

# RARE-EVENT PROBABILITY ESTIMATION WITH CONDITIONAL MONTE CARLO\*

Joshua C. C. Chan<sup>†</sup>

Dirk P. Kroese<sup>‡</sup>

April, 2009

## Abstract

Estimation of rare-event probabilities in high-dimensional settings via importance sampling is a difficult problem due to the degeneracy of the likelihood ratio. In fact, it is generally recommended that Monte Carlo estimators involving likelihood ratios should not be used in such settings. In view of this, we develop efficient algorithms based on conditional Monte Carlo to estimate rare-event probabilities in situations where the degeneracy problem is expected to be severe. By utilizing an asymptotic description of how the rare event occurs, we derive algorithms that involve generating random variables only from the nominal distributions, thus avoiding any likelihood ratio. We consider two settings that occur frequently in applied probability: systems involving bottleneck elements and models involving heavy-tailed random variables. We first consider the problem of estimating  $\mathbb{P}(X_1 + \dots + X_n > \gamma)$ , where  $X_1, \dots, X_n$  are independent but not identically distributed (ind) heavy-tailed random variables. Guided by insights obtained from this model, we then study a variety of more general settings. Specifically, we consider a complex bridge network and a generalization of the widely popular normal copula model used in managing portfolio credit risk, both of which involve hundreds of random variables. We show that the same conditioning idea, guided by an asymptotic description of the way in which the rare event happens, can be used to derive estimators that outperform existing ones.

*Keywords:* bounded relative error, conditional Monte Carlo, cross-entropy, screening, bottlenecks, degeneracy, heavy-tailed distribution, rare event, subexponential distribution, normal copula,  $t$ -copula, credit risks

---

\*Department of Mathematics, University of Queensland, Australia. This work is supported by the Australian Research Council (Discovery Grant DP0985177).

<sup>†</sup>This author would like to acknowledge financial support from the Australian Government and the University of Queensland through IPRS and UQRS scholarships. Email: [chancc@maths.uq.edu.au](mailto:chancc@maths.uq.edu.au)

<sup>‡</sup>Email: [kroese@maths.uq.edu.au](mailto:kroese@maths.uq.edu.au)

# 1 Introduction

Most stochastic models of practical interest involve a large number of random variables interacting in a complex way, where relevant performance measures of the models are not available analytically. In such cases, Monte Carlo methods are indispensable for estimating various quantities of interest, for example the estimation of rare-event probabilities. However, in high-dimensional models the estimation of rare-event probability is a particularly difficult problem, due to the degeneracy properties of the likelihood ratio. In fact, Rubinstein (2007) comments that importance sampling (IS) with high-dimensional likelihood ratio should not be used in those problems. In view of this, instead of considering IS we develop efficient algorithms for estimating rare-event probabilities in high-dimensional settings based on another popular variance reduction idea, namely, conditional Monte Carlo (Asmussen and Glynn, 2007; Rubinstein and Kroese, 2007). By utilizing an asymptotic description of how the rare event occurs, we derive algorithms that involve generating random variables only from the nominal distributions, thus avoiding the problem of choosing a suitable proposal density.

We consider two general settings. The first of these concerns models with *bottleneck elements*. That is, the parameters of those random variables which are critical in the occurrence of the rare event. Although the *screening method* (Rubinstein, 2007) is especially developed for such a setting, it can only mitigate the degeneracy problem. The second general setting involves *heavy-tailed* random variables, in which case the popular exponential change of measure technique (Asmussen and Glynn, 2007) cannot be directly applied. In both settings, however, the asymptotic description of the rare event occurrence can be easily exploited to develop efficient conditional Monte Carlo algorithms.

In terms of specific models, the simplest and most well-studied problem is the estimation of the probability  $\mathbb{P}(S_n > \gamma)$ , where  $S_n = X_1 + \dots + X_n$  and  $X_1, \dots, X_n$  are independent and identically distributed (iid) random variables. This particular problem has attracted considerable attention due to its relevance in queueing theory and various applications in finance and risk management (see, e.g., Cruz, 2002; Asmussen and Glynn, 2007). The literature on the case where the random variables are thin-tailed is extensive, but not until recently has substantial progress been made on the heavy-tailed case. The first logarithmically efficient algorithm in the setting of sums of iid heavy-tailed random variables appears to be Asmussen and Binswanger (1997), which is based on the conditional Monte Carlo method, conditioning on the order statistics  $X_{(1)}, \dots, X_{(n)}$ . Other efficient IS estimators are reported in Asmussen et al. (2000) and Juneja and Shahabuddin (2002), where the former involves order statistics and large deviation theory, and the latter uses the idea of delayed hazard rate twisting. The cross-entropy (CE) method is later utilized to derive another efficient IS estimator in Asmussen et al. (2005). However, the first estimator with bounded relative error is as recent as Asmussen and Kroese (2006), where the authors propose a conditional Monte Carlo estimator based upon the identity  $\mathbb{P}(S_n > \gamma) = n\mathbb{P}(S_n > \gamma, M_n = X_n)$ , where  $M_n = \max(X_1, \dots, X_n)$ . Most recently, Juneja (2007) proposes an estimator with asymptotically vanishing relative error by decomposing  $\mathbb{P}(S_n > \gamma)$  into two parts where the dominant part is available analytically and only the minor part needs to be estimated by simulation.

Although the setting of the sum of iid random variables has numerous applications and is interesting in its own right, many important problems in various fields involve more complex and sophisticated frameworks. In particular, problems often take the form of estimating the

quantity  $\mathbb{P}(S(\mathbf{X}) > \gamma)$ , where  $S(\cdot)$  is some specified function and  $\mathbf{X} = (X_1, \dots, X_n)$  is a vector of random variables that are not necessarily identical nor even independent (see, e.g., Glasserman et al., 2002; Glasserman, 2004; Bassmaboo et al., 2008, among many others). To gain insights into more general settings, we first extend the results in Asmussen and Kroese (2006) to estimate  $\mathbb{P}(X_1 + \dots + X_n > \gamma)$ , when  $X_1, \dots, X_n$  are independent but not identically distributed (ind) random variables. Guided by these insights, we consider a variety of more general settings than the sum of ind random variables. In particular, we consider the bridge network studied in Rubinstein (2007) and the  $t$ -copula model introduced by Bassmaboo et al. (2008), both of which involve hundreds of random variables, so that the problem of degeneracy is expected to be severe for any Monte Carlo estimator utilizing likelihood ratios. We show that the same conditioning idea, motivated by an asymptotic description of the way in which the rare event happens, can be used to derive estimators that outperform existing algorithms. In addition, the proposed algorithms involve generating random variables only from the nominal distributions, thus avoiding the potential degeneracy problem of the likelihood ratio. It is worth noting that for all proposed algorithms the extra computation effort is negligible. Firstly, in terms of random variable generation, the proposed algorithms require less CPU time than the crude Monte Carlo method, as fewer random variables are needed. Secondly, even though the proposed algorithms require evaluations of the distribution function of the random variables, the computation effort is trivial, as analytic formulas (for the case of Pareto and Weibull distributions) or fast routines (for the case of  $t$  distribution) are available. In fact, when analytic formulas are available, evaluations of the distribution function take less than 1 % of the total computation time; when analytic formulas are not available, evaluations take less than 10 % of the total computation time.

The rest of the article is organized as follows. Section 2 generalizes the conditional Monte Carlo estimator of Asmussen and Kroese (2006) to cover the case of sums of ind random variables. The performance of the proposed estimator is demonstrated via various simulation experiments. In particular, we consider the two most important examples of subexponential distributions: the Pareto and Weibull cases. Next, we apply the same conditioning ideas to more complicated settings than sums of ind random variables. Specifically, we first consider the stochastic shortest path within a simple bridge network in Section 3. Then we extend our studies to an  $m \times n$  bridge network in Section 4 and show how the proposed algorithm can be applied to this much more complicated setting with only minor modifications. In Section 5, we investigate a generalization of the *normal copula model*, a popular model in finance for estimating the probability of large portfolio losses. We show that the ideas developed in previous sections can be applied in a straightforward manner.

## 2 Sum of Ind Random Variables

A non-negative random variable  $X$  (or its distribution  $F$ ) is said to be *heavy-tailed* when its exponential moments fail to exist, i.e.,  $\mathbb{E}e^{tX} = \infty$  for  $t > 0$ . Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a vector of ind heavy-tailed random variables with distributions  $F_1, \dots, F_n$ . Consider the problem of estimating the probability of the form

$$\ell = \mathbb{P}(S(\mathbf{X}) > \gamma), \tag{1}$$

where  $S(\mathbf{X}) = X_1 + \dots + X_n$  and  $\gamma > 0$  is large so that  $\ell$  is small. Most often in the literature a narrower class of distributions is considered, namely the class of subexponential distributions, where  $F$  is said to be *subexponential* if

$$\frac{\mathbb{P}(X_1 + \dots + X_n > \gamma)}{\mathbb{P}(X_1 > \gamma)} \rightarrow n \text{ as } \gamma \rightarrow \infty,$$

and  $X_1, \dots, X_n$  are iid with distribution  $F$ . In the numerical examples in this section, we will consider the two most important examples of subexponential distributions: Pareto and Weibull distributions.

In the context of rare event simulation, two measures of efficiency are widely used. An estimator  $Z \equiv Z(\gamma)$  (i.e., some function of the random variables  $\mathbf{X}$ ) is said to be *logarithmically efficient*, or simply *efficient*, if

$$\limsup_{\gamma \rightarrow \infty} \frac{\text{Var}Z(\gamma)}{[\mathbb{E}Z(\gamma)]^{2-\varepsilon}} = 0$$

for all  $\varepsilon > 0$ . A stronger condition is the concept of *bounded relative error*:

$$\limsup_{\gamma \rightarrow \infty} \frac{\text{Var}Z(\gamma)}{[\mathbb{E}Z(\gamma)]^2} < \infty.$$

In the setting of iid heavy-tailed random variables ( $F_1 = \dots = F_n = F$ ), the asymptotic distribution of  $X_1, \dots, X_n$  given the rare event  $\{S(\mathbf{X}) > \gamma\}$  is such that with probability  $1/n$  one of the variables, say  $X_1$ , is distributed according to the conditional distribution  $X_1 | X_1 > \gamma$ , while the other variables  $X_2, \dots, X_n$  are distributed according to the original (nominal) distributions. To utilize this asymptotic description of the rare event, Asmussen and Kroese (2006) write (1) as

$$\ell = n\mathbb{P}(S(\mathbf{X}) > \gamma, M_n = X_n) = n\mathbb{E}\bar{F} \left( \left( \gamma - \sum_{j=1}^{n-1} X_j \right) \vee M_{n-1} \right), \quad (2)$$

where  $M_j = \max_{i \leq j} X_i$ ,  $\bar{F}(x) = 1 - F(x)$  and  $a \vee b$  is a shorthand notation for  $\max(a, b)$ . The identity in (2) suggests the following conditional Monte Carlo estimator:

$$\hat{\ell} = \frac{1}{N} \sum_{i=1}^N n\bar{F} \left( \left( \gamma - \sum_{j=1}^{n-1} X_j^{(i)} \right) \vee M_{n-1}^{(i)} \right), \quad (3)$$

where  $M_k^{(i)} = \max_{j \leq k} X_j^{(i)}$  and  $X_j^{(i)}$   $j = 1, \dots, n-1$ ,  $i = 1, \dots, N$  are iid random variables generated from the common distribution  $F$ . The superior performance of the above conditional MC estimator is illustrated in Asmussen and Kroese (2006) via various simulation studies. In addition, it appears to be the first estimator with bounded relative error for the regularly varying heavy-tailed distributions.

The above estimator can be generalized to cover the ind case. To set the stage, let  $\mathbf{X}_{-i}$  denote the vector  $\mathbf{X}$  with the  $i$ th element removed, i.e.,  $\mathbf{X}_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ , and  $\widetilde{M}_i = \max(\mathbf{X}_{-i})$ . Since for the general ind case we can write the rare-event probability in (1) as

$$\ell = \sum_{i=1}^n \mathbb{P}(S(\mathbf{X}) > \gamma, X_i = M_n) = \sum_{i=1}^n \mathbb{E}\bar{F}_i \left( \left( \gamma - \sum_{j \neq i} X_j \right) \vee \widetilde{M}_i \right), \quad (4)$$

we consider the estimator

$$\widehat{\ell} = \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^n \overline{F}_i \left( (\gamma - \sum_{j \neq i} X_j^{(k)}) \vee \widetilde{M}_i \right), \quad (5)$$

where  $X_j^{(k)}$ ,  $j = 1, \dots, n$ ,  $k = 1, \dots, N$  are random variables with distribution  $F_j$ . In other words, we generate  $N$  copies of  $X_1, \dots, X_n$  from the nominal distributions  $F_1, \dots, F_n$ , and for each  $i = 1, \dots, n$ , we use the sample  $\mathbf{X}_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$  to estimate the  $i$ th component of the summand in (4), i.e., the quantity  $\overline{F}_i \left( (\gamma - \sum_{j \neq i} X_j) \vee \widetilde{M}_i \right)$ . It is important to note that even though we use overlapping samples to compute the various components of the summand, it is obvious that the conditional MC estimator in (5) is still unbiased. Moreover, in order to get  $N$  samples, one only needs a total of  $N \times n$  random draws. In the next subsections we will illustrate the performance of the proposed estimator via various simulated experiments. In particular we consider the two most important examples of subexponential distributions: the Pareto and Weibull distributions.

## 2.1 Pareto Case

The first example of a subexponential distribution is the Pareto distribution, defined by the probability density function (pdf)

$$f(x; \alpha, \lambda) = \alpha \lambda (1 + \lambda x)^{-(\alpha+1)}, \quad x \geq 0, \quad (\alpha > 0, \lambda > 0). \quad (6)$$

We will show that for the Pareto case, the conditional MC estimator (5) has bounded relative error. We first recall a result proved in Asmussen and Kroese (2006).

**Lemma 2.1** *Let  $X_1, \dots, X_n$  be independent but not necessarily identically distributed Pareto random variables with distributions  $F_1, \dots, F_n$ . Then for  $1 \leq i \leq n$ ,*

$$\mathbb{E} \overline{F}_i \left( (\gamma - \sum_{j \neq i} X_j) \vee \widetilde{M}_i \right)^2 \leq c_i \overline{F}_i(\gamma)^2, \quad (7)$$

where the expectation is taken with respect to  $F_1, \dots, F_{i-1}, F_{i+1}, \dots, F_n$ ,  $c_i$  is a constant not dependent on  $\gamma$ , and  $\widetilde{M}_i = \max(\mathbf{X}_{-i})$ .

In the original proof, Asmussen and Kroese (2006) only consider the case where  $X_1, \dots, X_n$  have identical distributions. Nevertheless, the same arguments work even when the distributions are non-identical. With the above result, it is straightforward to show the estimator (5) has bounded relative error.

**Proposition 2.2** *The conditional Monte Carlo estimator (5) has bounded relative error under the Pareto case.*

**Proof** For notational convenience, let  $\delta_i = (\gamma - \sum_{j \neq i} X_j) \vee \widetilde{M}_i$ . The claim now follows from:

$$\begin{aligned} \mathbb{E} \left( \sum_{i=1}^n \overline{F}_i(\delta_i) \right)^2 &= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \overline{F}_i(\delta_i) \overline{F}_j(\delta_j) \leq \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} (\overline{F}_i(\delta_i)^2 + \overline{F}_j(\delta_j)^2) \\ &\leq n \sum_{i=1}^n c_i \overline{F}_i(\gamma)^2 \leq C \sum_{i=1}^n \overline{F}_i(\gamma)^2 \leq C \left( \sum_{i=1}^n \overline{F}_i(\gamma) \right)^2, \end{aligned}$$

where  $C = n \max(c_1, \dots, c_n)$ , and  $c_i, i = 1, \dots, n$  are constants independent of  $\gamma$ .

We provide two numerical examples for the sum of  $n = 10$  ind Pareto random variables  $X_i \stackrel{ind}{\sim} \text{Pareto}(\alpha_i, \lambda_i)$ ,  $i = 1, \dots, n$ . In the first example we set  $\alpha_i = 2 + i/10$  and  $\lambda_i = 1$ , and  $\alpha_i = 2.5$  and  $\lambda_i = 0.5 + i/n$  in the second. We estimate the rare-event probability (1) for various  $\gamma$  using the proposed conditional MC estimator (5), with a sample of size  $N = 100,000$ . For comparison, we also estimate the same probabilities by the *transform likelihood method* (Kroese and Rubinstein, 2004). In particular, we keep  $\alpha_i$  fixed and sequentially update  $\lambda_i$  via the TLR procedure, to obtain the optimal parameters for the IS estimator. We use a sample size of  $N_1 = 100,000$  for the estimation of the optimal parameters and  $N = 100,000$  for the main importance sampling run. The estimated probabilities for the conditional MC estimator and the TLR estimator, together with their relative errors, are reported in Tables 1–2. In the tables we also give the factor of variance reduction for both estimators compared with the crude Monte Carlo estimator. As is apparent from the tables, both estimators offer substantial variance reduction when  $\gamma$  is large. Moreover, the proposed estimator compares favorably with the TLR estimator and outperforms it (in terms of variance reduction) by a factor of  $10^4$  to  $10^7$  in the first example and  $10^4$  to  $10^9$  in the second. Also note that for large  $\gamma$ , the TLR estimator consistently underestimates the true probabilities.

Table 1: Performance of the proposed conditional Monte Carlo for rare-event probability estimation of Pareto ind random variables with  $n = 10$ ,  $\alpha_i = 2 + i/n$  and  $\lambda_i = 1$ . Variance reduction is compared with the crude Monte Carlo estimator.

$\gamma$	TLR			conditional Monte Carlo		
	$\widehat{\ell}$	RE (%)	Var. reduction	$\widehat{\ell}$	RE (%)	Var. reduction
100	$1.89 \times 10^{-4}$	7	10	$1.91 \times 10^{-4}$	0.04	$3.32 \times 10^5$
500	$4.51 \times 10^{-6}$	9	277	$4.74 \times 10^{-6}$	$7.1 \times 10^{-3}$	$4.14 \times 10^8$
1,000	$9.59 \times 10^{-7}$	11	908	$1.01 \times 10^{-6}$	$3.4 \times 10^{-3}$	$8.57 \times 10^9$
5,000	$1.97 \times 10^{-7}$	11	4,360	$2.21 \times 10^{-7}$	$1.5 \times 10^{-3}$	$2.03 \times 10^{11}$

Table 2: Performance of the proposed conditional Monte Carlo estimator for rare-event probability estimation of Pareto ind random variables with  $n = 10$ ,  $\alpha_i = 2.5$  and  $\lambda_i = 0.5 + i/10$ . Variance reduction is compared with the crude Monte Carlo estimator.

$\gamma$	TLR			conditional Monte Carlo		
	$\widehat{\ell}$	RE (%)	Var. reduction	$\widehat{\ell}$	RE (%)	Var. reduction
100	$1.48 \times 10^{-4}$	11	9	$1.46 \times 10^{-4}$	0.05	$3.16 \times 10^5$
500	$1.88 \times 10^{-6}$	22	348	$2.35 \times 10^{-6}$	$5.9 \times 10^{-3}$	$6.21 \times 10^8$
1,000	$3.15 \times 10^{-7}$	26	1,830	$4.10 \times 10^{-7}$	$2.6 \times 10^{-3}$	$1.85 \times 10^{10}$
5,000	$5.82 \times 10^{-8}$	13	5,450	$7.26 \times 10^{-9}$	$4.8 \times 10^{-4}$	$2.97 \times 10^{13}$

## 2.2 Weibull Case

For the second example of a subexponential distribution, we consider the Weibull distribution, which is defined by the pdf

$$f(x; \alpha, \lambda) = \alpha \lambda (\lambda x)^{\alpha-1} e^{-(\lambda x)^\alpha}, \quad x \geq 0, \quad (\alpha > 0, \lambda > 0). \quad (8)$$

Tables 3 and 4 present the results of a numerical study similar to that for the Pareto case. In particular, we consider the sum of  $n = 10$  ind Weibull random variables  $X_i \stackrel{ind}{\sim} \text{Weib}(\alpha_i, \lambda_i)$ ,  $i = 1, \dots, n$ , with  $\alpha_i = 0.25$  and  $\lambda_i = 0.5 + i/n$  in the first numerical example, and  $\alpha_i = 0.75$  and  $\lambda_i = 0.5 + i/n$  in the second. For the proposed conditional MC estimator (5), we again use a sample of size  $N = 100,000$ . For comparison, we also estimate the same probabilities using the *cross-entropy method* (CE) (Rubinstein and Kroese, 2004), performing a two-parameter update ( $\alpha_i$  and  $\lambda_i$ ) in each iteration. We use a sample size of  $N_1 = 100,000$  for the estimation of the optimal parameters and  $N = 100,000$  for the main importance sampling run. The results show that both estimators offer substantial variance reduction when  $\gamma$  is large. It is also worth noting that the conditional MC estimator performs much better when  $\alpha_i$ 's are small ( $\alpha_i = 0.25$ ) than when they are large ( $\alpha_i = 0.75$ ). This is not surprising since the conditional MC estimator is motivated by the asymptotic description for the sum of heavy-tailed random variables, and when the  $\alpha_i$ 's approach 1 the tails of Weibull random variables become thinner and more exponential-like.

Table 3: Performance of the proposed conditional Monte Carlo for rare-event probability estimation of Weibull ind random variables with  $n = 10$ ,  $\alpha_i = 0.25$  and  $\lambda_i = 0.5 + i/10$ . Variance reduction is compared with the crude Monte Carlo estimator.

$\gamma$	CE			conditional Monte Carlo		
	$\widehat{\ell}$	RE (%)	Var. reduction	$\widehat{\ell}$	RE (%)	Var. reduction
10,000	$5.97 \times 10^{-4}$	2.1	36	$5.96 \times 10^{-4}$	0.06	54,200
20,000	$9.68 \times 10^{-5}$	3.1	107	$9.64 \times 10^{-5}$	0.04	$6.47 \times 10^5$
50,000	$5.36 \times 10^{-6}$	4.9	791	$5.32 \times 10^{-6}$	0.02	$4.33 \times 10^7$
100,000	$3.83 \times 10^{-7}$	6.0	7,190	$3.81 \times 10^{-7}$	0.01	$2.22 \times 10^9$

Table 4: Performance of the proposed conditional Monte Carlo estimator for rare-event probability estimation of Pareto ind random variables with  $n = 10$ ,  $\alpha_i = 0.75$  and  $\lambda_i = 0.5 + i/10$ . Variance reduction is compared with the crude Monte Carlo estimator.

$\gamma$	CE			conditional Monte Carlo		
	$\widehat{\ell}$	RE (%)	Var. reduction	$\widehat{\ell}$	RE (%)	Var. reduction
40	$7.99 \times 10^{-4}$	0.95	139	$7.96 \times 10^{-4}$	0.98	131
50	$8.10 \times 10^{-5}$	1.2	834	$8.19 \times 10^{-5}$	1.4	614
70	$1.21 \times 10^{-6}$	2.4	14,500	$1.21 \times 10^{-6}$	2.5	13,100
100	$4.64 \times 10^{-9}$	5.8	$6.79 \times 10^5$	$4.62 \times 10^{-9}$	2.0	$5.19 \times 10^6$

### 3 A Simple Bridge Network

We now explore various more general settings than  $\mathbb{P}(X_1 + \dots + X_n > \gamma)$ . We will demonstrate that the same conditioning ideas, motivated by the asymptotic description of how the rare event occurs, can be used to derive efficient estimators in those settings. In this section, we first consider the problem of estimating the probability that the shortest path from node  $A$  to node  $B$  in the network of Figure 1 has a length of at least  $\gamma$ , where the random lengths  $X_1, \dots, X_5$  of the links are assumed to be independent.

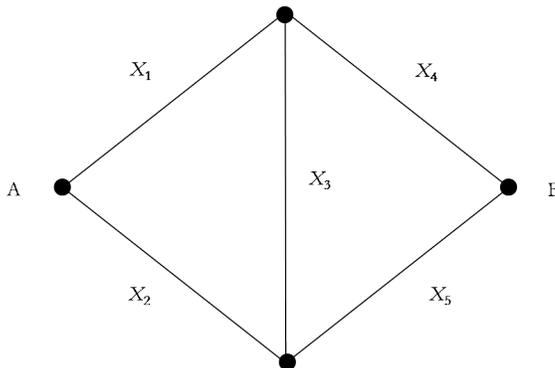


Figure 1: A simple bridge network.

Defining  $\mathbf{X} = (X_1, \dots, X_5)$  and

$$S(\mathbf{X}) = \min\{X_1 + X_4, X_1 + X_3 + X_5, X_2 + X_5, X_2 + X_3 + X_4\}, \quad (9)$$

we are interested in estimating the probability

$$\ell = \mathbb{P}(S(\mathbf{X}) > \gamma). \quad (10)$$

As a motivating example, let us first consider the special case where  $X_i \sim \text{Exp}(\lambda_i)$ ,  $i = 1, \dots, 5$  with pdf  $f(x; \theta) = \theta e^{-\theta x}$  and  $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = (1, 1, 3, 2, 10)$ . Note that in this particular example, the parameter values  $\lambda$  are chosen in such a way that the rare event  $\{S(\mathbf{X}) > \gamma\}$

happens primarily when both  $X_1$  and  $X_2$  are large, while the other variables are in some sense ‘typical’ (in the language of Rubinstein (2007), only  $\lambda_1$  and  $\lambda_2$  are *bottleneck elements*). Put differently, if one derives the conditional distribution given the rare event  $\{S(\mathbf{X}) > \gamma\}$ , only  $X_1$  and  $X_2$  are affected significantly by the conditioning. Therefore, the general principle in importance sampling — of taking the proposal density to be as close as possible to the conditional distribution given the rare event — would suggest twisting only the distributions of  $X_1$  and  $X_2$ , while leaving the others unchanged. Further variance reduction can be achieved by ‘integrating out’  $X_1$  and  $X_2$  via conditional Monte Carlo. More specifically, note that given  $X_i = x_i$ ,  $i = 3, 4, 5$ , we can write equation (10) as follows:

$$\begin{aligned} & \mathbb{P}(S(\mathbf{X}) > \gamma | X_3 = x_3, X_4 = x_4, X_5 = x_5) \\ &= \mathbb{P}(X_1 > (\gamma - x_4) \vee (\gamma - x_3 - x_5), X_2 > (\gamma - x_5) \vee (\gamma - x_3 - x_4)) \\ &= \bar{F}_1((\gamma - x_4) \vee (\gamma - x_3 - x_5)) \bar{F}_2((\gamma - x_5) \vee (\gamma - x_3 - x_4)), \end{aligned}$$

where  $a \vee b = \max(a, b)$ ,  $\bar{F}_i(x) = 1 - F_i(x)$  and  $F_i$  is the distribution of  $X_i$ . Therefore, we consider the following conditional Monte Carlo estimator

$$\hat{\ell}_s = \frac{1}{N} \sum_{k=1}^N \bar{F}_1 \left( (\gamma - X_4^{(k)}) \vee (\gamma - X_3^{(k)} - X_5^{(k)}) \right) \bar{F}_2 \left( (\gamma - X_5^{(k)}) \vee (\gamma - X_3^{(k)} - X_4^{(k)}) \right), \quad (11)$$

where  $X_i^{(k)} \stackrel{\text{ind}}{\sim} \text{Exp}(\lambda_i)$ ,  $i = 3, 4, 5$ ,  $k = 1 \dots, N$ . For comparison, we also estimate the rare-event probability  $\ell$  using the *screening method* (Rubinstein, 2007; Rubinstein and Kroese, 2007), which is shown to outperform the standard cross-entropy method and *variance minimization method* (Rubinstein and Kroese, 2007, pp.132-136) in certain settings. In the estimation, we only twist the distributions of  $X_1$  and  $X_2$ . The proposal density is of the form of a product of exponential densities where the parameters are determined by the CE method. We use a sample size of  $N_1 = 10,000$  to estimate the optimal parameters for the proposal density and  $N = 100,000$  for the main importance sampling run. For the conditional MC estimator we also use a sample size of  $N = 100,000$ . Table 5 reports the CE with screening (CE-SCR) and conditional MC estimates and their corresponding relative errors for the case  $X_i \sim \text{Exp}(\lambda_i)$ ,  $i = 1, \dots, 5$  and  $\lambda = (1, 1, 3, 2, 10)$ .

Table 5: Performance of the proposed conditional Monte Carlo estimator for the simple bridge example where the random lengths are exponentially distributed random variables with  $\lambda = (1, 1, 3, 2, 10)$ . Variance reduction is compared with the crude Monte Carlo estimator.

$\gamma$	CE with screening			conditional Monte Carlo		
	$\hat{\ell}$	RE (%)	Var. reduction	$\hat{\ell}$	RE (%)	Var. reduction
4	$4.37 \times 10^{-4}$	1.9	64	$4.33 \times 10^{-4}$	0.06	56,800
6	$7.94 \times 10^{-6}$	2.8	1,620	$7.92 \times 10^{-6}$	0.06	$3.18 \times 10^6$
8	$1.46 \times 10^{-7}$	3.7	49,300	$1.45 \times 10^{-7}$	0.06	$1.70 \times 10^8$
10	$2.65 \times 10^{-9}$	4.4	$1.86 \times 10^6$	$2.66 \times 10^{-9}$	0.06	$9.38 \times 10^9$

As a second numerical example, we consider the same network but with the path lengths  $X_1, \dots, X_5$  being independent  $\text{Weib}(\alpha, \lambda_i)$  random variables with  $\alpha = 0.2$  and  $\lambda = (1, 1, 3, 2, 10)$ . The proposal density for the screening method is of the form of a product of Weibull densities

with  $\alpha$  fixed at 0.2, and  $\lambda$  determined by the CE method. We use the same sample size as in the previous example, and the results are reported in Table 6. In both cases, the proposed conditional MC estimator performs remarkably well and outperforms the CE-SCR estimator (in terms of variance reduction) by a factor of 900 to 5,000 in the light-tailed case and  $10^6$  to  $10^8$  in the heavy-tailed one. It is also evident from Tables 5 and 6 that the proposed algorithm performs much better in the heavy-tailed case in terms of much lower relative errors.

Table 6: Performance of the proposed conditional Monte Carlo estimator for the simple bridge example where the random lengths are Weibull distributed random variables with  $\alpha = 0.2$  and  $\lambda = (1, 1, 3, 2, 10)$ . Variance reduction is compared with the crude Monte Carlo estimator.

$\gamma$	CE with screening			conditional Monte Carlo		
	$\hat{\ell}$	RE (%)	Var. reduction	$\hat{\ell}$	RE (%)	Var. reduction
5,000	$1.71 \times 10^{-5}$	2.6	881	$1.70 \times 10^{-5}$	$2.4 \times 10^{-3}$	$1.02 \times 10^9$
10,000	$3.35 \times 10^{-6}$	2.9	3,550	$3.31 \times 10^{-6}$	$1.1 \times 10^{-3}$	$2.71 \times 10^{10}$
20,000	$5.02 \times 10^{-7}$	3.3	17,800	$5.07 \times 10^{-7}$	$5.3 \times 10^{-4}$	$7.03 \times 10^{11}$
50,000	$2.69 \times 10^{-8}$	4.0	$2.31 \times 10^5$	$2.74 \times 10^{-8}$	$3.0 \times 10^{-4}$	$3.94 \times 10^{13}$

### 3.1 Independent and Identically Distributed Heavy-tailed Case

We now consider the case where the random path lengths are iid heavy-tailed random variables with common distribution  $F$ . Recall that in the previous two examples the rare event  $\{S(\mathbf{X}) > \gamma\}$  occurs primarily when both  $X_1$  and  $X_2$  are large, due to the bottleneck parameter  $\lambda_1$  and  $\lambda_2$ . In other words, only  $X_1$  and  $X_2$  are much affected by conditioning on the rare event  $\{S(\mathbf{X}) > \gamma\}$ , while the other variables are not. Therefore, substantial variance reduction is achieved by ‘integrating out’  $X_1$  and  $X_2$ . For the case of iid heavy-tailed random variables, the rare event  $\{S(\mathbf{X}) > \gamma\}$  occurs when either both  $X_1$  and  $X_2$  or both  $X_4$  and  $X_5$  are large. To utilize the asymptotic description of the rare event under the iid case, consider the identity

$$\mathbb{P}(S(\mathbf{X}) > \gamma) = 2\mathbb{P}(S(\mathbf{X}) > \gamma, X_1 > X_4, X_2 > X_5) + 2\mathbb{P}(S(\mathbf{X}) > \gamma, X_1 > X_4, X_2 < X_5). \quad (12)$$

Hence, given  $X_i = x_i$ ,  $i = 3, 4, 5$ , the rare-event probability can be written as

$$\begin{aligned} & \mathbb{P}(S(\mathbf{X}) > \gamma | X_3 = x_3, X_4 = x_4, X_5 = x_5) \\ &= 2\mathbb{P}(X_1 > \gamma_1 \vee \gamma_2 \vee x_4) [\mathbb{P}(X_2 > \gamma_3 \vee \gamma_4 \vee x_5) + \mathbb{P}(x_5 > X_2 > \gamma_3 \vee \gamma_4)], \end{aligned}$$

where

$$\gamma_1 = \gamma - x_4, \quad \gamma_2 = \gamma - x_3 - x_5, \quad \gamma_3 = \gamma - x_5, \quad \gamma_4 = \gamma - x_3 - x_4.$$

Therefore, we consider the estimator

$$\hat{\ell} = \frac{2}{N} \sum_{k=1}^N \bar{F}(\gamma_1^{(k)} \vee \gamma_2^{(k)} \vee X_4^{(k)}) \left( \bar{F}(\gamma_3^{(k)} \vee \gamma_4^{(k)} \vee X_5^{(k)}) + (\bar{F}(\gamma_3^{(k)} \vee \gamma_4^{(k)}) - \bar{F}(X_5^{(k)})) \vee 0 \right), \quad (13)$$

where  $X_i^{(k)} \stackrel{iid}{\sim} F$ ,  $i = 3, 4, 5$ ,  $k = 1, \dots, N$ .

As a numerical example, consider the simple bridge network where the  $X_i \stackrel{iid}{\sim} \text{Weib}(\alpha, \lambda)$ ,  $i = 1, \dots, 5$  with  $\alpha = 0.2$  and  $\lambda = 1$ . For comparison, we also estimate the rare-event probability  $\ell$  via the CE method. For the CE method, we use a sample size of  $N_1 = 50,000$  for the estimation of the optimal parameters and  $N = 100,000$  for the main importance sampling run. For the conditional MC estimator, we use the same sample size of  $N = 100,000$ . Table 7 presents the CE and conditional MC estimates and the corresponding relative errors. Variance reduction is compared with the crude Monte Carlo estimator. The proposed estimator performs remarkably well in terms of variance reduction and compares most favorably with the CE estimator.

Table 7: Performance of the proposed conditional Monte Carlo estimator (13) for the simple bridge example where the random lengths are iid Weibull distributed random variables with  $\alpha = 0.2$  and  $\lambda = 1$ . Variance reduction is compared with the crude Monte Carlo estimator.

$\gamma$	CE			conditional Monte Carlo		
	$\hat{\ell}$	RE (%)	Var. reduction	$\hat{\ell}$	RE (%)	Var. reduction
5,000	$3.42 \times 10^{-5}$	5.5	97	$3.41 \times 10^{-5}$	0.37	21,000
10,000	$6.75 \times 10^{-6}$	6.9	311	$6.64 \times 10^{-6}$	0.35	$1.24 \times 10^5$
20,000	$1.03 \times 10^{-6}$	7.4	1,790	$1.02 \times 10^{-6}$	0.29	$5.01 \times 10^5$
50,000	$5.21 \times 10^{-8}$	10	18,600	$5.49 \times 10^{-8}$	$3.3 \times 10^{-3}$	$1.68 \times 10^{11}$

### 3.2 Independent but Not Identically Distributed Heavy-tailed Case

We now generalize the conditional MC estimator (13) to the ind case. Notice that if, say,  $X_1$  and  $X_2$  have much heavier tails than  $X_4$  and  $X_5$ , then the rare event  $\{S(\mathbf{X}) > \gamma\}$  happens primarily when both  $X_1$  and  $X_2$  are large, and we only need to consider the estimator  $\hat{\ell}_s$  in (11). Thus the relevant setting here is when  $X_1, \dots, X_5$  are non-identically distributed but with similar tail indices. Observe that for the ind case, identity (12) becomes

$$\begin{aligned} \mathbb{P}(S(\mathbf{X}) > \gamma) &= \mathbb{P}(S(\mathbf{X}) > \gamma, X_1 > X_4, X_2 > X_5) + \mathbb{P}(S(\mathbf{X}) > \gamma, X_1 > X_4, X_2 < X_5) \\ &\quad + \mathbb{P}(S(\mathbf{X}) > \gamma, X_4 > X_1, X_2 > X_5) + \mathbb{P}(S(\mathbf{X}) > \gamma, X_4 > X_1, X_5 > X_2), \end{aligned}$$

which suggests the estimator

$$\begin{aligned} \hat{\ell} &= \frac{1}{N} \sum_{k=1}^N \left[ \bar{F}_1(\gamma_1^{(k)} \vee \gamma_2^{(k)} \vee X_4^{(k)}) \left( \bar{F}_2(\gamma_3^{(k)} \vee \gamma_4^{(k)} \vee X_5^{(k)}) + (\bar{F}_2(\gamma_3^{(k)} \vee \gamma_4^{(k)}) - \bar{F}_2(X_5^{(k)})) \vee 0 \right) \right. \\ &\quad \left. + \bar{F}_4(\gamma_5^{(k)} \vee \gamma_6^{(k)} \vee X_1^{(k)}) \left( \bar{F}_5(\gamma_7^{(k)} \vee \gamma_8^{(k)} \vee X_2^{(k)}) + (\bar{F}_5(\gamma_7^{(k)} \vee \gamma_8^{(k)}) - \bar{F}_5(X_2^{(k)})) \vee 0 \right) \right], \quad (14) \end{aligned}$$

where

$$\begin{aligned} \gamma_1^{(k)} &= \gamma - X_4^{(k)}, \quad \gamma_2^{(k)} = \gamma - X_3^{(k)} - X_5^{(k)}, \quad \gamma_3^{(k)} = \gamma - X_5^{(k)}, \quad \gamma_4^{(k)} = \gamma - X_3 - X_4^{(k)}, \\ \gamma_5^{(k)} &= \gamma - X_1^{(k)}, \quad \gamma_6^{(k)} = \gamma - X_2^{(k)} - X_3^{(k)}, \quad \gamma_7^{(k)} = \gamma - X_2^{(k)}, \quad \gamma_8^{(k)} = \gamma - X_1 - X_3^{(k)}, \end{aligned}$$

and  $X_i^{(k)} \sim F_i$ ,  $i = 1, \dots, 5$ ,  $k = 1, \dots, N$ . It is also obvious that if  $X_i$ ,  $i = 1, \dots, 5$  are iid random variables, the above estimator (14) reduces to (13).

As a numerical example, consider the simple bridge network where  $X_i \stackrel{ind}{\sim} \text{Weib}(\alpha, \lambda_i)$ ,  $i = 1, \dots, 5$  with  $\alpha = 0.2$  and  $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = (1.2, 0.8, 1, 0.9, 1.1)$ . Table 8 presents the CE and conditional Monte Carlo estimates, as well as the corresponding relative errors. Variance reduction is compared with the crude Monte Carlo estimator. Again the proposed algorithm gives more accurate estimates and outperforms the CE estimator.

Table 8: Performance of the proposed conditional Monte Carlo estimator (14) for the simple bridge example where the random lengths are iid Weibull distributed random variables with  $\alpha = 0.2$  and  $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = (1.2, 0.8, 1, 0.9, 1.1)$ . Variance reduction is compared with the crude Monte Carlo estimator.

$\gamma$	CE			conditional Monte Carlo		
	$\hat{\ell}$	RE (%)	Var. reduction	$\hat{\ell}$	RE (%)	Var. reduction
5,000	$3.53 \times 10^{-5}$	4.5	137	$3.50 \times 10^{-5}$	0.28	37,800
10,000	$6.73 \times 10^{-6}$	6.0	415	$6.82 \times 10^{-6}$	0.30	$1.67 \times 10^5$
20,000	$1.04 \times 10^{-6}$	8.4	1,350	$1.06 \times 10^{-6}$	0.51	$3.65 \times 10^5$
50,000	$5.72 \times 10^{-8}$	11	14,100	$5.69 \times 10^{-8}$	$2.3 \times 10^{-3}$	$3.42 \times 10^{11}$

## 4 An $m \times n$ Bridge System

In this section we extend our investigation to the  $m \times n$  bridge system in Figure 2, where  $m$  denotes the number of parallel paths and  $n$  represents the number of bridges in each path. Observe that for the  $m \times n$  bridge system, there are altogether  $5mn$  random variables, and this high dimensionality makes it a particularly difficult problem for simulation. In fact, Rubinstein (2007) comments that IS should not be used in high-dimensional problems because of the degeneracy properties of the likelihood ratio. However, in the presence of bottleneck elements, Rubinstein (2007) shows substantial variance reduction can still be achieved via the screening method. This problem is also investigated in Botev and Kroese (2008), who propose a novel simulation approach that circumvents the degeneracy problem by avoiding the likelihood ratio altogether. In the following we will discuss a conditional MC estimator that outperforms (in terms of variance reduction) both approaches.

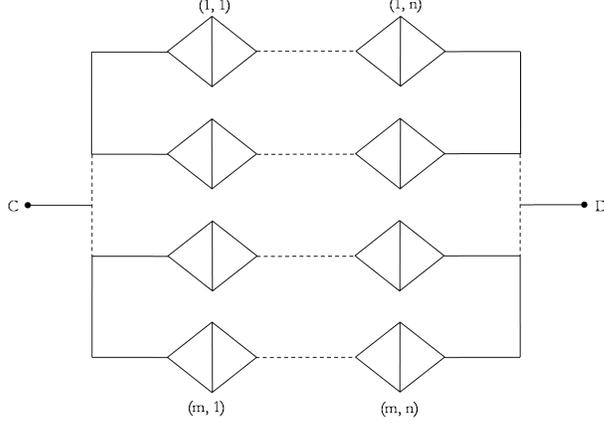


Figure 2: An  $m \times n$  bridge network.

Denote the lengths of the edges within the  $(i, j)$ th bridge by  $X_{ij1}, \dots, X_{ij5}$ . We are interested in estimating the rare-event probability that the shortest path from node  $C$  to node  $D$  in the network has a length of at least  $\gamma$ . Define

$$S(\mathbf{X}) = \min\{Y_{11} + \dots + Y_{1n}, \dots, Y_{m1} + \dots + Y_{mn}\},$$

where

$$Y_{ij} = \min\{X_{ij1} + X_{ij4}, X_{ij1} + X_{ij3} + X_{ij5}, X_{ij2} + X_{ij5}, X_{ij2} + X_{ij3} + X_{ij4}\}.$$

We are interested in estimating  $\ell = \mathbb{P}(S(\mathbf{X}) > \gamma)$  for some large constant  $\gamma$ . We note that the rare event  $\{S(\mathbf{X}) > \gamma\}$  happens if and only if  $Y_{i1} + \dots + Y_{in} > \gamma$ , for  $i = 1, \dots, m$ . Hence, it suffices to derive conditional Monte Carlo estimators for estimating the probabilities

$$\ell_i = \mathbb{P}(Y_{i1} + \dots + Y_{in} > \gamma), \quad i = 1, \dots, m,$$

because

$$\ell = \prod_{i=1}^m \ell_i.$$

As a motivating example, consider the case where  $X_{ijk} \sim \text{Exp}(\lambda_{ijk})$ , with  $\lambda_{111} = \lambda_{112} = \lambda_{211} = \lambda_{212} = \lambda_{311} = \lambda_{312} = 1$ , while the remaining parameters are set to 4. This particular example was studied by Rubinstein (2007) and Rubinstein and Kroese (2007) for illustrating the screening method. Specifically, the parameters are chosen in such a way that the rare event  $\{S(\mathbf{X}) > \gamma\}$  occurs primarily when all  $X_{i1j}$ ,  $i = 1, 2, 3$ ,  $j = 1, 2$  are large. In other words, only  $\lambda_{i1j}$ ,  $i = 1, 2, 3$ ,  $j = 1, 2$  are bottleneck elements and one only needs to twist the distributions of the corresponding random variables. For our conditional Monte Carlo estimator, since only  $X_{i1j}$ ,  $i = 1, 2, 3$ ,  $j = 1, 2$  are affected by conditioning on the rare event  $\{S(\mathbf{X}) > \gamma\}$ , we can simulate other variables from their nominal distributions, and only ‘integrate out’  $X_{i1j}$ ,  $i = 1, 2, 3$ ,  $j = 1, 2$  via conditional Monte Carlo methods. To this end, let us consider the estimator

$$\hat{\ell}_i = \frac{1}{N} \sum_{k=1}^N \bar{F}_{i11}((\tilde{\gamma}^{(k)} - X_{i14}^{(k)}) \vee (\tilde{\gamma}^{(k)} - X_{i13}^{(k)} - X_{i15}^{(k)})) \bar{F}_{i12}((\tilde{\gamma}^{(k)} - X_{i15}^{(k)}) \vee (\tilde{\gamma}^{(k)} - X_{i13}^{(k)} - X_{i14}^{(k)})),$$

where  $\tilde{\gamma}^{(k)} = \max(\gamma - \sum_{j=2}^n Y_{ij}^{(k)}, 0)$  and  $X_{ijk}^{(l)} \stackrel{iid}{\sim} \text{Exp}(4)$ . Then we deliver the unbiased estimator (since each  $\hat{\ell}_i$  is estimated by an independent sample):

$$\hat{\ell} = \prod_{i=1}^n \hat{\ell}_i. \quad (15)$$

The variance of the above estimator can be estimated readily by the formula (Goodman, 1962):

$$\widehat{\text{Var}} \left( \prod_{i=1}^n X_i \right) = \prod_{i=1}^n \hat{\mu}_i^2 \left( \prod_{i=1}^n \left( \frac{\hat{\sigma}_i^2}{\hat{\mu}_i^2} + 1 \right) - 1 \right), \quad (16)$$

where  $\hat{\mu}_i$  and  $\hat{\sigma}_i^2$  are respectively the sample mean and variance of  $X_i$ .

As a numerical example, we use a sample size of  $N = 400,000$  for the conditional Monte Carlo estimator. As a comparison, we also estimate the rare-event probability via the CE method with screening (CE-SCR), where only the aforementioned six bottleneck elements are tilted. For the CE-SCR estimator, we use a sample size of  $N_1 = 100,000$  for estimation of the optimal parameters and  $N = 400,000$  for the main importance sampling run. Table 9 presents the conditional Monte Carlo and CE-SCR estimates for the  $3 \times 10$  bridge system, as well as the corresponding relative errors. As is apparent from the table, the proposed conditional Monte Carlo estimator performs remarkably well even in this complex example with 150 random variables, and it outperforms (in terms of variance reduction) the CE-SCR estimator by a factor of 1,000 to 40,000.

Table 9: Performance of the proposed conditional Monte Carlo estimator for the  $3 \times 10$  bridge network where the random lengths are ind exponentially distributed random variables with  $\lambda_{i1j} = 1$ ,  $i = 1, 2, 3$ ,  $j = 1, 2$ , while the remaining parameters are set to 4. Variance reduction is compared with the crude Monte Carlo estimator.

$\gamma$	CE with screening			conditional Monte Carlo		
	$\hat{\ell}$	RE (%)	Var. reduction	$\hat{\ell}$	RE (%)	Var. reduction
5.0	$2.3 \times 10^{-5}$	20	3	$2.22 \times 10^{-5}$	0.60	3,100
5.5	$1.2 \times 10^{-6}$	26	31	$1.17 \times 10^{-6}$	0.71	42,000
6.0	$5.7 \times 10^{-8}$	50	173	$5.97 \times 10^{-8}$	0.82	$6.16 \times 10^5$
6.5	$4.0 \times 10^{-9}$	55	276	$3.00 \times 10^{-9}$	0.86	$1.13 \times 10^7$

#### 4.1 Independent and Identically Distributed Case

It is important to realize that the reason why the above conditional Monte Carlo estimator performs well is that the problem is set up in such a way that the rare event  $\{S(\mathbf{X}) > \gamma\}$  occurs primarily when all  $X_{i1j}$ ,  $i = 1, 2, 3$ ,  $j = 1, 2$  are large. In this section we consider a more general setting in which the asymptotic description of the rare event occurrence is different, and similar considerations (and additional conditioning) as in Section 3.1 are needed. Specifically, let us consider the case where  $Y_{ij}$ —the shortest lengths of the  $(i, j)$ th bridge—are iid random variables; in particular, we assume  $X_{ijk} \sim F_k$ ,  $k = 1, \dots, 5$ , where  $F_k$  is the distribution of a heavy-tailed random variable. Under this setting, the rare event  $\{Y_{i1} + \dots + Y_{in} > \gamma\}$  occurs

primarily when one of the  $Y_{ij}$ ,  $j = 1, \dots, n$ , say,  $Y_{in}$  is large, while the others are ‘typical’. With this in mind, we consider the identity

$$\ell_i = \mathbb{P}(Y_{i1} + \dots + Y_{in} > \gamma) = n\mathbb{P}(Y_{i1} + \dots + Y_{in} > \gamma, Y_{in} = M_{in}) \quad (17)$$

$$= n\mathbb{P}\left(Y_{in} > \left(\gamma - \sum_{j=1}^{n-1} Y_{ij}\right) \vee M_{i,n-1}\right), \quad (18)$$

where  $M_{ij} = \max(Y_{i1}, \dots, Y_{ij})$ . Given  $Y_{i1}, \dots, Y_{i,n-1}$ , the problem is now reduced to the case of a simple bridge with ind random lengths. In other words, we can estimate  $\ell_i$  as follows: first generate  $N$  copies of  $Y_{i1}, \dots, Y_{i,n-1}$  (by drawing  $X_{ij1}, \dots, X_{ij5}$ ). Then use the estimator (14) to estimate  $\ell_i$  simply by replacing  $\gamma$  by  $(\gamma - \sum_{j=1}^{n-1} Y_{ij}) \vee M_{i,n-1}$  (and multiplying the quantity by  $n$ ).

As a numerical example, consider a  $5 \times 20$  bridge network where the random path lengths  $X_{ijk} \stackrel{iid}{\sim} \text{Weib}(\alpha, \lambda)$ ,  $i = 1, \dots, 5$ ,  $j = 1, \dots, 20$ ,  $k = 1, \dots, 5$  with  $\alpha = 0.2$  and  $\lambda = 1$ . Note that in this example there are a total of 500 variables and estimation of the optimal parameters via the CE method is not practical (and the screening method does not apply). Thus we do not report the CE estimates. For the conditional Monte Carlo estimator, we use a sample size of  $N = 100,000$ . The conditional Monte Carlo estimates and the corresponding relative errors are reported in Table 10. Variance reduction is compared with the crude Monte Carlo estimator. Observe that even in this fairly large example the proposed conditional Monte Carlo estimator performs quite well.

Table 10: Performance of the proposed conditional Monte Carlo estimator for the  $5 \times 20$  bridge network where the random lengths are iid Weibull random variables with  $\alpha = 0.2$  and  $\lambda = 1$ . Variance reduction is compared with the crude Monte Carlo estimator.

$\gamma$	$\widehat{\ell}$	RE (%)	Var. reduction
200	$3.30 \times 10^{-4}$	0.82	447
300	$2.43 \times 10^{-5}$	0.84	5,830
500	$5.89 \times 10^{-7}$	0.86	$2.30 \times 10^5$
1000	$1.98 \times 10^{-9}$	0.89	$6.45 \times 10^7$

## 4.2 Independent but Not Identically Distributed Case

Finally we consider the most general case where  $Y_{ij}$ — the length of the shortest path of the  $(i, j)$ th bridge— are ind random variables; in particular, we assume  $X_{ijk} \stackrel{iid}{\sim} F_{ijk}$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ ,  $k = 1, \dots, 5$ , where  $F_{ijk}$  is the distribution of a heavy-tailed random variable. As in the previous case, we write  $\ell_i$  by conditioning on certain events to take advantage of the asymptotic description of the rare event occurrence. To this end, consider the

identity

$$\ell_i = \mathbb{P}(Y_{i1} + \dots + Y_{in} > \gamma) = \sum_{j=1}^n \mathbb{P}(Y_{i1} + \dots + Y_{in} > \gamma, Y_{ij} = \widetilde{M}_{ij}) \quad (19)$$

$$= \sum_{j=1}^n \mathbb{P} \left( Y_{ij} > \left( \gamma - \sum_{k \neq j} Y_{ik} \right) \vee \widetilde{M}_{ij} \right), \quad (20)$$

where  $\widetilde{M}_{ij} = \max(\mathbf{Y}_{-ij})$  and  $\mathbf{Y}_{-ij} = (Y_{i1}, \dots, Y_{i,j-1}, Y_{i,j+1}, \dots, Y_{in})$ . Given  $\mathbf{Y}_{-ij}$ , the problem is now reduced to the case of a simple bridge with ind random lengths. Therefore, we can estimate  $\ell_i$  as follows: first generate  $N$  copies of  $Y_{i1}, \dots, Y_{in}$  (by drawing  $X_{ij1}, \dots, X_{ij5}$ ). For  $l = 1, \dots, n$ , use the sample  $\mathbf{Y}_{-il}$  and estimator (14) to estimate the  $l$ th component of the summand of  $\ell_i$ , simply by replacing  $\gamma$  by  $(\gamma - \sum_{k \neq j} Y_{ik}) \vee \widetilde{M}_{ij}$ . It is essential to note that even though we use the same sample to estimate the  $n$  components of  $\ell_i$ , the proposed estimator is unbiased.

To illustrate the effectiveness of the above estimator, we consider a  $5 \times 20$  bridge network where the random path lengths are ind Weibull random variables  $X_{ijk} \stackrel{ind}{\sim} \text{Weib}(\alpha_{ijk}, \lambda_{ijk})$ , with  $\alpha_{1j1} = 0.2 + j/100$ , and  $\lambda_{1j4} = 1 + j/10$ , while the other parameters  $\alpha_{ijk}$  and  $\lambda_{ijk}$  are respectively 0.2 and 1. In the numerical example we use a sample size of  $N = 100,000$ , and Table 11 presents the conditional Monte Carlo estimates and the corresponding relative errors. As can be seen from the table, substantial variance reduction is achieved even in this fairly complex example with 500 non-identically distributed variables.

Table 11: Performance of the proposed conditional Monte Carlo estimator for the  $5 \times 20$  bridge network where the random lengths are ind Weibull random variables with  $\alpha_{1j1} = 0.2 + j/100$ , and  $\lambda_{1j4} = 1 + j/10$  while other  $\alpha_{ijk}$  and  $\lambda_{ijk}$  are respectively 0.2 and 1. Variance reduction is compared with the crude Monte Carlo estimator.

$\gamma$	$\widehat{\ell}$	RE (%)	Var. reduction
200	$1.51 \times 10^{-4}$	0.16	6,730
300	$1.00 \times 10^{-5}$	0.15	$1.10 \times 10^5$
500	$2.16 \times 10^{-7}$	0.15	$5.63 \times 10^6$
1000	$6.44 \times 10^{-10}$	0.13	$2.31 \times 10^9$

## 5 The $t$ -copula Model

In this section, we turn our attention to study a popular model in finance for estimating the probability of large portfolio losses. Suppose we have a portfolio of loans consisting of  $n$  obligors, each of which has a given probability of defaulting, which is denoted as  $p_i \in (0, 1)$ ,  $i = 1, \dots, n$ . Introduce a vector of underlying latent variables  $\mathbf{X} = (X_1, \dots, X_n)$  such that the  $i$ th obligor defaults if  $X_i$  exceeds some given threshold level  $x_i$ , i.e.,  $p_i = \mathbb{P}(X_i > x_i)$ . The portfolio loss incurred from defaults is thus given by

$$S(\mathbf{X}) = c_1 I_{\{X_1 > x_1\}} + \dots + c_n I_{\{X_n > x_n\}}, \quad (21)$$

where  $I_{\{\cdot\}}$  denotes the indicator function, and  $c_i$  is the monetary loss associated with the default of the  $i$ th obligor. The primary interest is to estimate accurately the probability of large losses of the form

$$\ell(\gamma) = \mathbb{P}(S(\mathbf{X}) > \gamma), \quad (22)$$

where  $\gamma = bn$  for some  $b > 0$ . To complete the model specifications, one needs to specify the joint distribution of the latent variables  $\mathbf{X} = (X_1, \dots, X_n)$ . One popular model that is widely used in the financial industry, and forms the basis of the CreditMetrics and other related models (Gupton et al., 1997; Li, 2000), is the *normal copula model*, under which the dependence among obligors is modeled by assuming the vector of latent variables follows a multivariate normal distribution. The underlying correlations are often specified through a linear factor model as follows

$$X_i = w_{i1}Z_1 + \dots + w_{im}Z_m + w_i\eta_i, \quad (23)$$

where  $Z_1, \dots, Z_m$  are iid standard normal variables known as *factors* that capture the systemic risk common to all the obligors and  $\eta_i$  is a normal random variable independent of the factors that captures the idiosyncratic risk of the  $i$ th obligor. In addition, it is often assumed (without loss of generality) that  $w_{i1}^2 + \dots + w_{im}^2 + w_i^2 = 1$ . Since the normal distribution has ‘thin tails’, Glasserman and Li (2005) propose an importance sampling estimator based on an exponential change of measure to estimate the probability of large portfolio loss in (22).

Nevertheless, the normal copula model, though mathematically simple, does not capture some important features of financial variables. In particular, it assumes the latent variables have a multivariate normal distribution, which does not agree with the empirical observation that many financial series exhibit extremal dependence and have tails heavier than those of the normal distribution (see, e.g. Geweke, 1993; Mashal and Zeevi, 2002, among many others). Therefore, the normal copula model may underestimate the probability of large portfolio losses. In view of this, Bassmaboo et al. (2008) introduce the *t-copula model* that generalizes the normal copula model by assuming the latent variables follow a multivariate Student- $t$  distribution.

Unlike Bassmaboo et al. (2008), who restrict their analysis to a single factor ( $m=1$ ), we consider the general  $m$ -factor model. Specifically, we assume that the latent factors  $Z_1, \dots, Z_m$  are iid standard Student- $t$  random variables with degree of freedom parameter  $\nu_Z$ , and that the  $\eta_i$ ,  $i = 1 \dots, n$  are iid Student- $t$  random variables independent of the factors  $Z_1, \dots, Z_m$  with mean 0, scale parameter  $\sigma_\eta^2$  and degree of freedom parameter  $\nu_\eta$ .

To derive a conditional Monte Carlo estimator for  $\ell$ , we first observe that the rare event  $\{S(\mathbf{X}) > \gamma\}$  occurs primarily when one of the factors, say  $Z_i$ , is large, while all the other variables are typical. Therefore, substantial variance reduction can be achieved simply by integrating out the maximal factor  $Z_i$ . To this end, we write the rare-event probability in (22) as

$$\ell = \sum_{i=1}^m \mathbb{P}(S(\mathbf{X}) > \gamma, Z_i = M_m), \quad (24)$$

where  $M_m = \max(Z_1, \dots, Z_m)$ . As in the previous examples, we generate  $N$  copies of  $\mathbf{Y} = (\mathbf{Z}, \eta)$ , where  $\mathbf{Z} = (Z_1, \dots, Z_m)$  and  $\eta = (\eta_1, \dots, \eta_n)$ , from the nominal distributions and use  $\mathbf{Y}_{-i} = (\mathbf{Z}_{-i}, \eta)$  to estimate the  $i$ th component of the above summand. It is therefore sufficient to derive an estimator for, say,  $\mathbb{P}(S(\mathbf{X}) > \gamma, Z_i = M_m)$ . Fix the index  $i$  and define

$$h_{ij} = (x_j - \sum_{l \neq i} w_{jl}Z_l - w_j\eta_j)w_{ji}^{-1}, \quad j = 1, \dots, n. \quad (25)$$

Now order  $h_{i1}, \dots, h_{in}$  in ascending order and call the ordered values  $g_{ij}, j = 1, \dots, n$ . Arrange the monetary losses  $c_1, \dots, c_n$  associated to the obligors' defaults in the same order as  $g_{ij}$  and denote them as  $d_1, \dots, d_n$ . Heuristically, the event  $\{S(\mathbf{X}) > \gamma, Z_i = M_m\}$  happens when the maximal factor  $Z_i$  is so large that a "sufficient" number of obligors defaults. More precisely, let  $r = \min\{l : \sum_{j=1}^l d_j > \gamma\}$ . Then the event  $\{S(\mathbf{X}) > \gamma, Z_i = M_m\}$  happens if and only if  $Z_i > g_{ir} \vee \widetilde{M}$ , where  $\widetilde{M}_i = \max(\mathbf{Z}_{-i})$ . In particular, if  $c_j = c$  for all  $j = 1, \dots, n$  (and therefore  $d_j = d$ ), then  $r = \lfloor \gamma/d \rfloor + 1$ , where  $\lfloor \cdot \rfloor$  indicates the integer part. Thus if we have  $N$  copies of  $\mathbf{Y}_{-i}^{(1)}, \dots, \mathbf{Y}_{-i}^{(N)}$ , we can estimate  $\mathbb{P}(S(\mathbf{X}) > \gamma, Z_i = M_m)$  by

$$\widehat{\ell}_i = \frac{1}{N} \sum_{k=1}^N \overline{F}_{\nu_Z} \left( g_{ir}^{(k)} \vee \widetilde{M}_i^{(k)} \right), \quad (26)$$

where  $\widetilde{M}_i^{(k)} = \max(\mathbf{Z}_{-i}^{(k)})$ , and  $F_{\nu_Z}(\cdot)$  is the distribution of the standard Student- $t$  distribution with degree of freedom parameter  $\nu_Z$ .

In the following numerical examples, we assume the obligors have the same individual threshold of defaulting ( $x_1 = \dots = x_n = 0.5 \times \sqrt{n}$ ), and the monetary losses associated to the obligors' defaults are the same ( $c_1 = \dots = c_n = 1$ ). In addition, the factors  $Z_1, \dots, Z_m$  are assumed to be iid standard Student- $t$  random variables with degree of freedom parameter  $\nu_Z = 6$ , while the individual risk factors  $\eta_1, \dots, \eta_n$  are iid Student- $t$  random variables with mean zero, scale parameter  $\sigma_\eta^2 = 9$  and degree of freedom parameter  $\nu_\eta = 6$ . Furthermore, we set the number of factors  $m$  to be 5 and for simplicity we assume all obligors have the same sensitivities to the factors:  $(w_{i1}, \dots, w_{im}) = (0.1, 0.2, 0.3, 0.4, 0.5)$ ,  $i = 1, \dots, n$ . Recall that the overall threshold is defined to be  $\gamma = bn$ . In the first numerical example, we fix  $n = 250$  and estimate the rare-event probability  $\ell$  for various default rates  $b$ . In the second numerical example, we fix  $b = 0.25$  and estimate  $\ell$  for various numbers of obligors  $n$ . For both examples we use a sample size of  $N = 50,000$ . Tables 12 and 13 report the estimates and their corresponding relative errors, as well as the variance reduction compared with the crude Monte Carlo estimator. As is clear from the tables, the proposed conditional Monte Carlo estimator gives fairly accurate estimates for all the scenarios studied, even though the sample size is rather small. In addition, even when the model involves more than a thousand random variables (the case when  $n = 1,000$ ), the proposed algorithm performs remarkably well.

Table 12: Performance of the proposed conditional Monte Carlo estimator for the  $t$ -copula model with  $n = 250$  obligors for various default rates  $b$ . Variance reduction is compared with the crude Monte Carlo estimator.

$b$	$\widehat{\ell}$	RE (%)	Var. reduction
0.1	$6.46 \times 10^{-5}$	0.89	3,910
0.2	$1.73 \times 10^{-5}$	0.53	41,000
0.3	$8.41 \times 10^{-6}$	0.41	$1.42 \times 10^5$
0.4	$5.00 \times 10^{-6}$	0.36	$3.13 \times 10^5$

Table 13: Performance of the proposed conditional Monte Carlo estimator for the  $t$ -copula model with various numbers of obligors  $n$ . Variance reduction is compared with the crude Monte Carlo estimator.

$n$	$\widehat{\ell}$	RE (%)	Var. reduction
100	$6.91 \times 10^{-4}$	1.0	265
250	$1.18 \times 10^{-5}$	0.46	80,700
500	$8.76 \times 10^{-7}$	0.23	$4.34 \times 10^6$
1000	$8.13 \times 10^{-8}$	0.13	$1.56 \times 10^8$

## 6 Concluding Remarks and Future Research

In this article we demonstrate how to utilize the asymptotic description of the way in which the rare event occurs to derive efficient conditional Monte Carlo estimators in a variety of settings. We show that the conditioning ideas discussed in Asmussen and Kroese (2006) can be applied to various far more complex models than the setting of sums of iid random variables. The effectiveness of the proposed algorithms are then illustrated by extensive simulation studies. By utilizing the same conditioning ideas, we are currently developing efficient algorithms for estimating the probabilities of large portfolio losses in the context of  $t$ -copula and other related models. For future research, it would be interesting to investigate possible ways to modify the proposed algorithm to estimate the *mean excess*, defined by  $\mathbb{E}(X - \gamma | X > \gamma)$ , or more generally,  $\mathbb{E}(S(\mathbf{X}) - \gamma | S(\mathbf{X}) > \gamma)$  for some specified function  $S(\cdot)$ . In addition, it would also be interesting to explore other applications, such as deriving efficient algorithms for computing portfolio Value-at-Risk involving heavy-tailed random variables.

## References

- S. Asmussen and K. Binswanger. Simulation of ruin probabilities for subexponential claims. *ASTIN Bulletin*, 27:297–318, 1997.
- S. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*. Springer, New York, 2007.
- S. Asmussen and D. P. Kroese. Improved algorithms for rare event simulation with heavy tails. *Advances in Applied Probability*, 38:545–558, 2006.
- S. Asmussen, K. Binswanger, and B. Højgaard. Rare events simulation for heavy-tailed distributions. *Bernoulli*, 6:303–322, 2000.
- S. Asmussen, R. Y. Rubinstein, and D. P. Kroese. Heavy tails, importance sampling and cross-entropy. *Stochastic Models*, 21:57–76, 2005.
- A. Bassmaboo, S. Juneja, and A. Zeevi. Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research*, 56(3):593–606, 2008.

- Z. I. Botev and D. P. Kroese. An efficient algorithm for rare-event probability estimation, combinatorial optimization, and counting. *Methodology and Computing in Applied Probability*, 10(4):471–505, 2008.
- M. Cruz. *Modeling, Measuring, and Hedging Operational Risk*. John Wiley & Sons, Chichester, 2002.
- J. Geweke. Bayesian treatment of the independent Student- $t$  linear model. *Journal of Applied Econometrics*, 8:S19–S40, 1993.
- P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.
- P. Glasserman and J. Li. Importance sampling for portfolio credit risk. *Management Science*, 51(11):1643–1656, 2005.
- P. Glasserman, P. Heidelberger, and P. Shahabuddin. Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, 12(3):239–269, 2002.
- L. A. Goodman. The variance of the product of  $k$  random variables. *Journal of the American Statistical Association*, 57(297):54–60, 1962.
- G. Gupton, C. Finger, and M. Bhatia. Creditmetrics technical document. Technical report, J. P. Morgan & Co., New York, 1997.
- S. Juneja. Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error. *Queueing Systems*, 57(2-3):115–127, 2007.
- S. Juneja and P. Shahabuddin. Simulating heavy tailed processes using delayed hazard rate twisting. *ACM Trans. Model. Comput. Simul.*, 12(2):94–118, 2002. ISSN 1049-3301.
- D. P. Kroese and R. Y. Rubinstein. The transform likelihood ratio method for rare event simulation with heavy tails. *Queueing Systems*, 46:317–351, 2004.
- D. Li. On default correlations: a copula function approach. *Journal of Fixed Income*, 9:43–54, 2000.
- R. Mashal and A. Zeevi. Beyond correlation: Extreme co-movements between financial assets. Technical report, Columbia University, New York, 2002.
- R. Y. Rubinstein. How to deal with the curse of dimensionality of likelihood ratios in Monte Carlo simulation. Technical report, Technion, Haifa, Israel, 2007.
- R. Y. Rubinstein and D. P. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization Monte-Carlo Simulation, and Machine Learning*. Springer-Verlag, New York, 2004.
- R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo Method Second Edition*. John Wiley & Sons, New York, 2007.