ORIGINAL PAPER

State-dependent importance sampling for estimating expectations of functionals of sums of independent random variables

Eya Ben Amar[1] · Nadhir Ben Rached[2] · Abdul-Lateef Haji-Ali[3] · Raúl Tempone[1,4]

Abstract
Estimating the expectations of functionals applied to sums of random variables (RVs) is a well-known problem encountered in many challenging applications. Generally, closed-form expressions of these quantities are out of reach. A naive Monte Carlo simulation is an alternative approach. However, this method requires numerous samples for rare event problems. Therefore, it is paramount to use variance reduction techniques to develop fast and efficient estimation methods. In this work, we use importance sampling (IS), known for its efficiency in requiring fewer computations to achieve the same accuracy requirements. We propose a state-dependent IS scheme based on a stochastic optimal control formulation, where the control is dependent on state and time. We aim to calculate rare event quantities that could be written as an expectation of a functional of the sums of independent RVs. The proposed algorithm is generic and can be applied without restrictions on the univariate distributions of RVs or the functional applied to the sum. We apply this approach to the log-normal distribution to compute the left tail and cumulative distribution of the ratio of independent RVs. For each case, we numerically demonstrate that the proposed state-dependent IS algorithm compares favorably to most well-known estimators dealing with similar problems.

1 Introduction

1.1 Motivation

In a probabilistic model, rare events are important events that infrequently happen with very small probabilities. Estimating these probabilities has become a substantial area of research because of its many applications, such as queuing systems, insurance risk, financial engineering, and wireless

✉ Eya Ben Amar
  eya.benamar@kaust.edu.sa

1 Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

2 School of Mathematics, University of Leeds, Leeds, UK

3 School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK

4 Alexander von Humboldt Professor in Mathematics for Uncertainty Quantification, RWTH Aachen University, Aachen, Germany

communication (Asmussen and Glynn 2007; Juneja and Shahabuddin 2006; Glasserman 2004; Ben Rached et al. 2015a). Typical examples occur in the context of communication systems, where the rare event could be an event in which the system fails to operate properly. For illustration, one can encounter the problem of estimating failure probabilities of the order of 10^{-9} for sophisticated networks, such as ultra-reliable fifth or sixth generation (5G or 6G) systems (Ben Rached et al. 2020a).

Calculating rare event quantities that could be written as an expectation of a functional of the sums of independent RVs is of paramount practical interest in many challenging applications. For instance, in financial engineering, calculating the value-at-risk (VaR) requires computing the left tails of the sums of RVs (i.e., the probability that the sum is less than a sufficiently small threshold). Another relevant example is calculating the probability that the signal-to-interference-plus-noise ratio is less than a given threshold in communication systems. Under some particular fading environments, this probability can be expressed as a cumulative distribution function (CDF) of the ratio of independent RVs.

1.2 Literature review

Various researchers have proposed closed-form approximations of the left and right tails of sums of RVs (López-Salcedo 2009; Xiao et al. 2019; Chatterjee et al. 2018; Beaulieu and Luan 2020; Zhu et al. 2020; Constantinescu et al. 2016; Singh 2018). However, these approximation methods are not generic. Moreover, their accuracy is not always guaranteed for all scenarios, as it can degrade for certain system parameters. The Monte Carlo (MC) method can be used as a generic tool to cope with these problems. However, it is well acknowledged that estimating rare event quantities with the naive MC sampler requires a prohibitively large number of simulation runs (Kroese et al. 2011). Variance reduction techniques have been used extensively to improve the computational work of the naive MC method. In this context, importance sampling (IS) is among the most popular variance reduction techniques that provide accurate estimates of rare event probabilities with a reduced number of simulation runs when appropriately used (Kroese et al. 2011).

Variance reduction techniques have been widely discussed in the literature, and particular focus has been devoted to proposing algorithms for the efficient simulation of the right tail of sums of RVs (i.e., the probability that the sum exceeds a sufficiently large threshold). In particular, for distributions with light right tails (i.e., decaying at an exponential rate or faster), under some regularity assumptions, the popular exponential twisting IS approach (Asmussen and Glynn 2007) satisfies the logarithmic efficiency property, which is a useful metric used to assess the efficiency of an estimator. In contrast, for heavy-tailed distributions, such as the case of log normals and Weibulls with shape parameters strictly less than 1, the exponential twisting method is inapplicable. Therefore, efficient algorithms have been developed for estimating tail probabilities involving heavy-tailed RVs. In this context, Asmussen and Binswanger (1997) provided the first logarithmically efficient estimator for such probabilities using the conditional MC idea. Other authors (Asmussen and Kroese 2006) have proposed an estimator with a bounded relative error under distributions with regularly varying tails, which was further extended to more general scenarios (e.g., see Hartinger and Kortschak 2009; Chan and Kroese 2011; Asmussen and Kortschak 2012, 2015; Ghamami and Ross 2012). In addition to estimators based on the conditional MC, various state-independent IS techniques have been proposed (Juneja and Shahabuddin 2002; Juneja 2007; Karthyek Rajhaa and Juneja 2012; Murthy et al. 2015).

State-independent changes of measure for estimating certain rare events involving sums of heavy-tailed RVs cannot achieve logarithmic efficiency (Bassamboo et al. 2007). Therefore, more complex state-dependent IS algorithms have been proposed in the literature over the last few years to esti-

mate probabilities for sums of heavy-tailed independent RVs. Of valuable interest are studies developed in Dupuis et al. (2007), Dupuis and Wang (2007), Blanchet and Liu (2006, 2008), Blanchet and Li (2011) and Blanchet and Lam (2012). The researchers in Blanchet and Liu (2006) developed an efficient state-dependent IS estimator with a bounded relative error under distributions with regularly varying heavy tails. The estimator can also be adapted to provide strongly efficient algorithms in light-tailed situations. A related approach, based on the construction of Lyapunov inequalities, has also been developed (Blanchet and Liu 2008) for constructing strongly efficient estimators for large deviation probabilities of regularly varying random walks. These algorithms use a parametric family of change of measure based on mixtures that are appropriately selected using Lyapunov bounds. Moreover, stochastic control and game theory have been used to build efficient state-dependent IS schemes to simulate rare events (Dupuis and Wang 2004; Dupuis et al. 2005, 2007). For instance, in the heavy-tailed setting, the authors in Dupuis et al. (2007) constructed dynamic IS estimators with a nearly asymptotically optimal relative error for independent and identically distributed (i.i.d.) nonnegative regularly varying RVs. They considered a parametric family of change of measure whose parameters are determined by solving a deterministic, discrete-time control problem. The closest work to the proposed approach is in Dupuis and Wang (2004), where the authors proposed an approach based on connecting IS with stochastic optimal control (SOC). The scope of Dupuis and Wang (2004) is limited to the i.i.d. case and distributions with finite moment generating functions. In this work, independence is the only assumption we make. The connection between IS and SOC has been investigated for other scenarios, such as diffusions (Hartmann et al. 2017), and for stochastic reaction networks approximated by the Tau-Leap scheme (Hammouda et al. 2021). The dynamics in Hammouda et al. (2021) evolve according to discrete-time discrete-space Markov chains, whereas the dynamics in this work evolve according to discrete-time continuous-space Markov chains.

Few researchers have recently addressed the left-tail region [i.e., the probability that sums of nonnegative RVs fall below a sufficiently small threshold (Asmussen et al. 2016; Ben Issaid et al. 2017; Ben Rached et al. 2015b, 2020a, b, 2021)]. For instance, Asmussen et al. (2016) considered the specific setting of the i.i.d. sum of log-normal RVs. The approach was based on the exponential twisting technique and is logarithmically efficient. The work of Ben Rached et al. (2015b) proposed two unified hazard rate twisting (HRT)-based approaches that estimate the outage capacity values for generalized independent fading channels. The first estimator achieves logarithmic efficiency for arbitrary fading models, whereas the second achieves the bounded relative error criterion for most well-known fading variates and log-

arithmic efficiency for the log-normal case. Recently, Ben Rached et al. (2020a) proposed an IS scheme based on sample rejection applied to the case of the independent Rayleigh, correlated Rayleigh, and i.i.d. Rice fading models, showing that the estimator satisfies the bounded relative error property.

1.3 Main contributions

In this paper, we propose a generic state-dependent IS approach to estimate rare event probabilities that could be written as an expectation of a functional of the sums of independent RVs. We adopt a SOC formulation to determine the optimal IS parameters, minimizing the variance or, equivalently, the second moment of the estimator within a preselected class of measures. After formulating the SOC problem and describing the algorithm used to derive the optimal controls, which are optimal IS parameters, we apply the proposed algorithm to two examples: the computation of the left-tail probability in a log-normal setting, and the computation of the CDF of the ratio of independent log-normal RVs. The proposed algorithm is generic and not restricted to the log-normal environment. The algorithm can be applied to compute the quantity of interest without restrictions on the distribution of the univariate RVs in the sum or the expression of the functional applied to the sum. Numerical simulations demonstrate the superior performance of the proposed estimator in terms of the number of samples and computational work to meet a given prescribed tolerance (TOL) compared to the existing state-of-the-art estimators dealing with similar problems.

The rest of the paper is organized as follows. Section 2 describes the problem setting, presents applications that fall within the scope of applicability of the proposed approach, and introduces the concept of IS. Section 3 contains the main work, explaining the state-dependent IS scheme via a novel SOC formulation, followed by presenting the algorithm. Section 4 applies the proposed algorithm to two applications in wireless communications. The proposed algorithm compares favorably to some well-known estimators dealing with similar problems.

2 Problem setting

This section states the objective of the method and enumerates some applications that fall within the scope of its applicability. Next, this section introduces the concept of the naive MC method. Finally, it presents the IS technique, one of the most popular variance reduction techniques.

2.1 Objective

We consider $\mathbf{X} = (X_1, X_2, \ldots, X_N)^t$ to be a random vector comprising independent positive components with probability density functions (PDFs) $f_{X_1}(.), f_{X_2}(.), \ldots, f_{X_N}(.)$ and a joint PDF $f(\mathbf{x}) = \prod_{n=1}^{N} f_{X_n}(x_n)$. In this work, $X_i, i = 1, \ldots, N$ are one-dimensional vectors. However, this approach is still applicable to the multidimensional case. We let $S_N = \sum_{n=1}^{N} X_i$ and $g : \mathbb{R}_+ \to \mathbb{R}$ be a given function. We aim to develop a state-dependent IS algorithm via a connection to an SOC formulation to estimate rare event quantities that could be written as follows:

$$\alpha = \mathbb{E}[g(S_N)]. \tag{1}$$

2.2 Applications

2.2.1 Right and left tail

One of the problems that can be written as (1) is calculating the right-tail probability of sums of RVs (i.e., the probability that the sum is larger than a sufficiently large threshold), which arises in many areas of engineering. This probability can be expressed as

$$\alpha = \mathbb{P}(S_N \geq \gamma_{th}) = \mathbb{E}\left[\mathbb{1}_{(S_N \geq \gamma_{th})}\right], \tag{2}$$

corresponding to (1), where $g(x) = \mathbb{1}_{(x \geq \gamma_{th})}$.

As a practical example, the right-tail probability of the sums of RVs may represent the ruin probability of an insurance company. In this setting, S_N represents the total sum of claims, and γ_{th} is the initial reserve. The claims $X_1, \ldots, X_N$ can be modeled by heavy-tailed distributions (Asmussen et al. 2000). In the Cramer–Lundberg model, this probability can be expressed as (2) (Asmussen and Glynn 2007). Moreover, calculating left-tail probabilities occurs extensively in many applications. In these cases, the quantity of interest can be expressed as

$$\alpha = \mathbb{E}\left[\mathbb{1}_{(S_N \leq \gamma_{th})}\right], \tag{3}$$

which is in the form of (1), where $g(x) = \mathbb{1}_{(x \leq \gamma_{th})}$.

One of the most relevant examples is estimating the VaR, defined as the $1 - \alpha$ quantile of the loss distribution, for a sufficiently small value of α. We let a portfolio be based on N assets with upcoming prices $X_1 \ldots, X_N$, which can be modeled using log-normal distributions (Asmussen et al. 2016). The VaR of S_N at the level of α (VaR$_\alpha$ (S_N)) is defined as the value such that the probability of a loss larger than that value is equal to $1 - \alpha$ (Alemany et al. 2013; Sun and Hong 2009). In other words, VaR$_\alpha$ (S_N) $= F_{S_N}^{-1}(\alpha)$, where $F_{S_N}(.)$ is the CDF of S_N.

Another challenging application is the analysis of wireless communication systems. The outage probability (OP), defined as the probability that the signal-to-noise ratio (SNR) falls below a given threshold γ_{th} (Yilmaz and Alouini 2012), is equivalent to computing the CDF of sums of the SNRs of the received signals. Hence, it can be expressed as in (3).

2.2.2 CDF of the ratio of independent RVs

Another performance metric that can be expressed as in (1) is the OP in the presence of co-channel interference and noise. For single-input, single-output (SISO) systems, the OP is expressed as (Ben Rached et al. 2017)

$$P_{out} = \mathbb{P}(\text{SINR} \leq \gamma_{\text{th}}) = \mathbb{P}\left(\frac{X_0}{\sum_{n=1}^{N} X_n + \eta} \leq \gamma_{\text{th}}\right), \quad (4)$$

where X_0 denotes the desired user signal power, $X_1, \ldots, X_N$ represent the received powers of the N interfering signals, and η indicates the variance of the additive white Gaussian noise. We assume that $X_0, \ldots, X_N$ are independent. Through conditioning on $X_1, X_2, \ldots, X_N$ and using the law of total expectation, we write (4) as

$$\mathbb{E}\left[F_{X_0}\left(\gamma_{\text{th}} \sum_{n=1}^{N} X_n + \gamma_{\text{th}}\eta\right)\right], \quad (5)$$

where $F_{X_0}(\cdot)$ is the CDF of the RV X_0, corresponding to the form in (1) with $g(x) = F_{X_0}(\gamma_{\text{th}}(x + \eta))$.

2.3 Importance sampling

The naive MC estimator of the quantity of interest in (1) is

$$\hat{\alpha}_{mc} = \frac{1}{M} \sum_{k=1}^{M} g\left(S_N^{(k)}\right), \quad (6)$$

where M denotes the number of simulation runs, and $\{S_N^{(k)}\}_{k=1}^{M}$ represents independent realizations of the RV $S_N = \sum_{i=1}^{N} X_i$.

However, the naive MC method is computationally expensive, requiring a substantial number of simulation runs to meet a given accuracy when considering rare event probabilities. Using appropriate variance reduction techniques, such as IS, is necessary to overcome the failure of naive MC simulations and considerably reduce the computational work. The idea is to perform a change of measure under which the rare event is generated with a higher probability than under the original distribution (Kroese et al. 2011). The IS technique consists of writing α as

$$\alpha = \mathbb{E}_{\tilde{f}}\left[\tilde{g}(\mathbf{X})\right], \quad (7)$$

where

$$\tilde{g}(\mathbf{x}) = g\left(\sum_{n=1}^{N} x_n\right) \frac{f(\mathbf{x})}{\tilde{f}(\mathbf{x})}, \quad (8)$$

and $\mathbb{E}_{\tilde{f}}[\cdot]$ denotes the expectation under which the vector $\mathbf{X}$ has the joint PDF $\tilde{f}(\cdot)$. The IS estimator is expressed as

$$\hat{\alpha}_{IS} = \frac{1}{M} \sum_{k=1}^{M} \tilde{g}\left(\mathbf{X}^{(k)}\right), \quad (9)$$

where $\{\mathbf{X}^{(k)}\}_{k=1}^{M}$ represents independent realizations of $\mathbf{X}$ sampled according to $\tilde{f}(\cdot)$. When $g(x) > 0, x \in \mathbb{R}_+$, the optimal change of measure minimizing the variance of the IS estimator is given by

$$f^*(\mathbf{x}) = \frac{f(\mathbf{x})g\left(\sum_{n=1}^{N} x_i\right)}{\alpha}, \quad \mathbf{x} \in \mathbb{R}_+^N. \quad (10)$$

This optimal change of measure yields zero variance; thus, it is called the zero variance change of measure. However, using such a change of measure is impractical because it assumes the knowledge of α, which is the unknown quantity.

3 IS via an SOC formulation

This section explains the SOC formulation and how to link it to IS to construct the state-dependent IS estimator. Then, it introduces the HRT family as a change of measure. Finally, this section describes the steps of the proposed state-dependent IS algorithm.

3.1 State-dependent IS approach

The idea we adopt is to link the problem of finding an efficient change of measure to a SOC problem. To apply SOC to the current static problem, we embed it with the evolution of a Markov chain with the following dynamics:

$$S_{n+1} = S_n + X_{n+1}, \quad n = 0, 1, \ldots, N-1, \quad (11)$$

where $S_0 = 0$. Instead of sampling X_{n+1} according to $f_{X_{n+1}}(\cdot)$, we perform a change of measure such that, given S_n, X_{n+1} is distributed according to $\tilde{f}_{X_{n+1}}(\cdot; \mu_{n+1}(S_n))$, where μ_{n+1} is a function of S_n. With this idea, the new joint PDF can be written as

$$\tilde{f}(\mathbf{x}) = \prod_{n=1}^{N} \tilde{f}_{X_n}(x_n; \mu_n(s_{n-1})), \quad (12)$$

where $s_{n-1} = \sum_{i=1}^{n-1} x_i$. The objective is to determine the optimal controls $\mu_n : \mathbb{R}_+ \to A \subset \mathbb{R}$, $n = 1, 2, \ldots, N$ that minimize the second moment of the IS estimator. Therefore, assuming that the second moment of the estimator is finite, we define the cost function for $\mu_{n+1}, \ldots, \mu_N \in \mathcal{D}^{N-n}$, $n = 0, \ldots, N-1$ as

$$C_{n,s}(\mu_{n+1}, \ldots, \mu_N)$$
$$= \mathbb{E}_{\tilde{f}}\left[(g(S_N))^2 \prod_{i=n+1}^{N} \left(\frac{f_{X_i}(X_i)}{\tilde{f}_{X_i}(X_i; \mu_i(S_{i-1}))} \right)^2 \mid S_n = s \right], \quad (13)$$

where $\mathcal{D} = \{\mu : \mathbb{R}^+ \to A\}$ represents the set of admissible Markov controls. More precisely,

for $\mu \in \mathcal{D}$, $\mu : \mathbb{R}_+ \to A$.

We also define the value function as follows:

$$u(n, s) = \inf_{(\mu_{n+1}, \ldots, \mu_N) \in \mathcal{D}^{N-n}} C_{n,s}(\mu_{n+1}, \ldots, \mu_N). \quad (14)$$

The above SOC formulation is flexible because the RVs are dependent. The same observation holds for the optimal change of measure (10). Therefore, if the family of PDFs $\tilde{f}_{X_n}(.; \mu_n)$ is sufficiently large, we can expect the SOC formulation to deliver an estimator with a performance close to that of the optimal estimator.

Next, the question is how to solve the minimization problem and determine the optimal controls μ_n, $n = 1, 2, \ldots, N$. The idea is to solve it sequentially by going backward in time. In Proposition 1, we state the dynamic programming equation solved by the value function u.

Proposition 1 *For all $n \in \{0, 1, \ldots, N-1\}$ and $s \geq 0$, we obtain*

$$u(n, s)$$
$$= \inf_{\mu \in A} \mathbb{E}_{\tilde{f}}\left[\left(\frac{f_{X_{n+1}}(X_{n+1})}{\tilde{f}_{X_{n+1}}(X_{n+1}; \mu)} \right)^2 u(n+1, S_{n+1}) \mid S_n = s \right]. \quad (15)$$

If the minimum is attained, we have

$$\mu_{n+1}(s)$$
$$= \arg\min_{\mu \in A} \mathbb{E}_{\tilde{f}}\left[\left(\frac{f_{X_{n+1}}(X_{n+1})}{\tilde{f}_{X_{n+1}}(X_{n+1}; \mu)} \right)^2 u(n+1, S_{n+1}) \mid S_n = s \right], \quad (16)$$

where $u(N, x) = (g(x))^2$, $S_{n+1} = s + X_{n+1}$, and X_{n+1} is distributed according to $\tilde{f}_{X_{n+1}}(\cdot; \mu_{n+1}(s))$.

Proof For simplicity, we assume that the optimal control is attained:

$$u(n, s) = \min_{(\mu_{n+1}, \ldots, \mu_N) \in \mathcal{D}^{N-n}} C_{n,s}(\mu_{n+1}, \ldots, \mu_N). \quad (17)$$

Step 1 We let $\mu_{n+1}^*, \ldots, \mu_N^*$ be the optimal control minimizing (17). Then, we obtain

$$u(n, s) = \mathbb{E}_{\tilde{f}}\left[(g(S_N))^2 \prod_{i=n+1}^{N} \left(\frac{f_{X_i}(X_i)}{\tilde{f}_{X_i}(X_i; \mu_i^*(S_{i-1}))} \right)^2 \mid S_n = s \right]$$
$$= \mathbb{E}_{\tilde{f}}\left[\mathbb{E}_{\tilde{f}}\left[(g(S_N))^2 \prod_{i=n+2}^{N} \left(\frac{f_{X_i}(X_i)}{\tilde{f}_{X_i}(X_i; \mu_i^*(S_{i-1}))} \right)^2 \right.\right. \quad (18)$$
$$\times \left.\left. \left(\frac{f_{X_{n+1}}(X_{n+1})}{\tilde{f}_{X_{n+1}}(X_{n+1}; \mu_{n+1}^*(S_n))} \right)^2 \mid S_n = s, X_{n+1} \right] \mid S_n = s \right].$$

Knowing X_{n+1} and S_n, $\left(\frac{f_{X_{n+1}}(X_{n+1})}{\tilde{f}_{X_{n+1}}(X_{n+1}; \mu_{n+1}^*(S_n))} \right)^2$ is deterministic. Thus, using the Markov property of S_n, we obtain

$$\mathbb{E}_{\tilde{f}}\left[(g(S_N))^2 \prod_{i=n+2}^{N} \left(\frac{f_{X_i}(X_i)}{\tilde{f}_{X_i}(X_i; \mu_i^*(S_{i-1}))} \right)^2 \mid S_n = s, X_{n+1} \right]$$
$$= C_{n+1, S_{n+1}}(\mu_{n+2}^*, \ldots, \mu_N^*) \quad (19)$$
$$\geq u(n+1, S_{n+1}).$$

Hence, the following inequality holds:

$$u(n, s) \geq \mathbb{E}_{\tilde{f}}\left[\left(\frac{f_{X_{n+1}}(X_{n+1})}{\tilde{f}_{X_{n+1}}(X_{n+1}; \mu_{n+1}^*(s))} \right)^2 u(n+1, S_{n+1}) \mid S_n = s \right]$$
$$\geq \min_{\mu \in A} \mathbb{E}_{\tilde{f}}\left[\left(\frac{f_{X_{n+1}}(X_{n+1})}{\tilde{f}_{X_{n+1}}(X_{n+1}; \mu)} \right)^2 u(n+1, S_{n+1}) \mid S_n = s \right]. \quad (20)$$

Step 2 We choose the control μ_{n+1}^+ to be arbitrary and, given the value of S_{n+1}, we select the optimal controls $\mu_{n+2}^*, \ldots, \mu_N^*$. Then, the following lower bound holds:

$$u(n, s) \leq \mathbb{E}_{\tilde{f}}\left[\left(\frac{f_{X_{n+1}}(X_{n+1})}{\tilde{f}_{X_{n+1}}(X_{n+1}; \mu_{n+1}^+(s))} \right)^2 (g(S_N))^2 \right.$$
$$\prod_{i=n+2}^{N} \left(\frac{f_{X_i}(X_i)}{\tilde{f}_{X_i}(X_i; \mu_i^*(S_{i-1}))} \right)^2 \mid S_n = s \right]$$
$$\leq \mathbb{E}_{\tilde{f}}\left[\left(\frac{f_{X_{n+1}}(X_{n+1})}{\tilde{f}_{X_{n+1}}(X_{n+1}; \mu_{n+1}^+(s))} \right)^2 \right.$$
$$\times \mathbb{E}_{\tilde{f}}\left[(g(S_N))^2 \prod_{i=n+2}^{N} \left(\frac{f_{X_i}(X_i)}{\tilde{f}_{X_i}(X_i; \mu_i^*(S_{i-1}))} \right)^2 \right.$$
$$\left.\left. \mid S_n = s, X_{n+1} \right] \mid S_n = s \right]$$

$$= \mathbb{E}_{\tilde{f}}\left[\left(\frac{f_{X_{n+1}}(X_{n+1})}{\tilde{f}_{X_{n+1}}(X_{n+1}; \mu_{n+1}^+(s))}\right)^2 u(n+1, S_{n+1}) \mid S_n=s\right]. \quad (21)$$

Taking the minimum over all controls $\mu_{n+1}^+(s)$ yields

$$u(n, s) \le \min_{\mu \in A} \mathbb{E}_{\tilde{f}}\left[\left(\frac{f_{X_{n+1}}(X_{n+1})}{\tilde{f}_{X_{n+1}}(X_{n+1}; \mu)}\right)^2 \times u(n+1, S_{n+1}) \mid S_n=s\right]. \quad (22)$$

Hence, the proof is concluded using (21) and (22).   □

Remark 1 We can prove the proposition without the assumption that the minimum is attained. For that, we use a minimizing sequence μ_i of controls, satisfying

$$u(n, s) = \lim_{i \to \infty} C_{n,s}(\mu_i). \quad (23)$$

3.2 Hazard rate twisting family

The choice the family of PDFs $\tilde{f}_{X_n}(.; \mu_n)$, $n = 1, \ldots, N$ in this work is based on the well-known HRT. The HRT technique was originally developed to deal with the right tail of sums of heavy-tailed RVs (Juneja and Shahabuddin 2002; Ben Rached et al. 2018).
We define the hazard rate $\lambda_{X_i}(\cdot)$ associated with the RV X_i as

$$\lambda_{X_i}(x) = \frac{f_{X_i}(x)}{1 - F_{X_i}(x)}, \quad x > 0, \quad (24)$$

where $F_{X_i}(x) = \mathbb{P}(X_i \le x)$ is the CDF of X_i, $i = 1, \ldots, N$. We also define the hazard function as

$$\Lambda_{X_i}(x) = -\log\left(1 - F_{X_i}(x)\right), \quad x > 0. \quad (25)$$

From (24) and (25), the PDF of X_i can be expressed as

$$f_{X_i}(x) = \lambda_{X_i}(x) \exp\left(-\Lambda_{X_i}(x)\right), \quad x > 0. \quad (26)$$

The HRT change of measure is obtained by twisting the hazard rate of each component X_i, $i = 1, \ldots, N$ by a quantity $\mu_i < 1$ as follows:

$$\tilde{f}_{X_i}(x; \mu_i) = (1-\mu_i)\lambda_{X_i}(x) \exp\left(-(1-\mu_i)\Lambda_{X_i}(x)\right) = (1-\mu_i)f_{X_i}(x) \exp\left(\mu_i \Lambda_{X_i}(x)\right), \quad x > 0. \quad (27)$$

Moreover, μ_i should satisfy $0 \le \mu_i < 1$, $i = 1, \ldots, N$ to efficiently address the estimation of the right tail of the sum distribution. Consequently, the tail of the resulting distribution becomes much heavier to the right than the original. However, this feature is unsuitable for dealing with the left tail. Two approaches were proposed in Ben Rached et al.

(2015b) to adjust the HRT to handle the left-tail region. The first is based on twisting the RVs $-X_1, \ldots, -X_N$ instead of the original variates $X_1, \ldots, X_N$. The second approach applies the HRT approach to $X_1, \ldots, X_N$ using a negative twisting parameter.
Considering the appropriate twisting parameter, we employ the HRT change of measure given by (27), and the set A in this case is given by $A = (-\infty, 1)$. By doing so, the value function is given by

$$u(n, s) = \inf_{\mu \in A} \mathbb{E}_{\tilde{f}}\left[\frac{\exp\left(-2\mu\Lambda_{X_{n+1}}(X_{n+1})\right)}{(1-\mu)^2} u(n+1, S_{n+1}) \mid S_n=s\right]. \quad (28)$$

3.3 Algorithm

Based on the results stated in the proposition, we propose a numerical algorithm to approximate the optimal controls μ_n, where $n = 1, \ldots, N$. We start by truncating the space $\mathbb{R}_+$ and work in the interval $[0, \bar{S}]$, where $\bar{S}$ is a large number in $\mathbb{R}_+$. There are particular cases that we treat, where $\bar{S}$ is naturally chosen. For instance, when estimating $\mathbb{P}(S_N \le \gamma_{\text{th}})$, due to the nonnegativity of X_i, $u(n, s) = 0$ for $s \ge \gamma_{\text{th}}$ and $n = 0, \ldots, N$. In this case, $\bar{S}$ is set equal to γ_{th}. In the general case, $\bar{S}$ is selected to be sufficiently large. At each step of the backward algorithm, we use linear extrapolation to compute the value function for $s > \bar{S}$.
We consider a mesh in the one-dimensional s-space: $0 = s_0 < s_1, \cdots < s_K = \bar{S}$. The aim is to approximately compute $u(n, s_k)$ for all $n = 0, 1, \ldots, N-1$ and s_k, $k = 0, 1, \ldots, K$. The algorithm is summarized as follows:
Step 1 For each s_k in the mesh, we solve the following:

$$u(N-1, s_k) = \min_{\mu \in A} \mathbb{E}_{\tilde{f}}\left[\left(\frac{f_{X_N}(X_N)}{\tilde{f}_{X_N}(X_N; \mu)}\right)^2 (g(s_k + X_N))^2\right] = \min_{\mu \in A} \int_0^{+\infty} \frac{(f_{X_N}(t))^2}{\tilde{f}_{X_N}(t; \mu)} (g(s_k + t))^2 dt, \quad (29)$$

and

$$\mu_N(s_k) = \arg\min_{\mu \in A} \int_0^{+\infty} \frac{(f_{X_N}(t))^2}{\tilde{f}_{X_N}(t; \mu)} (g(s_k + t))^2 dt. \quad (30)$$

This step is not expensive because we must compute a one-dimensional integral for each point in the mesh and perform an optimization problem for the parameter μ. When the HRT family is used, the optimization problem becomes equivalent to determining the root of a nonlinear equation.

Step 2 After obtaining $u(N-1, s_k)$ for all s_k in the grid, the next step again applies the result of the proposition to obtain an approximation of $u(N-2, s_k)$ and $\mu_{N-1}(s_k)$

$$u(N-2, s_k) = \min_{\mu \in A} \int_0^{+\infty} \frac{\left(f_{X_{N-1}}(t)\right)^2}{\tilde{f}_{X_{N-1}}(t; \mu)} u(N-1, s_k + t)\, dt. \quad (31)$$

To perform this step, we must know $u(N-1, s)$ for all s that are not necessarily in the grid. To overcome this problem, we proceed by interpolating between the points $u(N-1, s_k)$, where $k = 0, 1, \ldots, K$. As mentioned, linear extrapolation is employed for $s > \bar{S}$ when needed.

Step 3 After computing $\mu_n(s_k)$ for $n = 1, 2, \ldots, N$ and all s_k in the grid $k = 0, 1, 2, \ldots, K$, the following step is to solve for $\mu_n, n = 1, 2, \ldots, N$ by going forward in time. More specifically, we start at $S_0 = 0$ and sample from $\tilde{f}_{X_1}(\cdot, \mu_1)$ to obtain S_1. Further, $\mu_1(0)$ was already computed in the resolution of the backward problem. We compute μ_2 as

$$\mu_2(\tilde{s}_1) = \arg\min_{\mu \in A} \int_0^\infty \frac{\left(f_{X_2}(t)\right)^2}{\tilde{f}_{X_2}(t; \mu)} u(2, \tilde{s}_1 + t)\, dt. \quad (32)$$

After computing μ_2, we simulate S_2 as $S_2 = \tilde{s}_1 + X_2$, with X_2 sampled from $\tilde{f}_{X_2}(.; \mu_2)$. We continue repeating this procedure until we reach μ_N and then sample X_N. In the case of smooth controls, the optimization problem (32) can be avoided using interpolation between controls, obtained in the backward step, on the grid, $s_1, \ldots, s_K$.

Step 4 The forward problem is repeated M times. The proposed IS estimator is given as

$$\hat{\alpha}_{IS} = \frac{1}{M} \sum_{k=1}^{M} g\left(S_N^{(k)}\right) \prod_{i=1}^{N} \frac{f_{X_i}\left(X_i^{(k)}\right)}{\tilde{f}_{X_i}\left(X_i^{(k)}, \mu_i\left(S_{i-1}^{(k)}\right)\right)}. \quad (33)$$

4 Numerical results

This section presents selected numerical results to illustrate the performance of the proposed IS scheme. First, the methodology adopted to demonstrate the performance of the proposed approach is discussed. The motivation for using the improved version of the proposed method, called the aggregate method, is explained. Then, the proposed algorithm is applied to estimate the OP at the output of diversity receivers with and without co-channel interference in the log-normal environment.

4.1 Methodology

Within the broad applicability of the proposed estimator, we focused on applying it to calculate the left-tail probability

and the CDF of the ratio of independent RVs. We used the proposed estimator to estimate the OP at the output of diversity receivers with and without co-channel interference. We considered the case in which the antennae are sufficiently spaced to assume that independent RVs can model fading channels. We considered the log-normal fading environment that exhibits a good fit for realistic propagation channels. We demonstrated that the proposed approach achieves a substantial reduction of the variance compared to other well-known IS algorithms.

In both applications, the objective was to efficiently estimate the following:

$$\alpha = \mathbb{E}\left[g\left(\sum_{i=1}^{N} X_i\right)\right], \quad (34)$$

where $X_1, \ldots, X_N$ denote i.i.d. log-normal RVs with parameters m and σ^2. The PDF of X_i, $i = 1, \ldots, N$, is expressed as follows:

$$f_{X_i}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - m)^2}{2\sigma^2}\right), \quad x > 0. \quad (35)$$

For the second application, we let X_0 be a log-normal RV with parameters m_0 and σ_0^2.

We employed the HRT change of measure in (27) to build the estimator. Hence, we call this approach the HRT-SOC IS approach, and the corresponding estimator is denoted by $T_{\text{HRT-SOC}}$ which is expressed as follows:

$$T_{\text{HRT-SOC}} = g(S_N) \prod_{i=1}^{N} \frac{e^{-\mu_i(S_{i-1})\Lambda_{X_i}(X_i)}}{(1 - \mu_i(S_{i-1}))}, \quad (36)$$

where $g(x) = \mathbb{1}_{(x \leq \gamma_{th})}$ in the first application, and $g(x) = F_{X_0}(\gamma_{th}(x + \eta))$ in the second application. In this setting, each step of the backward algorithm can be expressed, for $k = 0, \ldots, K$, as

$$u(n, s_k) = \min_{\mu \in (-\infty, 1)} \frac{1}{1-\mu} \int_0^{+\infty} u(n+1, s_k + t) f_{X_{n+1}}(t)$$
$$\times\, e^{-\mu\, \Lambda_{X_{n+1}}(t)} dt. \quad (37)$$

The controls $\mu_n(s_k)$ are obtained by solving the following equation:

$$1 - \mu_n(s_k)$$
$$= \frac{\int_0^{+\infty} u(n-1, s_k + t) f_{X_{n-1}}(t)\, e^{-\mu_n(s_k)\, \Lambda_{X_{n-1}}(t)} dt}{\int_0^{+\infty} \Lambda_{X_{n-1}}(t) u(n-1, s_k + t) f_{X_{n-1}}(t)\, e^{-\mu_n(s_k)\, \Lambda_{X_{n-1}}(t)} dt}. \quad (38)$$

For the forward step, assuming that the control is smooth (motivated by numerical observations), we can compute the

controls by interpolating between the points $\mu_n(s_k)$, $k = 0, \ldots, K$.

To sample from the change of measures $\tilde{f}_{X_i}(\cdot)$, $i = 1, \ldots, N$, we used the inverse CDF technique. In Ben Rached et al. (2015a), the authors revealed that the inverse CDF of the HRT of a log-normal RV X_i is given by

$$F_{X_i}^{-1}(y) = \exp\left(m + \sigma\,\Phi^{-1}\left(1 - (1 - y)^{-\frac{1}{\mu_i - 1}}\right)\right), \quad (39)$$

where μ_i is the twisting parameter corresponding to X_i and $\Phi(\cdot)$ is the CDF of the standard normal distribution. This formula can be generalized to other distributions as BenRached et al. (2015b, eq. (65)).

The relative error serves as a measure of the efficiency of the estimators. The relative error of the naive MC estimator and the proposed IS estimator are defined respectively through the central limit theorem (Asmussen and Glynn 2007) as

$$\epsilon_{MC} = C\frac{\sqrt{\alpha(1-\alpha)}}{\sqrt{M}\alpha}, \quad \epsilon_{HRT\text{-}SOC} = C\frac{\sqrt{\mathrm{Var}\left[T_{HRT\text{-}SOC}\right]}}{\sqrt{M}\alpha}, \quad (40)$$

where C is the confidence constant equal to 1.96 for the 95% confidence interval.

We compared the estimator defined in (36) to other existing estimators when calculating the OP at the output of diversity receivers with and without co-channel interference. For instance, using the log-normal setting with the HRT technique allows us to compare the estimator with the approach in Ben Rached et al. (2015a), which used the HRT without SOC (i.e., the control is constant, independent of the state and time). We denote this method as HRT. In the numerical experiments, the HRT-SOC technique reduces the variance substantially compared to other approaches. However, it requires additional time, called a backward cost, to determine the optimal controls.

We let M_{HRT} and $M_{HRT\text{-}SOC}$ be the number of required simulation runs for the HRT estimator T_{HRT} and the proposed estimator $T_{HRT\text{-}SOC}$, respectively, to ensure a relative error equal to TOL. The total costs of the HRT-SOC and HRT approaches are expressed as follows:

$$W_{HRT\text{-}SOC} = \underbrace{N \times K \times T_b}_{\text{Backward cost}} + \underbrace{M_{HRT\text{-}SOC} \times T_f}_{\text{Forward cost}}, \quad (41)$$

$$W_{HRT} = \underbrace{M_{HRT} \times T_f}_{\text{Forward cost}}, \quad (42)$$

where T_b is the time required in the backward algorithm to calculate a single control, and T_f represents the cost per sample in the forward step (approximately the same for both approaches). Figures 2 and 4 illustrate that the variance reduction compared to the HRT technique increases as

the quantity of interest becomes rarer. Thus, we determine that $M_{HRT} \gg M_{HRT\text{-}SOC}$, especially for rare regions. Consequently, we expect that, for the regime of rare events and a fixed N, the backward time can be neglected compared to the forward cost of the HRT, which is presented in Fig. 3.

When the backward time dominates the forward time of the HRT, we propose an improved version of the HRT-SOC estimator. We call this version the aggregate method (HRT-SOC-AG), which aims to reduce the backward cost without considerably affecting the variance reduction.

4.2 Aggregate method

The idea for the aggregate method is to divide the sum S_N into B blocks and compute the controls for each block rather than for each X_i, $i = 1, \ldots, N$. Doing so reduces the backward cost from $N \times K \times T_b$ to $B \times K \times T_b$. In other words, if we select B blocks, such that $B \leq N$, we consider the following dynamics:

$$S_{n_m + b_{m+1}} = S_{n_m} + \sum_{i=n_m+1}^{n_m + b_{m+1}} X_i, \quad m = 0, 1, \ldots, B-1, \quad (43)$$

where $n_m = \sum_{j=1}^{m} b_j$, and b_m, $m = 1, 2, \ldots, B$ are chosen such that $n_B = \sum_{j=1}^{B} b_j = N$. We adopted the same control $\mu_m(S_{n_{m-1}})$ for each X_i from $i = n_{m-1} + 1$ to $i = n_m$. Thus, the B new controls $\mu_1^X, \ldots, \mu_B^X$ are defined such that

$$\begin{aligned} \mu_i &= \mu_m^X \text{ for } n_{m-1} < i \leq n_m, \\ & i = 1, \ldots, N, \ m = 1, \ldots, B. \end{aligned} \quad (44)$$

With this proposed approach, we decreased the cost of the backward step with the price of increasing the variance.
To determine $\mu_1^X, \ldots, \mu_B^X$, we used the dynamics proposed in (43) instead of the initial dynamics (11) to define a reformulated dynamic programming equation. We employed the same steps as those followed in the proof of the proposition, but instead of conditioning on X_{n+1}, we conditioned on $X_{n_m+1}, \ldots, X_{n_m+b_m}$. Applying the same control μ_{m+1} for each X_i, $i = n_m + 1, \ldots, n_m + b_m$, as explained in (44), we obtain

$$\begin{aligned} u(m, s_k) &= \min_{\mu \in (-\infty, 1)} \int_{[0,+\infty[^{b_m}} \frac{e^{-\mu \sum_{j=n_m+1}^{n_m+b_m} \Lambda_{X_j}(t_j)}}{(1-\mu)^{b_m}} \\ &\times \prod_{j=n_m+1}^{n_m+b_m} f_{X_j}(t_j)\, u\left(m+1, s_k + t_{n_m+1} + \cdots + t_{n_m+b_m}\right) \\ & dt_{n_m+1} \cdots dt_{n_m+b_m}. \end{aligned} \quad (45)$$

Instead of solving the above equation, we propose minimizing its approximate upper bound, which becomes clearer in the next two subsections.

4.3 OP at the output of diversity receivers in a log-normal environment without co-channel interference

The computation of the OP at the output of diversity receivers is equivalent to evaluating the CDF of the sum of the SNRs. Therefore, the interest in the first application is in the estimation of the left-tail region of the following form:

$$\mathbb{P}\left(\sum_{i=1}^{N} X_i \leq \gamma_{\text{th}}\right). \tag{46}$$

We compared the approach to the HRT technique (see Ben Rached et al. 2015a) and the exponential twisting estimator (see Asmussen et al. 2016). We also used the improved version to achieve better results. When applying the aggregate method, instead of solving (45), we propose to minimize an approximate upper bound of it. More precisely, for the i.i.d. log-normal case,

$$\sum_{j=n_m+1}^{n_m+b_m} \Lambda_{X_j}(t_j) \leq \Lambda_X\left(\sum_{j=n_m+1}^{n_m+b_m} t_j\right), \quad t_j > 0. \tag{47}$$

holds asymptotically, i.e. when the sum $\sum_{j=n_m+1}^{n_m+b_m} t_j$ is sufficiently small, where X has the same distribution as X_j, $j = n_m + 1, \ldots, n_m + b_m$. This result can be proven using the asymptotic result of the tail of a Normal distribution in Asmussen et al. (2011). Using the inequality (47), the twisting parameters μ_{m+1}^X are then selected as the argmin of the following approximated upper bound

$$u(m, s_k) \gtrapprox \min_{\mu \in (-\infty, 1)} \int_{[0, S-s_k]} \frac{e^{-\mu \, \Lambda_X(y)}}{(1-\mu)^{b_m}} \, f_{\sum_{j=n_m+1}^{n_m+b_m} X_j}(y)$$
$$\times u(n_m + b_m, s_k + y) \, dy, \tag{48}$$

where $f_{\sum_{j=n_m+1}^{n_m+b_m} X_j}(\cdot)$ is the PDF of $\sum_{j=n_m+1}^{n_m+b_m} X_j$. Given that the PDF of sums of i.i.d. log-normal RVs is unknown, we suggest approximating it using a univariate log-normal PDF $f_{Y_{m+1}}(\cdot)$, whose parameters are computed using moment matching (see Cobb et al. 2012).

Finally, we obtain

$$u(m, s_k) \approx \min_{\mu \in (-\infty, 1)} \int_{[0, S-s_k]} \frac{e^{-\mu \, \Lambda_X(y)}}{(1-\mu)^{b_m}} \, f_{Y_{m+1}}(y)$$
$$\times u(n_m + b_m, s_k + y) \, dy, \quad m = 0, \cdots, B - 1. \tag{49}$$
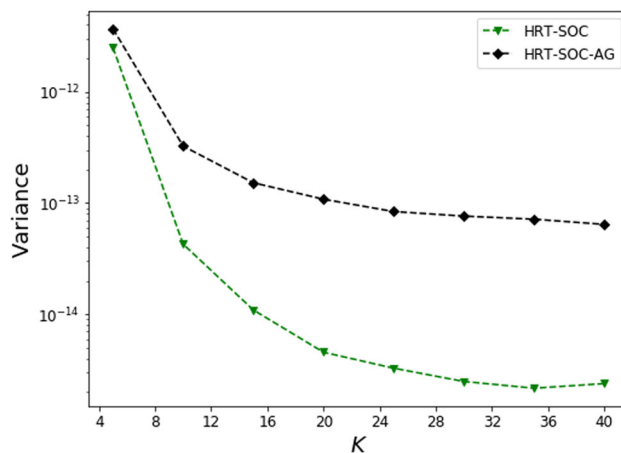

Fig. 1 Variance as a function of K with the following parameters: $N = 10$, $m = 0$ dB, $\sigma = 3$ dB, TOL $= 0.05$, and $b = 2$

Moreover, $\mu_1^X, \ldots, \mu_B^X$ are obtained as follows:

$$\mu_{m+1}^X(s_k) \approx \arg\min_{\mu \in (-\infty, 1)} \int_{[0, S-s_k]} \frac{e^{-\mu \, \Lambda_X(y)}}{(1-\mu)^{b_m}} \, f_{Y_{m+1}}(y)$$
$$\times u(n_m + b_m, s_k + y) \, dy, \quad m = 0, \ldots, B - 1. \tag{50}$$

Figure 2 plots the number of samples, required for the various approaches, to ensure TOL $= 5\%$ as a function of γ_{th}. The range of γ_{th} ensures a range of probabilities between 2×10^{-12} and 6×10^{-6}. For the aggregate method, we selected a constant parameter b (i.e., $b_m = 2$ for all $m = 1, \ldots, B$ with $B = \frac{N}{2}$).

The choice of the parameter $K = 20$ is motivated by Fig. 1, which plots the variance as a function of K. A larger K results in a smaller variance. The backward step is costly when K is large. Further, the variance reduction for $K > 20$ is minimal compared to the increased cost of solving the backward problem.

Figure 2 indicates that the number of samples required by naive MC simulations increases faster as the threshold decreases. In addition, the HRT-SOC approach requires the smallest number of simulation runs and saves a considerable number of samples compared to the HRT approach. For example, the number of simulations reduces by about 41,775 times for a small threshold (4 dB), corresponding to an OP value of 2×10^{-12}. In contrast, the HRT-SOC-AG requires an additional number of samples, compared to the HRT-SOC approach, to reach a 5% relative error, indicating that the variance has increased as expected. However, we still obtained better variance reduction compared to the HRT technique.

We further studied the computational work for each method. Figure 3 plots the total time required for the exponential twisting, HRT, HRT-SOC, and HRT-SOC-AG techniques to ensure a 5% relative error as a function of the threshold. We also plotted the time required by the HRT-SOC and HRT-
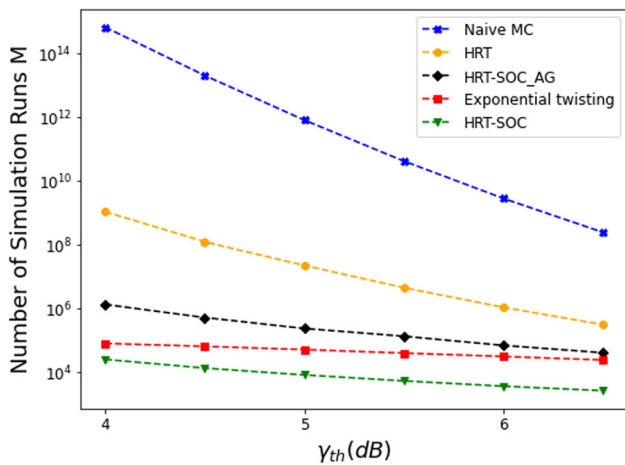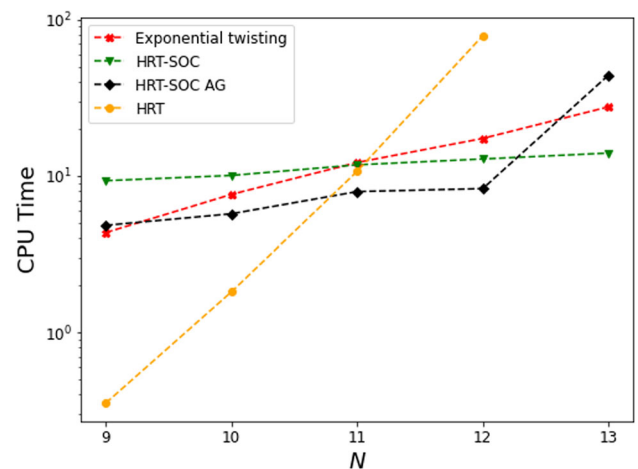
Fig. 2 Number of required simulation runs for a 5% relative error with the following parameters: $N = 10$, $K = 20$, $m = 0$ dB, $\sigma = 3$ dB, TOL = 0.05, and $b = 2$
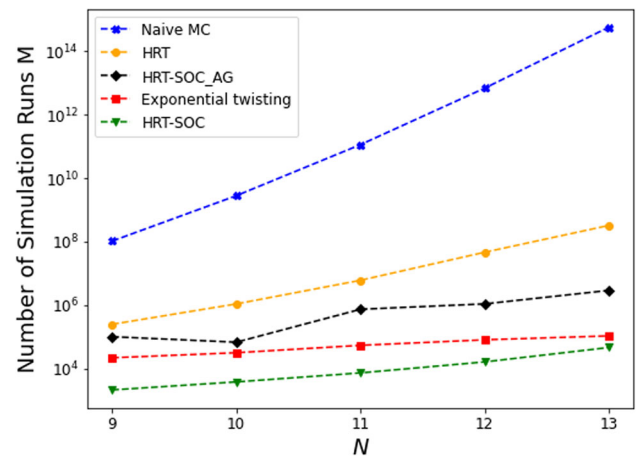
Fig. 4 Number of required simulation runs for a 5% relative error with the following parameters: $K = 20$, $\gamma_{\text{th}} = 6$ dB, $m = 0$ dB, $\sigma = 3$ dB, and TOL = 0.05
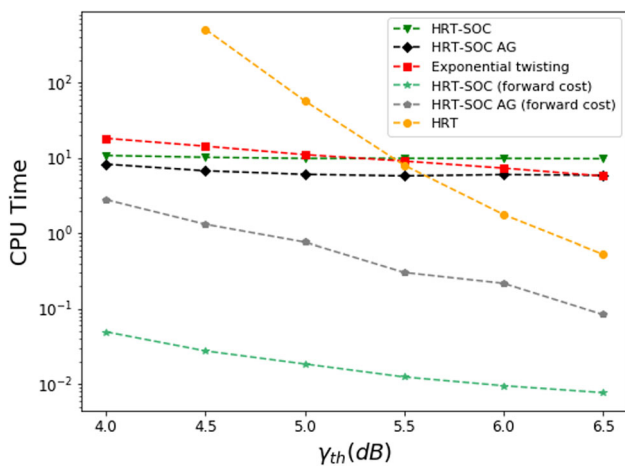
Fig. 3 CPU time required for a 5% relative error with the following parameters: $N = 10$, $K = 20$, $m = 0$ dB, $\sigma = 3$ dB, TOL = 0.05, and $b = 2$

Fig. 5 CPU time required for a 5% relative error with the following parameters: $K = 20$, $\gamma_{\text{th}} = 6$ dB, $m = 0$ dB, $\sigma = 3$ dB, and TOL = 0.05

SOC-AG techniques to demonstrate the time required for the backward step compared to that required for the forward step.

The proposed estimator is the best for computational time for small thresholds (corresponding to an OP of less than 3.6×10^{-8}). As the event becomes rarer, the time gap between the proposed approach and other IS techniques increases significantly. Additionally, Figs. 2 and 3 reveal that the HRT approach requires numerous samples to estimate the OP of the order of 2×10^{-12} with good accuracy. However, for an OP greater than 3.6×10^{-8}, the proposed approach is more expensive than others due to the additional computational time for the backward step for each threshold, which exceeds the time the remaining techniques when the number of samples is not sufficiently large. Nevertheless, this was enhanced when we used the improved version. The HRT-SOC-AG reduces the CPU time by about 1.7 times compared

to the HRT-SOC approach for $\gamma_{\text{th}} \geq 5$ dB. Thus, with this choice of b, the efficiency of the aggregate method regarding time reduction exceeds the loss in variance. This choice of $b_m, m = 1, \ldots, B$ is not optimal. Despite this, it provides better results than the HRT-SOC approach.

Another possible experiment is to study the efficiency as a function of the number N of antennae for a fixed threshold and investigate the number of simulation runs required for each method and the computational time (Figs. 4 and 5, respectively). The range of the OP is between 10^{-5} and 2.5×10^{-12} when using a range between nine and 13 antennae and a fixed threshold $\gamma_{\text{th}} = 6$ dB. For the aggregate method, we used $b_m = 2$, $m = 1, \ldots, \frac{N}{2}$ for an even-numbered N and $b_m = 2, m = 1, \ldots, \frac{N-3}{2}$, $b_{\frac{N-1}{2}} = 3$ for an odd-numbered N.

Figure 4 indicates that the HRT-SOC approach is more efficient and requires fewer simulation runs than the HRT and the exponential twisting approaches. For $N = 13$, the proposed method requires 7455 times fewer simulation runs than the HRT technique to meet the same accuracy requirements. In addition, the variance reduction for the HRT-SOC-AG technique depends on whether N is odd or even. Moreover, the HRT-SOC-AG method requires more simulation runs than the HRT-SOC technique to reach a fixed precision TOL, but it is more efficient in terms of CPU time for $N \leq 12$. When the event becomes rarer (for small γ_{th} and large N), the improved approach with a fixed choice of b becomes less efficient in terms of CPU time than the HRT-SOC approach. In these cases, the number of samples is large enough that the backward time is neglected. Thus, reducing the variance rather than the cost of the backward step is more efficient. These results demonstrate that the choice of b_m, $m = 1, \ldots, B$ is crucial and should be adaptively chosen to provide better results. More precisely, for fixed parameters γ_{th}, TOL and N, the following optimization problem should be solved:

$$\min_{b, M, K} \quad B \times K \times T_b + M_{\text{HRT-SOC-AG}}(b) \times T_f, \quad (51)$$

such that

$$C^2 \frac{\text{Var}\left[T_{\text{HRT-SOC-AG}}(b)\right]}{M_{\text{HRT-SOC-AG}}(b)\alpha^2} \leq \text{TOL}^2.$$

The above optimization problem reveals that an optimal choice of b_m in the case of a very rare event is $b_m = 1$, $m = 1, \ldots, B$, where $B = N$. However, when the event becomes less rare, an optimal choice of B is to take a single block (i.e., $b_1 = N$). By doing so, the HRT-SOC-AG technique reduces to the HRT technique because the controls are state-independent in this case. Future work can be devoted to solving the previous optimization problem. Using optimal values of b_m, we expect the HRT-SOC-AG estimator to achieve better performance.

4.4 OP in the presence of co-channel interference in a log-normal environment for SISO systems

We consider a SISO system and recall that the OP in the presence of co-channel interference and noise is expressed as follows:

$$P_{out} = \mathbb{E}\left[F_{X_0}\left(\gamma_{\text{th}} \left(\sum_{n=1}^{N} X_n + \eta \right) \right) \right],$$

where $X_1, \ldots, X_N$ are the interfering power signal and are assumed to be i.i.d. log-normal RVs with parameters m and σ^2.
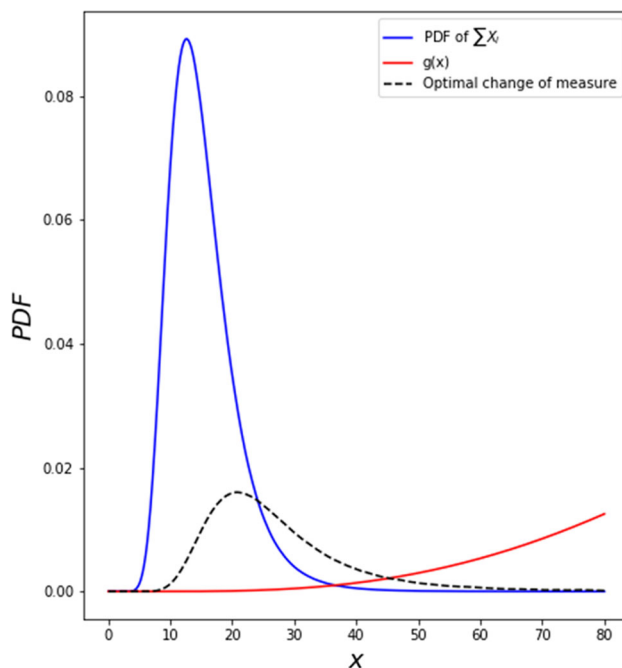

Fig. 6 Motivation for using IS with $N = 10$, $m_0 = 10$ dB, $\sigma_0 = 4$ dB, $m = 0$ dB, $\sigma = 4$ dB, $\gamma_{\text{th}} = -18$ dB, and $\eta = -10$ dB

The PDF of $\sum_{i=1}^{N} X_i$ is denoted by $f_{\sum_{i=1}^{N} X_i}(\cdot)$. To motivate the need for IS to efficiently estimate P_{out}, Fig. 6 plots the quantities $f_{\sum_{i=1}^{N} X_i}$, g, and the optimal IS PDF, which is proportional to $g f_{\sum_{i=1}^{N} X_i}$. The product $g f_{\sum_{i=1}^{N} X_i}$ in Fig. 6 is not normalized (i.e., it is an unnormalized PDF).

Sampling from the original PDF of $\sum_{i=1}^{N} X_i$ is not efficient (i.e., when sampling from the original PDF, most samples fall in the region where g takes almost zero values). Hence, the computation of P_{out} behaves like a rare event problem and can be addressed using the proposed HRT-SOC technique. The comparison is made concerning the estimator of Ben Rached et al. (2017), which is based on a covariance matrix scaling (CS) technique. It transforms the problem of evaluating the OP to computing the probability that a sum of correlated log-normal RVs exceeds a certain threshold. The estimator in Ben Rached et al. (2017) is given by

$$T_{\text{CS}}(\mathbf{Z}) = \mathbb{1}_{\left(\sum_{i=0}^{N} \exp(Z_i) \geq 1/\gamma_{th}\right)} L(Z_0, \ldots, Z_N), \quad (52)$$

where $\mathbf{Z} = (Z_0, Z_1, \ldots, Z_N)^t$,

$$Z_i = \begin{cases} \log(X_i) - \log(X_0) & i = 1, 2, \ldots, N \\ \log(\eta) - \log(X_0) & i = 0 \end{cases}, \text{ and}$$

$$L(Z_0, Z_1, \ldots, Z_N) = \frac{\exp\left(-\frac{\theta}{2}(\mathbf{Z} - \mathbf{m})^t \Sigma^{-1}(\mathbf{Z} - \mathbf{m})\right)}{(1 - \theta)^{(N+1)/2}}. \quad (53)$$

The expressions of $\mathbf{m}$, Σ, and θ are given in BenRached et al. (2017, eqs. (6), (7), (19)) respectively.

We also compared the proposed approach to the exponentially tilted (ET) estimator of Botev and l'Ecuyer (2017). We also used the HRT-SOC-AG method proposed in the previous subsection to further improve the computational work of the HRT-SOC technique. The reformulated dynamic programming equation is

$$
\begin{aligned}
u(m, s_k) = \min_{\mu \in (-\infty, 1)} & \int_{(0, +\infty)^{b_m}} \frac{e^{-\mu \sum_{j=n_m+1}^{n_m+b_m} \Lambda_{X_j}(t_j)}}{(1-\mu)^{b_m}} \\
& \times \prod_{j=n_m+1}^{n_m+b_m} f_{X_j}(t_j) u \left(m+1, s_k + \sum_{j=n_m+1}^{n_m+b_m} t_j \right) \\
& dt_{n_m+1} \dots dt_{n_m+b_m}.
\end{aligned}
\tag{54}
$$

Next, using the following inequality, proven in Juneja and Shahabuddin (2002), which is particularly satisfied in the case of i.i.d. log-normal RVs and holds for $\sum_{j=n_m+1}^{n_m+b_m} t_j$ that are large enough:

$$
\sum_{j=n_m+1}^{n_m+b_m} \Lambda_{X_j}(t_j) \geq \Lambda_X \left(\sum_{j=n_m+1}^{n_m+b_m} t_j \right) - \epsilon, \ t_j > 0,
$$
$$
\text{for all } \epsilon > 0,
\tag{55}
$$

we can write

$$
\begin{aligned}
u(m, s_k) \approx \min_{\mu \in (-\infty, 1)} & \int_{(0, +\infty)} \frac{e^{-\mu \Lambda_X(y)}}{(1-\mu)^{b_m}} f_{Y_{m+1}}(y) \\
& \times u(n_m + b_m, s_k + y) \ dy, \ m = 0, \dots, B-1.
\end{aligned}
\tag{56}
$$

The large value of $\sum_{j=n_m+1}^{n_m+b_m} t_j$ is motivated by Fig. 6, which illustrates that the change of measure tends to increase the value of the sum in the regime of rare events. We studied the efficiency of the four IS schemes regarding the number of samples necessary to ensure a fixed accuracy requirement. To this end, Fig. 7 plots the number of samples to ensure TOL = 5% as a function of γ_{th}. This figure reveals that the HRT-SOC approach saves numerous samples compared to other approaches. For instance, the CS technique requires approximately 2000 times as many simulations as the HRT-SOC scheme needs. The aggregate method did not affect the variance reduction.

We further investigated the gain in terms of the required computational time. Figure 8 presents the total CPU time needed by the four techniques to achieve the fixed accuracy TOL. The HRT-SOC approach requires less CPU time than the ET approach for the range of considered thresholds. In particular, when $\gamma_{\text{th}} = -30$ dB, it is 13 times more efficient than the ET scheme. Compared to the CS approach, the HRT-SOC technique is more efficient when $\gamma_{\text{th}} < -25$ dB, corresponding to an OP less than 3×10^{-8}.
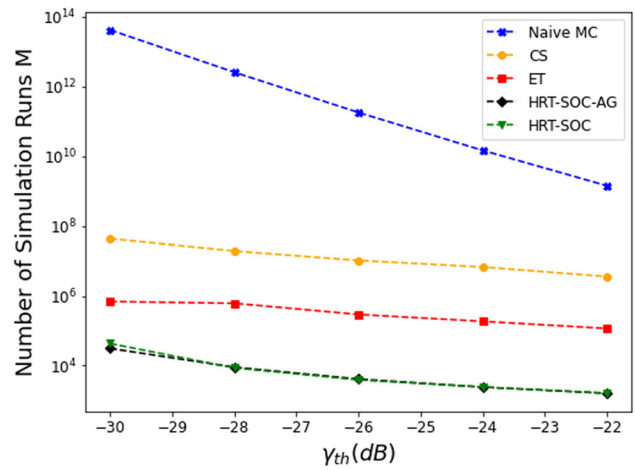

Fig. 7 Number of required simulation runs with the following parameters: $N = 10$, $K = 20$, $S = 40$, TOL = 0.05, $\eta = -10$ dB, $m_0 = 10$ dB, $\sigma_0 = 4$ dB, $m = 0$ dB, and $\sigma = 4$ dB
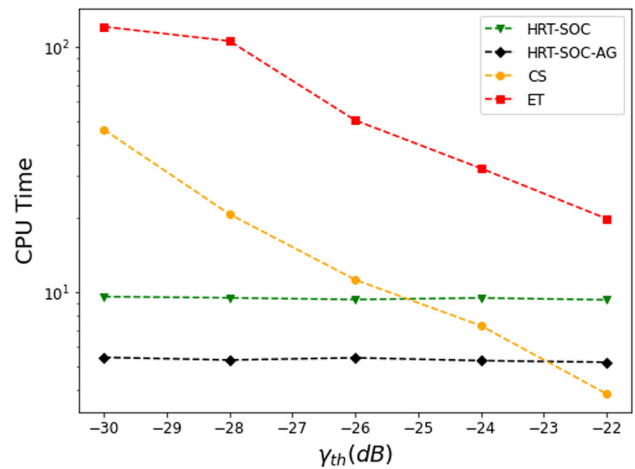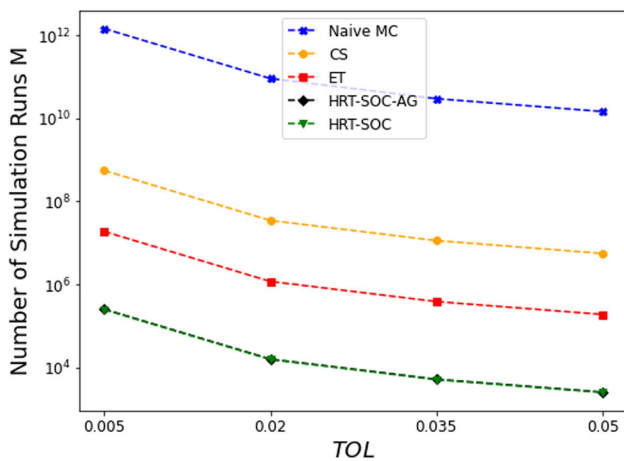

Fig. 8 CPU time required for a 5% relative error with the following parameters: $N = 10$, $K = 20$, $S = 40$, TOL = 0.05, $\eta = -10$ dB, $m_0 = 10$ dB, $\sigma_0 = 4$ dB, $m = 0$ dB, and $\sigma = 4$ dB

The required computational time for the HRT-SOC technique is almost the same in the considered threshold range, whereas the CS and ET approaches require much more time as the threshold decreases. Moreover, the HRT-SOC-AG technique requires less time than the HRT-SOC technique using $b = 2$ to estimate the quantity of interest α. Therefore, the improved approach widens the region over which the proposed approach outperforms the CS approach.

In the last experiment, we studied the influence of varying the accuracy TOL on the proposed and other IS approaches. To this end, Figs. 9 and 10 present the number of simulation runs and CPU time needed when varying TOL for a fixed γ_{th} and N. This choice makes the OP approximately equal to 10^{-7}.

Figure 9 confirms the high gains of the proposed methods compared to all other IS approaches. Our approaches are

Fig. 9 Number of required simulation runs with the following parameters: $N = 10$, $K = 20$, $S = 40$, $\gamma_{\text{th}} = -24$ dB, $\eta = -10$ dB, $m_0 = 10$ dB, $\sigma_0 = 4$ dB, $m = 0$ dB, and $\sigma = 4$ dB
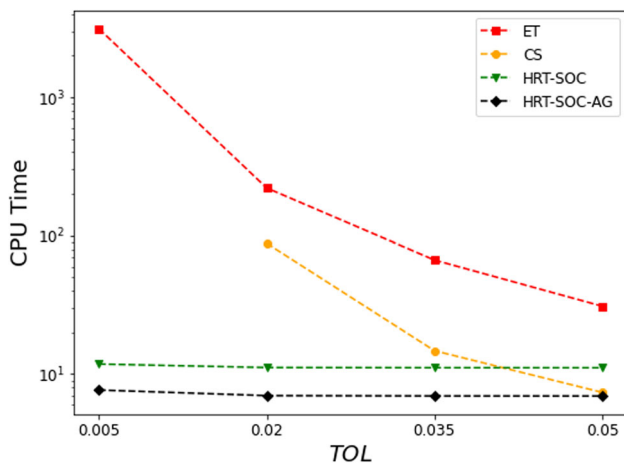
Fig. 10 CPU time for a 5% relative error with the following parameters: $N = 10$, $K = 20$, $S = 40$, $\gamma_{\text{th}} = -24$ dB, $\eta = -10$ dB, $m_0 = 10$ dB, $\sigma_0 = 4$ dB, $m = 0$ dB, and $\sigma = 4$ dB

2000 times (respectively 65 times) more efficient than the CS (respectively the ET) approaches for all values of TOL. Furthermore, Fig. 10 demonstrates that the required time for the proposed methods compared to the other algorithms remains unchanged for the considered range of TOL. Moreover, similarly to the previous conclusions, the computational time required by the proposed algorithm is less than that needed by the ET algorithm for all TOL. Additionally, the superior performance of the method compared to the CS approach is critical for small values of TOL. Finally, the HRT-SOC-AG method increases the threshold, below which the proposed method performs better than the CS approach, from 0.045 to 0.058.

5 Conclusions

Summary

We developed a generic state-dependent IS algorithm to efficiently estimate rare event quantities that could be written as an expectation of a functional of the sums of independent RVs. These problems have applications in the performance analysis of wireless communications systems operating over fading channels. Within a preselected class of a change of measures, the optimal IS parameters are determined via the connection to an SOC formulation. The numerical experiments verified the ability of the proposed approach to accurately and efficiently estimate the quantity of interest in the rare event regime. The proposed approach yields a substantial variance reduction compared with other well-known estimators. Additionally, the estimator requires less CPU time than the other proposed approaches in rare regions. We also proposed an aggregate method to improve efficiency further in terms of computational time.

Possible extensions

For future research, the present work can be extended in many directions. One possible direction is to optimize the aggregate method by solving the optimization problem (51).

A further interesting extension is to consider multivariate RVs when estimating the quantity of interest. In this case, RVs should be mutually independent, but the components of each RV are not necessarily independent. As the backward cost increases exponentially with the dimensions, we could employ cheaper approximation methods to calculate the controls, such as neural networks.

References

Alemany, R., Bolancé, C., Guillen, M.: A nonparametric approach to calculating value-at-risk. Insur. Math. Econ. **52**(2), 255–262 (2013)

Asmussen, S., Binswanger, K.: Simulation of ruin probabilities for subexponential claims. ASTIN Bull. J. IAA **27**(2), 297–318 (1997)

Asmussen, S., Glynn, P.W.: Stochastic Simulation: Algorithms and Analysis. Springer, Berlin (2007)

Asmussen, S., Kortschak, D.: On error rates in rare event simulation with heavy tails. In: Proceedings of the IEEE Winter Simulation Conference (WSC), pp. 1–11 (2012)

Asmussen, S., Kortschak, D.: Error rates and improved algorithms for rare event simulation with heavy Weibull tails. Methodol. Comput. Appl. Probab. **17**(2), 441–461 (2015)

Asmussen, S., Kroese, D.P.: Improved algorithms for rare event simulation with heavy tails. Adv. Appl. Probab. **38**(2), 545–558 (2006)

Asmussen, S., Binswanger, K., Højgaard, B., et al.: Rare events simulation for heavy-tailed distributions. Bernoulli **6**(2), 303–322 (2000)

Asmussen, S., Blanchet, J., Juneja, S., Rojas-Nandayapa, L.: Efficient simulation of tail probabilities of sums of correlated Lognormals. Ann. Oper. Res. **189**(1), 5–23 (2011)

Asmussen, S., Jensen, J.L., Rojas-Nandayapa, L.: Exponential family techniques for the lognormal left tail. Scand. J. Stat. **43**(3), 774–787 (2016)

Bassamboo, A., Juneja, S., Zeevi, A.: On the inefficiency of state-independent importance sampling in the presence of heavy tails. Oper. Res. Lett. **35**(2), 251–260 (2007)

Beaulieu, N.C., Luan, G.: On the Marcum Q-function behavior of the left tail probability of the lognormal sum distribution. In: Proceedings of the IEEE International Conference on Communications (ICC) (2020)

Ben Issaid, C., Ben Rached, N., Kammoun, A., Alouini, M.-S., Tempone, R.: On the efficient simulation of the distribution of the sum of Gamma-Gamma variates with application to the outage probability evaluation over fading channels. IEEE Trans. Commun. **65**(4), 1839–1848 (2017)

Ben Rached, N., Benkhelifa, F., Alouini, M.-S., Tempone, R.: A fast simulation method for the Log-normal sum distribution using a hazard rate twisting technique. In: Proceedings of the IEEE International Conference on Communications (ICC), pp. 4259–4264 (2015a)

Ben Rached, N., Kammoun, A., Alouini, M.-S., Tempone, R.: Unified importance sampling schemes for efficient simulation of outage capacity over generalized fading channels. IEEE J. Sel. Top. Signal Process. **10**(2), 376–388 (2015b)

Ben Rached, N., Kammoun, A., Alouini, M.-S., Tempone, R.: On the efficient simulation of outage probability in a Log-normal fading environment. IEEE Trans. Commun. **65**(6), 2583–2593 (2017)

Ben Rached, N., Benkhelifa, F., Kammoun, A., Alouini, M.-S., Tempone, R.: On the generalization of the hazard rate twisting-based simulation approach. Stat. Comput. **28**(1), 61–75 (2018)

Ben Rached, N., Kammoun, A., Alouini, M.-S., Tempone, R.: An accurate sample rejection estimator of the outage probability with equal gain combining. IEEE Open J. Commun. Soc. **1**, 1022–1034 (2020a)

Ben Rached, N., MacKinlay, D., Botev, Z., Tempone, R., Alouini, M.-S.: A universal splitting estimator for the performance evaluation of wireless communications systems. IEEE Trans. Wirel. Commun. **19**(7), 4353–4362 (2020b)

Ben Rached, N., Haji-Ali, A.-L., Rubino, G., Tempone, R.: Efficient importance sampling for large sums of independent and identically distributed random variables. Stat. Comput. **31**(6), 4353–4362 (2021)

Blanchet, J., Lam, H.: State-dependent importance sampling for rare-event simulation: an overview and recent advances. Surv. Oper. Res. Manag. Sci. **17**(1), 38–59 (2012)

Blanchet, J., Li, C.: Efficient rare event simulation for heavy-tailed compound sums. ACM Trans. Model. Comput. Simul. (TOMACS) **21**(2), 1–23 (2011)

Blanchet, J.H., Liu, J.: Efficient simulation for large deviation probabilities of sums of heavy-tailed increments. In: Proceedings of the IEEE Winter Simulation Conference, pp. 757–764 (2006)

Blanchet, J.H., Liu, J.: State-dependent importance sampling for regularly varying random walks. Adv. Appl. Probab. **40**(4), 1104–1128 (2008)

Botev, Z., l'Ecuyer, P.: Accurate computation of the right tail of the sum of dependent Log-normal variates. In: 2017 Winter Simulation Conference (WSC), pp. 1880–1890. IEEE (2017)

Chan, J.C.C., Kroese, D.P.: Rare-event probability estimation with conditional Monte Carlo. Ann. Oper. Res. **189**(1), 43–61 (2011)

Chatterjee, A., Parida, P., Das, S.S.: Downlink signal-to-interference ratio and spectral efficiency of MIMO cellular networks using truncated Lognormal approximation. IEEE Syst. J. **13**(1), 76–87 (2018)

Cobb, B.R., Rumí, R., Salmerón, A.: Approximating the distribution of a sum of Log-normal random variables. Stat. Comput. **16**(3), 293–308 (2012)

Constantinescu, C., Samorodnitsky, G., Zhu, W.: Closed-form ruin probabilities in classical risk models with gamma claims (2016)

Dupuis, P., Wang, H.: Importance sampling, large deviations, and differential games. Stoch. Int. J. Probab. Stoch. Process. **76**(6), 481–508 (2004)

Dupuis, P., Wang, H.: Subsolutions of an Isaacs equation and efficient schemes for importance sampling. Math. Oper. Res. **32**(3), 723–757 (2007)

Dupuis, P., Wang, H., et al.: Dynamic importance sampling for uniformly recurrent Markov chains. Ann. Appl. Probab. **15**(1A), 1–38 (2005)

Dupuis, P., Leder, K., Wang, H.: Importance sampling for sums of random variables with regularly varying tails. ACM Trans. Model. Comput. Simul. (TOMACS) **17**(3), 14 (2007)

Ghamami, S., Ross, S.M.: Improving the Asmussen–Kroese-type simulation estimators. J. Appl. Probab. **49**(4), 1188–1193 (2012)

Glasserman, P.: Monte Carlo Methods in Financial Engineering, vol. 53. Springer, Berlin (2004)

Hammouda, C.B., Rached, N.B., Tempone, R., Wiechert, S.: Optimal importance sampling via stochastic optimal control for stochastic reaction networks (2021). arXiv preprint arXiv:2110.14335

Hartinger, J., Kortschak, D.: On the efficiency of the Asmussen–Kroese-estimator and its application to stop-loss transforms. Blätter DGVFM **30**(2), 363 (2009)

Hartmann, C., Richter, L., Schütte, C., Zhang, W.: Variational characterization of free energy: theory and algorithms. Entropy **19**(11), 626 (2017)

Juneja, S.: Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error. Queueing Syst. **57**(2), 115–127 (2007)

Juneja, S., Shahabuddin, P.: Simulating heavy tailed processes using delayed hazard rate twisting. ACM Trans. Model. Comput. Simul. (TOMACS) **12**(2), 94–118 (2002)

Juneja, S., Shahabuddin, P.: Rare-event simulation techniques: an introduction and recent advances. Handb. Oper. Res. Manag. Sci. **13**, 291–350 (2006)

Karthyek Rajhaa, A.M., Juneja, S.: State-independent importance sampling for estimating large deviation probabilities in heavy-tailed random walks. In: Proceedings of the IEEE 6th International ICST Conference on Performance Evaluation Methodologies and Tools, pp. 127–135 (2012)

Kroese, D.P., Taimre, T., Botev, Z.I.: Handbook of Monte Carlo Methods. Wiley, Hoboken (2011)

López-Salcedo, J.A.: Simple closed-form approximation to Ricean sum distributions. IEEE Signal Process. Lett. **16**(3), 153–155 (2009)

Murthy, K.R.A., Juneja, S., Blanchet, J.: State-independent importance sampling for random walks with regularly varying increments. Stoch. Syst. **4**(2), 321–374 (2015)

Singh, R.K.: Exact analytical ber expression in FSO using α-μ distribution proposed as an approximation of Gamma–Gamma pdf. In: Proceedings of the 27th Wireless and Optical Communication Conference (WOCC) (2018)

Sun, L., Hong, L.J.: A general framework of importance sampling for value-at-risk and conditional value-at-risk. In: Proceedings of the 2009 Winter Simulation Conference (WSC), pp. 415–422. IEEE (2009)

Xiao, Z., Zhu, B., Cheng, J., Wang, Y.: Outage probability bounds of EGC over dual-branch non-identically distributed independent Lognormal fading channels with optimized parameters. IEEE Trans. Veh. Technol. **68**(8), 8232–8237 (2019)

Yilmaz, F., Alouini, M.-S.: A unified MGF-based capacity analysis of diversity combiners over generalized fading channels. IEEE Trans. Commun. **60**(3), 862–875 (2012)

Zhu, B., Zhang, Z., Wang, L., Dang, J., Wu, L., Cheng, J., Li, G.Y.: Right tail approximation for the distribution of Lognormal sum and its applications. In: Proceedings of the IEEE Globecom Workshops (2020)