121.7	135	26.4

Table 3 shows the ratio of the time*variance of CMC estimate to that of ITLR estimate. We can see that the time*variances of ITLR are 3.5 to 541.1 times less than those of CMC with varying thresholds for the case of $\eta = 0.25$. In other words, ITLR is 3.5 to 541.1 times faster than CMC in terms of simulation time to obtain the same estimation error. In the same table, we observe that the estimate for the case of $\eta = 0.75$ shows the similar behavior of the time*variance to that of $\eta = 0.25$. In this case, ITLR is 1.2 to 26.4 times faster than CMC. ITLR shows better performance than CMC for the cases of $\eta = 0.25$ and $\eta = 0.75$. However, in the case of $\eta = 0.5$, the ratio has values from 0.48 to 121.7 according to the values of γ. It depends on the value of threshold which method shows the better performance. We can see from Table 3 that ITLR shows the better performance than the CMC for the case of high threshold, and for the case of weak dependence of X_i's.

5. Conclusion

We consider a static problem to estimate the probability that a strictly increasing function of a multivariate random variable has values over a given threshold, and assume that the exact form of the marginal distributions of the multivariate random variable and that of its copula are known. In order to solve the problem, we applied the inverse transform likelihood ratio method, which was originally proposed for the case of one dimensional random variables. Confining importance sampling distributions to a family of multivariate distributions in which the marginal distribution of each component

follows the beta distribution and the dependence of a multivariate distribution is modeled by the Gaussian copula, we proposed an algorithm to find the pseudo-optimal parameter in an iterative manner, and showed the convergence of the proposed algorithm. Numerical study showed that the proposed scheme is more efficient than the crude Monte Carlo simulation in the case that the threshold is high or in the case that the dependence of each pair of components is weak. However, the performance of the proposed scheme deteriorates in the other case.

If the copula of the nominal distribution is not Gaussian and it shows strong tail dependence, then the losses over a high threshold typically occur when some variables have large values at the same time. Regardless of the correlation matrix, the Gaussian copula is asymptotically independent in both the upper and the lower tails, and the samples generated by the proposed method are unlikely to have any of their components having large values at the same time. This makes it hard to generate large losses frequently, and the proposed method will lose its efficiency. Therefore, when the tail dependence is strong, copulas such as Gumbel or generalized Clayton can be suitable for importance sampling distributions.

As for further research, we suggest to extend the proposed scheme to the case of strong tail dependence, the case of monotonically non-decreasing loss functions, and also the case that the each marginal distribution of an importance sampling distribution is the beta distribution with two free parameters.

Acknowledgements

The authors would like to thank the anonymous reviewers for their comments and suggestions on the first draft of this paper. Their suggestions have greatly improve the quality of the paper. This work was supported by the 2021 sabbatical year research grant of the University of Seoul.

References

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

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

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

Blanchet J and Rojas-Nandayapa L (2011). Efficient simulation of tail probabilities of sums of dependent random variables, *Journal of Applied Probability*, **48**, 147–164.

De Boer P-T, Kroese DP, Mannor S, and Rubinstein RY (2005). A tutorial on the cross-entropy method, *Annals of Operations Research*, **134**, 19–67.

Dupuis P, Leder K, and Wang H (2007). Importance sampling for sums of random variables with regularly varying tails, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **17**, 14-es.

Glynn PW and Iglehart DL (1989). Importance sampling for stochastic simulations, *Management Science*, **35**, 1367–1392.

Glynn PW and Whitt W (1992). The asymptotic efficiency of simulation estimators, *Operations Research*, **40**, 505–520.

Homem-de-Mello T and Rubinstein RY (2002). Estimation of rare event probabilities using cross-entropy. In *Proceedings of the Winter Simulation Conference*, San Diego, CA, USA, 310–319.

Juneja S and Shahabuddin P (2002). Simulating heavy tailed processes using delayed hazard rate

twisting, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **12**, 94–118.
Juneja S and Shahabuddin P (2006). Rare-event simulation techniques: An introduction and recent advances, *Handbooks in Operations Research and Management Science*, **13**, 291–350.
Kroese DP and Rubinstein RY (2004). The transform likelihood ratio method for rare event simulation with heavy tails, *Queueing Systems*, **46**, 317–351.
Lieber D (1998). Rare-event estimation via cross-entropy and importance sampling (Doctoral dissertation), Technion-Israel Institute of Technology, Haifa.
Nelsen RB (2006). *An Introduction to Copulas*, Springer, New York.
Rached NB, Benkhelifa F, Alouini M-S, and Tempone R (2015). A fast simulation method for the log-normal sum distribution using a hazard rate twisting technique. In *Proceedings of 2015 IEEE International Conference on Communications (ICC)*, London, 4259–4264.
Rached NB, Benkhelifa F, Kammoun A, Alouini M-S, and Tempone R (2018). On the generalization of the hazard rate twisting-based simulation approach, *Statistics and Computing*, **28**, 61–75.
Rubinstein RY (1999). The cross-entropy method for combinatorial and continuous optimization, *Methodology and Computing in Applied Probability*, **1**, 127–190.
Rubinstein RY (2002). Cross-entropy and rare events for maximal cut and partition problems, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **12**, 27–53.
Rubinstein RY and Kroese DP (2016). *Simulation and the Monte Carlo Method*, John Wiley & Sons, Hoboken, New Jersey.
Rubinstein RY and Shapiro A (1993). *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, Wiley, Chichester, England.
Sadowsky JS (1993). On the optimality and stability of exponential twisting in Monte Carlo estimation, *IEEE Transactions on Information Theory*, **39**, 119–128.
Sadowsky JS and Bucklew JA (1990). On large deviations theory and asymptotically efficient Monte Carlo estimation, *IEEE Transactions on Information Theory*, **36**, 579–588.
Sak H and Hörmann W (2012). Fast simulations in credit risk, *Quantitative Finance*, **12**, 1557–1569.

Received September 11, 2023; Revised October 23, 2023; Accepted October 24, 2023