Additional Results on the Hazard Rate Twisting-Based Simulation Approach

Nadhir Ben Rached, Fatma Benkhelifa, Abla Kammoun, Mohamed-Slim Alouini, and Raul Tempone
Computer, Electrical and Mathematical Science and Engineering (CEMSE) Division
King Abdullah University of Science and Technology (KAUST)
Thuwal, Makkah Province, Saudi Arabia
{nadhir.benrached, fatma.benkhelifa, abla.kammoun, slim.alouini, raul.tempone}@kaust.edu.sa

***Abstract*—Estimating the probability that a sum of random variables (RVs) exceeds a given threshold is a well-known challenging problem. Closed-form expressions for the sum distribution do not generally exist, which has led to an increasing interest in simulation approaches. A crude Monte Carlo (MC) simulation is the standard technique for the estimation of this type of probability. However, this approach is computationally expensive, especially when dealing with rare events (i.e events with very small probabilities). An alternative approach to naive MC simulations is represented by the use of variance reduction techniques, known for their efficiency in requiring less computations for achieving the same accuracy requirement. Most of these methods have thus far been proposed to deal with specific settings under which the RVs belong to particular classes of distributions. In this paper, we propose a new Importance Sampling (IS) simulation technique based on the well-known hazard rate twisting approach, that presents the advantage of being asymptotically optimal for any arbitrary RVs. The wide scope of applicability of the proposed method is mainly due to our particular way of selecting the twisting parameter. It is worth observing that this interesting feature is rarely satisfied by variance reduction algorithms whose performances were only proven under some restrictive assumptions. It comes along with a good efficiency, illustrated by some selected simulation results comparing the performance of our method with that of an algorithm based on a conditional MC technique.**

***Index Terms*—Crude Monte Carlo, rare events, Importance Sampling, hazard rate twisting, asymptotically optimal, twisting parameter.**

I. Introduction

The performance analysis of communication systems generally requires analyzing the statistics of sums of Random Variables (RVs). For instance, the received signal-to-noise-ratio (SNR) at the output of some schemes based on diversity techniques, such as Maximum ratio Combining (MRC) and Equal Gain combining (EGC), is modelled as a sum of fading variates [1]. Unfortunately, the statistics of the sum distribution for most of the challenging problems are generally intractable and unknown. This is for instance the case of the Log-normal and the Weibull RVs, which are frequently encountered in various applications of digital communications [2]–[7]. In order to tackle this issue, several analytical approaches, which consists in determining accurate closed-form approximations, approaching the statistics of the sum of these RVs were proposed [8]–[14]. However, these analytical approaches presents the inconvenient of being specific to the problem under study, thereby limiting their practical interest. An alternative to these analytical methods is constituted by the class of Monte Carlo methods.

The naive Monte Carlo (MC) simulation is the standard technique to estimate the probability that a sum of RVs exceeds a given threshold. However, this approach requires substantial computational simulations, especially when extremely small probabilities are considered. To fill these gaps, a simulation approach termed as Importance Sampling (IS) which helps improve the computational efficiency of the naive MC simulation technique [15] was proposed. The basic idea behind the IS techniques is to change the underlying sampling distribution in a way to achieve a substantial variance reduction of the IS estimator. Many research efforts have been carried out to propose efficient IS algorithms. For instance, among the first works dealing with the application of IS to the field of digital communications are those in [16] and [17] which have proposed methods based respectively on scaling the variance and shifting the mean of the original probability measure. An extension of [16] was performed in [18] where a composite IS technique was derived. In [19], the asymptotic efficiency of five different IS techniques estimating the Bit Error Rate (BER) was studied. The exponential twisting technique, derived from the large deviation theory, is an interesting IS change of measure since in most of the cases it yields "optimal" asymptotic results [20] [21]. For instance, it was used to estimate the BER of direct-detection optical systems employing avalanche photodiode (APD) receivers in [22].

However, the scope of the exponential twisting is limited to that of distributions with finite Moment Generating Function (MGF). Thus, in the heavy-tailed setting where the MGF is infinite, this approach is not applicable. Given that many heavy-tailed distributions, such as the Log-normal and the Weibull RVs, are frequently encountered in various applications in digital communications [2]–[7], many research efforts have been focused on developing alternatives methods which can be used to deal with the sum of RVs drawn from heavy-tailed distributions. This has pushed forward the development of efficient variance reduction algorithms dealing with distributions in the heavy-tailed class. [23]–[27]. Among the different variance reduction techniques, one of the most popular is that based on twisting the hazard rate of the original probability measure of each component in the summation. This technique has been applied by the authors in [28] to estimate

The authors are members of the KAUST Strategic Research Initiative on Uncertainty Quantification in Science and Engineering (SRI-UQ).

the sum of independent and identically distributed (i.i.d) RVs with subexponential decay.

In this paper, we propose to generalize the hazard rate twisting IS-based approach [28] to the problem of estimating the probability that a sum of independent but not necessarily identically distributed any RVs exceeds a given threshold. Unlike the previously cited works where the estimators' performances have been established when the RVs belong to particular classes of distributions, the proposed algorithm is proven to be applicable, without any restriction, to any arbitrary RVs. The rest of the paper is organized as follows. In section II, we state the problem setting and enumerate the main contributions. In Section III, the proposed hazard rate twisting technique is presented and the main result proving the asymptotic optimality criterion is stated in Theorem 1. In the same section, a study case is analysed with details. Finally, some selected simulation results are shown in Section IV to assess the performance of the proposed IS scheme.

II. Mathematical Background

A. Problem Setting

Let $X_1, X_2, ..., X_N$ be a sequence of independent but not necessarily identically distributed RVs. Let us denote the Probability Density Function (PDF) of each X_i by $f_i(\cdot)$, $i = 1, 2, ..., N$. Our objective is to efficiently estimate:

$$\alpha = \mathbb{P}\left(\sum_{i=1}^{N} X_i > \gamma_{th}\right) = P\left(S_N > \gamma_{th}\right), \tag{1}$$

for a sufficiently large threshold γ_{th}. The standard technique to estimate α is to use the naive MC estimator defined as:

$$\hat{\alpha}_{MC} = \frac{1}{M}\sum_{j=1}^{M} \mathbf{1}_{(S_N(\omega_j)>\gamma_{th})}, \tag{2}$$

where M is the number of simulation runs and $\mathbf{1}_{(\cdot)}$ defines the indicator function. $\{S_N(\omega_j)\}_{j=1}^{M}$ represent independent realizations of the RV $S_N = \sum_{i=1}^{N} X_i$ where for each realization, $j = 1, 2, ..., M$, the sequence $X_1(\omega_j), X_2(\omega_j), ..., X_N(\omega_j)$ are sampled independently according to the distributions $f_i(\cdot)$, $i = 1, 2, ..., N$, respectively. It is widely known that the naive MC simulation is extensively expensive for the estimation of rare events. In fact, from the Central Limit Theorem (CLT), it can be shown that the naive MC estimation with 10% relative precision requires more than $100/\alpha$ simulation runs. For instance, the number of samples to estimate a probability of order 10^{-9} should be more than 10^{11}, with an accuracy requirement of 10%. This has triggered the need for alternative methods to naive MC simulations with improved computational efficiency.

B. Importance Sampling

IS is a variance reduction technique which aims to increase the computational efficiency of the naive MC simulation [15]. The general concept of IS is to construct an unbiased estimator of the desired probability with much smaller variance than the naive estimator. In fact, this technique is based on performing a suitable change of the sampling distribution as follows

$$\begin{aligned}\alpha &= \int_{\mathbb{R}^N} \mathbf{1}_{(S_N>\gamma_{th})} \prod_{i=1}^{N} f_i(x_i) dx_1 dx_2 ... dx_N \\ &= \int_{\mathbb{R}^N} \mathbf{1}_{(S_N>\gamma_{th})} L\left(x_1, x_2, ..., x_N\right) \prod_{i=1}^{N} g_i(x_i) dx_1 dx_2 ... dx_N \\ &= \mathbb{E}_{p^*}\left[\mathbf{1}_{(S_N>\gamma_{th})} L\left(X_1, X_2, ..., X_N\right)\right],\end{aligned} \tag{3}$$

where the expectation is taken with respect to the new probability measure p^* under which the PDF of each X_i is $g_i(\cdot)$, and L is the likelihood ratio defined as

$$L\left(X_1, X_2, ..., X_N\right) = \prod_{i=1}^{N} \frac{f_i(X_i)}{g_i(X_i)}. \tag{4}$$

The rationale behind this change of measure is to enhance sampling important values which have more impact on the desired probability. Hence, emphasizing that important values are sampled frequently will result in a decrease of the variance of the IS estimator. The new IS estimator is defined as

$$\hat{\alpha}_{IS} = \frac{1}{M}\sum_{j=1}^{M} \mathbf{1}_{(S_N(\omega_j)>\gamma_{th})} L(X_1(\omega_j), ..., X_N(\omega_j)). \tag{5}$$

where $X_1(\omega_j), X_2(\omega_j), ..., X_N(\omega_j)$ are sampled, for each realization $j = 1, 2, ..., M$, independently according to the new sampling distributions whose PDFs are $g_i(\cdot)$, $i = 1, 2, ..., N$, respectively.

Generally, it is not obvious how to construct a new probability measure which results in a decrease of the variance of the IS estimator and hence an improvement of the computational efficiency. Besides, in order to evaluate the efficiency of the proposed approach, a criterion is required to be defined. Several criteria have been used in the literature, among which we distinguish the bounded relative error property [25], [27] and the asymptotic optimality property [25], [28]. In practice, it is it is difficult to achieve the bounded relative error property, unless working with RVs drawn from a specific set of distributions. Since, we address in the present paper the case of arbitrary RVs, we will rather consider the second criterion which is the asymptotic optimality. In the sequel, we denote by efficient algorithm, an algorithm which satisfies the asymptotic optimality criterion. Let us consider the sequence of the RVs $\{T_{\gamma_{th}}\}$ defined as

$$T_{\gamma_{th}} = \mathbf{1}_{(S_N>\gamma_{th})} L\left(X_1, ..., X_N\right). \tag{6}$$

From the non-negativity of the variance of $T_{\gamma_{th}}$, we get

$$\mathbb{E}_{p^*}\left[T_{\gamma_{th}}^2\right] \geq (\mathbb{P}(S_N > \gamma_{th}))^2. \tag{7}$$

Applying the logarithm on both side, we conclude that, for all p^*, we have

$$\frac{\log\left(\mathbb{E}_{p^*}\left[T_{\gamma_{th}}^2\right]\right)}{\log\left(\mathbb{P}\left(S_N > \gamma_{th}\right)\right)} \leq 2. \tag{8}$$

Hence, we say that α is asymptotically optimally estimated under the probability measure p^* if the above equation holds with equality as $\gamma_{th} \to +\infty$, that is

$$\lim_{\gamma_{th}\to\infty} \frac{\log\left(\mathbb{E}_{p^*}\left[T^2_{\gamma_{th}}\right]\right)}{\log\left(\mathbb{P}\left(S_N > \gamma_{th}\right)\right)} = 2. \tag{9}$$

It is worth mentioning that the naive simulation is not asymptotically optimal for the estimation of α since, in this case, the ratio in (9) is equal to 1.

The exponential twisting technique, which is derived from the large deviation theory, is the main IS framework dealing with light-tailed distributions, that is distributions whose tails decay at an exponential rate or faster. The exponential twisting by an amount $\theta > 0$ is given by

$$g_i\left(x\right) \triangleq f_{i,\theta}(x) = \frac{f_i(x)\exp(\theta x)}{M_{X_i}(\theta)}, \tag{10}$$

where $M_{X_i}(\theta)$ denotes the moment generating function (MGF) of the RV X_i, $i = 1, 2, ..., N$. In most of the cases, this technique achieves optimal efficiency results [20] [21].

In the case when the sequence $X_1, X_2, ..., X_N$ contains some heavy-tailed components, the exponential twisting change of measure is not feasible and alternative techniques are needed, the MGFs being infinite for distributions with heavy tails. In [28], an efficient hazard rate twisting IS-based approach was developed for the estimation of α in the case of i.i.d sum of RVs with subexponential decay. In [29] and [30], it has been shown that, in the heavy-tailed setting, the hazard rate twisting approach plays a similar role to the one played by the exponential twisting approach in the light-tailed setting. Of tight relation to this work is the one in [28] which applies the hazard rate twisting method to subexponential distributed random variables. In particular, we propose to extend their approach to the general framework of sums involving independent but non identically distributed arbitrary random variables. To this end, we shall introduce the following objects. We define the hazard rate $\lambda_i(\cdot)$ associated to the RV X_i as:

$$\lambda_i(x) = \frac{f_i(x)}{1 - F_i(x)}, \quad x > 0, \tag{11}$$

where $F_i(\cdot)$ is the CDF of X_i , $i = 1, ..., N$. Besides, we define also the hazard function as:

$$\begin{aligned}\Lambda_i(x) &= \int_0^x \lambda_i(t)dt \\ &= -\log\left(1 - F_i(x)\right), \quad x > 0.\end{aligned} \tag{12}$$

From (11) and (12), the PDF of X_i is related to the hazard rate and function as:

$$\begin{aligned}f_i(x) &= \lambda_i(x)\exp\left(-\int_0^x \lambda_i(t)dt\right) \\ &= \lambda_i(x)\exp\left(-\Lambda_i(x)\right).\end{aligned} \tag{13}$$

The change of probability measure is obtained by twisting the hazard rate of the underlying distribution by a quantity $0 < \theta < 1$ as follows:

$$\begin{aligned}g_i(x) \triangleq f_{i,\theta}(x) &= \left(1-\theta\right)\lambda_i(x)\exp\left(-\left(1-\theta\right)\Lambda_i(x)\right) \\ &= \left(1-\theta\right)f_i(x)\exp\left(\theta\Lambda_i\left(x\right)\right).\end{aligned} \tag{14}$$

Consequently, the RV $T_{\gamma_{th}}$ has the following expression:

$$T_{\gamma_{th}} = \frac{1}{\left(1-\theta\right)^N}\exp\left(-\theta\sum_{i=1}^N \Lambda_i(X_i)\right)\mathbf{1}_{(S_N>\gamma_{th})}. \tag{15}$$

Note that the expectation under the probability measure p^*, that is $\mathbb{E}_{p^*}\left[\cdot\right]$, will be re-denoted by $\mathbb{E}_\theta\left[\cdot\right]$ in the rest of this work.

C. Main Contributions

A lot of research efforts have been devoted to develop efficient algorithms for the estimation of $\alpha = P\left(S_N > \gamma_{th}\right)$, when the underlying distributions are heavy-tailed. Of valuable interest are for instance the works developed in [25]–[28]. In effect, the work of [25] was the first to propose an estimator with bounded relative error under distributions with regularly varying tails. This method was based on the use of the conditional MC technique and dealt with sums of i.i.d random variables. It was then generalized in [31] to the non i.i.d case but its efficiency was only proven under the setting of Pareto-distributed RVs. In addition to methods based on the artifice of conditional MC, we distinguish the dynamic IS scheme [26] whose efficiency was proven only for regularly varying distributions and that of [27] which lies in the intersection of IS and conditional MC techniques. While based on different approaches, all these works present the common denominator of being specific to particular settings. It is thus not clear whether these methods will keep the same performances when applied to other scenarios which does not fall within their original scope of applicability. This motivates our work. In particular, inspired by [28], we generalize the hazard rate twisting IS technique to the general setting of independent and not necessarily identically distributed. Unlike most of the variance reduction techniques, we do not require any assumption on the RVs. To the best of our knowledge, this is a major finding in the context of variance reduction techniques for the following main reasons:

- As mentioned above, efficiency results of most the existing algorithms have often been derived under the assumption that the underlying RVs are drawn from a particular set of distributions. Our work is among the very few ones establishing the asymptotic optimality for any arbitrary RVs, regardless its sign or the nature of its tail.
- The previously cited works [25]–[27] were shown to exhibit good performances in the i.i.d case. While their generalization to the non i.i.d setting is straightforward, it is not clear whether their efficiency will be kept. The authors of [31] who extend the condition MC algorithm [25] to the non i.i.d case have only been able to prove the efficiency of the corresponding algorithm for the particular case of RV following a Pareto distribution.
- We establish a powerful result, in that the asymptotic optimality of the proposed method holds for any arbitrary RVs. This includes, for instance, the interesting cases of summations S_N containing a mixture of heavy and

light tailed distributions or also those involving light-tailed distributions whose MGFs are not known to possess closed-forms.

- In our work, a minmax approach, based on the sequential resolution of minimization and maximization problems is employed in order to determine the closed-form expression of the twisting parameter. The maximization problem produces an upper-bound for the variance of the hazard rate estimator. This bound, depending on the twisting parameter, is then minimized with respect to it, in order to achieve the most possible lower value of the considered upper-bound. It will be shown that this approach of selection the twisting parameter ensures the asymptotic optimality of the proposed hazard-rate estimation for any arbitrary RVs.

III. Proposed Hazard Rate Twisting

A. Minmax Approach

In this subsection, we present the minmax procedure for the determination of the twisting parameter. The minmax choice of θ is divided into two steps. In the first step, we construct an upper bound of the second moment of $T_{\gamma_{th}}$ which is achieved by solving the following maximization problem (P):

$$\begin{aligned} (P): \max_{X_1,\dots,X_N} \quad & L(X_1,X_2,...,X_N) \\ \text{Subject to } & \sum_{i=1}^{N} X_i \geq \gamma_{th}, \\ & X_i > 0, \quad i=1,...,N, \end{aligned} \tag{16}$$

where the likelihood ratio is given by (4) and (14) as follows

$$L(X_1,X_2,...,X_N) = \frac{1}{(1-\theta)^N} \exp\left(-\theta \sum_{i=1}^{N} \Lambda_i(X_i)\right). \tag{17}$$

Hence, solving the problem (P) is equivalent to solving the following minimization problem (P'):

$$\begin{aligned} (P'): \min_{X_1,\dots,X_N} \quad & \sum_{i=1}^{N} \Lambda_i(X_i) \\ \text{Subject to } & \sum_{i=1}^{N} X_i \geq \gamma_{th}, \\ & X_i > 0, \quad i=1,...,N. \end{aligned} \tag{18}$$

Let us denote the optimal solution of (P) by $X_1^*(\gamma_{th}), X_2^*(\gamma_{th}), ..., X_N^*(\gamma_{th})$. Then, we have

$$\begin{aligned} \mathbb{E}_\theta\left[T^2_{\gamma_{th}}\right] &= \mathbb{E}_\theta\left[L^2\left(X_1,X_2,...,X_N\right)\mathbf{1}_{(S_N>\gamma_{th})}\right] \\ &\leq \frac{1}{(1-\theta)^{2N}} \exp\left(-2\theta \sum_{i=1}^{N} \Lambda_i(X_i^*(\gamma_{th}))\right). \end{aligned} \tag{19}$$

The second step is to minimize (19) to get the optimal twisting parameter θ^*. The resulting minimization problem is simple and leads to:

$$\theta^* = 1 - \frac{N}{\sum_{i=1}^{N} \Lambda_i(X_i^*(\gamma_{th}))}. \tag{20}$$

The value of the twisting given in (20) represents the minmax optimal choice among all values of θ, and for all threshold values. The newly derived closed-form expression (20) will ensure, as we will see in the following subsection, that the asymptotic optimality criterion holds for arbitrary sum of independent and not necessarily identically distributed RVs.

B. Asymptotic Optimality Criterion

This section is devoted to the proof of our main result. In particular, we prove that by using twisting parameter θ^* given in (20), the asymptotic optimality of the corresponding estimator holds for any arbitrary RVs. Our main result is based on a careful investigation of the behaviour of the solution to the minimization problem $(P^{'})$. Prior to stating the main theorem, the following lemma is required:

Lemma 1. Let $X_1^*(\gamma_{th}), X_2^*(\gamma_{th}), ..., X_N^*(\gamma_{th})$ be the solution to the minimization problem (P'). Then,

$$\lim_{\gamma_{th} \to +\infty} A(\gamma_{th}) = +\infty, \tag{21}$$

where $A(\gamma_{th}) = \sum_{i=1}^{N} \Lambda_i(X_i^*(\gamma_{th}))$.

Proof: From the inequality constraint of the minimization problem (P'), we have

$$\cap_{i=1}^{N} \left\{X_i \geq X_i^*(\gamma_{th})\right\} \subset \left\{\sum_{i=1}^{N} X_i \geq \gamma_{th}\right\}. \tag{22}$$

Using the independence of $X_1, X_2, ..., X_N$, we get

$$\prod_{i=1}^{N} P\left(X_i \geq X_i^*(\gamma_{th})\right) \leq \alpha.$$

Hence, upon applying the Logarithm function of both sides, it follows that

$$A\left(\gamma_{th}\right) \geq -\log\left(\alpha\right). \tag{23}$$

Finally, since $\alpha \to 0$ as $\gamma_{th} \to +\infty$, the proof is concluded. ■

The convergence result in Lemma 1 represents the key ingredient that underlies the proof of our main result. With Lemma 1 at hand, we prove the following theorem:

Theorem 1. *For any arbitrary independent sum of RVs, the quantity of interest α is asymptotically optimally estimated using the proposed hazard rate twisting IS-based approach with the minmax optimal parameter θ^* given in (20).*

Proof: By replacing the expression of the minmax optimal twisting parameter (20) into (19), we get

$$\mathbb{E}_{\theta^*}\left[T^2_{\gamma_{th}}\right] \leq \left(\frac{A(\gamma_{th})}{N}\right)^{2N} \exp\left(2N - 2A(\gamma_{th})\right) \tag{24}$$

Taking the logarithm of both sides of the above inequality, we get:

$$\log\left(\mathbb{E}_{\theta^*}\left[T^2_{\gamma_{th}}\right]\right) \leq 2N\left(1 + \log\left(\frac{A(\gamma_{th})}{N}\right)\right) - 2A(\gamma_{th}) \tag{25}$$

Now, combining (23) and (25) and using the fact that the right-hand side of (25) is negative for a sufficiently large γ_{th} (this follows from Lemma 1), we get

$$\frac{\log\left(\mathbb{E}_{\theta^*}\left[T^2_{\gamma_{th}}\right]\right)}{\log\left(\alpha\right)} \geq \frac{2N\left(1+\log\left(\frac{A(\gamma_{th})}{N}\right)\right)-2A(\gamma_{th})}{-A(\gamma_{th})} \tag{26}$$

Finally, resorting again to the result of Lemma 1, we obtain:

$$\lim_{\gamma_{th}\to+\infty}\frac{\log\left(\mathbb{E}_{\theta^*}\left[T^2_{\gamma_{th}}\right]\right)}{\log\left(\alpha\right)} \geq 2. \tag{27}$$

Hence, from (8), the asymptotic optimality (9) holds thereby ending the proof. ■

Remark 1. Since the hazard functions $\Lambda_i(\cdot)$ are increasing functions, the inequality constraint is actually satisfied with equality

$$\sum_{i=1}^{N} X_i^*(\gamma_{th}) = \gamma_{th} \tag{28}$$

C. Case Study

Theorem 1 establishes the asymptotic optimality criterion of the proposed IS estimator which uses θ^* as the twisting parameter. While the asymptotic optimality holds for arbitrary sums of RVs, achieving this criterion requires solving the optimization problem (P'). This step strongly depends on the nature of the underlying distribution and thus has to be studied on a case by case basis. For instance, the case of distributions with convex hazard rate functions including Weibull RVs with shape parameter greater than 1 can be handled using convex optimization algorithms. If the convexity of the hazard-rate functions is not satisfied, one can opt for standard numerical optimization methods which might produce local optimal solutions. In order to avoid such situations, some additional results serving to approach the solutions of problem (P') can be of fundamental practical interest. This is the main objective of this section. In particular, we will consider positive RVs belonging to the same family of distributions (a sum of Weibull RVs with different shape and scale parameters) with hazard rate functions being eventually concave, i.e, satisfying the following condition:

$$\exists \eta_i \text{ such that } \Lambda_i(\cdot) \text{ is concave in } [\eta_i,+\infty), i\in\{1,2,...,N\}. \tag{29}$$

Several commonly used distributions satisfy (29) including the Log-normal RV [32]. Moreover, through a simple computation, we can show that the hazard functions of the Weibull (with shape parameter less than 1) and the Pareto distributions are concave on the whole interval $[0,+\infty)$ and hence (29) is in particular satisfied. A similar result is satisfied by the Gamma RV with shape parameter less than 1 [33]. Note in passing that in this case, problem $(P^{'})$ turns out to be a concave minimization problem. The minimum can be thus analytically characterized as one of the extreme points of the domain of $(P^{'})$. While a similar analytical characterization seems to be out of reach when (29) is strictly satisfied (One of the η_i is strictly positive), the eventually concavity behaviour of $\Lambda_i(\cdot)$ can help find a close point to the optimal solution for large threshold values. This is the objective of the following Lemma:

Lemma 2. Under (29), there exists a fixed index $i_0 \in \{1,2,...,N\}$ such that the minimizers of (P') satisfy for a sufficiently large γ_{th}

$$\gamma_{th} - \sum_{i\neq i_0}\eta_i \leq X^*_{i_0}(\gamma_{th}) \leq \gamma_{th}, \tag{30}$$

$$X_i^*(\gamma_{th}) \leq \eta_i, \text{ for all } i\neq i_0, \tag{31}$$

and hence as $\gamma_{th}\to+\infty$, we have

$$X^*_{i_0} \underset{+\infty}{\sim} \gamma_{th}, \text{ as } \gamma_{th}\to\infty, \tag{32}$$

$$X_i^* = \mathcal{O}(1), \text{ for all } i\neq i_0. \tag{33}$$

Proof: Let us consider $S(N,\gamma_{th})$ the set of all feasible solutions:

$$S(N,\gamma_{th}) = \{X=(X_1,X_2,...,X_N)\in(\mathbb{R}^+)^N, \sum_{i=1}^{N} X_i = \gamma_{th}\}. \tag{34}$$

Through the use of (29), the objective function of (P') is concave on the subset:

$$\tilde{S}(N,\gamma_{th}) = \{X=(X_1,X_2,...,X_N)\in(\mathbb{R}^+)^N, \sum_{i=1}^{N} X_i = \gamma_{th}, X_i\geq\eta_i, \text{ for each } i\in\{1,2,...,N\}\}. \tag{35}$$

Thus, the minimum of the objective function of (P') over $\tilde{S}(N,\gamma_{th})$ is achieved in at least one of its extreme points. More precisely, the extreme points of $\tilde{S}(N,\gamma_{th})$ are $e_1,e_2,...,e_N$ such that $e_i = (\eta_1,\eta_2,...,\eta_{i-1},\gamma_{th}-\sum_{j\neq i}\eta_j,\eta_{i+1},...,\eta_N)$. Therefore the minimum of (P') over $S(N,\gamma_{th})$ is either achieved in one of the extreme point e_i, $i=1,2,...,N$, or on the set

$$\begin{aligned}\bar{S}(N,\gamma_{th}) &= S(N,\gamma_{th})\backslash\tilde{S}(N,\gamma_{th})\\ &= \{X=(X_1,X_2,...,X_N)\in(\mathbb{R}^+)^N, \sum_{i=1}^{N} X_i = \gamma_{th},\\ &\exists i \text{ such that } X_i<\eta_i\}.\end{aligned} \tag{36}$$

In both cases, there exists at least one index $i\in\{1,2...,N\}$ such that $X_i^*(\gamma_{th})\leq\eta_i$. In addition, in order to satisfy the equality constraint $\sum_{i=1}^{N}X_i^*(\gamma_{th}) = \gamma_{th}$ for a sufficiently large γ_{th}, there should exist an index $j\in\{1,2,...,N\}$ such that $X_j^*(\gamma_{th})\geq\eta_j$. In order to prove the result in Lemma 2, we proceed iteratively by dimension reduction. In fact, without loss of generality, we assume that $X_N^*(\gamma_{th})\leq\eta_N$ (through an index permutation). It follows that

$$\min_{S(N,\gamma_{th})}\sum_{i=1}^{N}\Lambda_i(X_i) = \min_{X_N\leq\eta_N}\ \min_{S(N-1,\gamma_{th,N-1})}\sum_{i=1}^{N}\Lambda_i(X_i), \tag{37}$$

where $\gamma_{th,N-1}=\gamma_{th}-X_N$. Hence, we get

$$\min_{S(N,\gamma_{th})}\sum_{i=1}^{N}\Lambda_i(X_i) = \Lambda_N(X_N^*(\gamma_{th})) + \min_{S(N-1,\gamma^*_{th,N-1})}\sum_{i=1}^{N-1}\Lambda_i(X_i), \tag{38}$$

Consequently, we can see that we have reduced the number of optimization variables to be $N-1$, while we have kept the same structure of the minimization problem (P') with $\gamma^*_{th,N-1} = \gamma_{th} - X^*_N(\gamma_{th})$. Hence the previous procedure could be repeated again. In fact, using the same argument as before, there exists another index $i \in \{1, 2, ..., N-1\}$ such that $X^*_i(\gamma_{th}) \leq \eta_i$. Without loss of generality, we assume that $i = N-1$ which leads to

$$\min_{S(N,\gamma_{th})} \sum_{i=1}^{N} \Lambda_i(X_i) = \Lambda_N(X^*_N(\gamma_{th})) + \Lambda_{N-1}(X^*_{N-1}(\gamma_{th})) + \min_{S(N-2,\gamma^*_{th,N-2})} \sum_{i=1}^{N-2} \Lambda_i(X_i), \quad (39)$$

where $\gamma^*_{th,N-2} = \gamma_{th} - X^*_N(\gamma_{th}) - X^*_{N-1}(\gamma_{th})$. After $N-2$ steps, we get

$$\min_{S(N,\gamma_{th})} \sum_{i=1}^{N} \Lambda_i(X_i) = \sum_{i=1}^{N-2} \Lambda_{N+1-i}(X^*_{N+1-i}(\gamma_{th})) + \min_{S(2,\gamma^*_{th,2})} \sum_{i=1}^{2} \Lambda_i(X_i), \quad (40)$$

with $X^*_i(\gamma_{th}) \leq \eta_i$, for $i = 3, 4...., N$, and $\gamma^*_{th,2} = \gamma_{th} - \sum_{i=3}^{N} X^*_i(\gamma_{th})$. Thus, we end up with a two dimensional minimization problem. Again, there should exist an index $i = 2$ (through a possible permutation) such that $X^*_2(\gamma_{th}) \leq \eta_2$. Therefore, using the equality constraint $\sum_{i=1}^{N} X^*_i(\gamma_{th}) = \gamma_{th}$, we get

$$X^*_i(\gamma_{th}) \leq \eta_i, \; i = 2, 3, ..., N, \quad (41)$$

$$\gamma^*_{th,2} - \eta_2 \leq X^*_1(\gamma_{th}) \leq \gamma^*_{th,2}. \quad (42)$$

The previous result follows also from the non-negativity of $X_1, X_2, ..., X_N$. Hence, it follows that

$$\gamma_{th} - \sum_{i=2}^{N} \eta_i \leq X^*_1(\gamma_{th}) \leq \gamma_{th}. \quad (43)$$

Thus, as γ_{th} goes to infinity, and using the fact that η_i, $i = 2, 3, ..., N$ are independent of γ_{th}, we have

$$X^*_1(\gamma_{th}) \underset{+\infty}{\sim} \gamma_{th} \quad (44)$$

$$X^*_i(\gamma_{th}) = \mathcal{O}(1), \; \forall i \in \{2, 3, ..., N\}. \quad (45)$$

■

It is important to note that in the particular i.i.d case, the index i_0 could be any index in $\{1, 2, ..., N\}$, and the minimum is achieved in N different points. A direct consequence of Lemma 2 is presented in the following lemma.

Lemma 3. Under (29), the objective function of (P') has the following asymptotic behaviour

$$\sum_{i=1}^{N} \Lambda_i(X^*_i(\gamma_{th})) \underset{+\infty}{\sim} \Lambda_{i_0}(\gamma_{th}), \text{ as } \gamma_{th} \to +\infty. \quad (46)$$

Proof: Using Lemma 2 and the fact that $\Lambda_{i_0}(\gamma_{th})$ tends to infinity as $\gamma_{th} \to +\infty$, we have

$$\frac{\Lambda_i(X^*_i(\gamma_{th}))}{\Lambda_{i_0}(\gamma_{th})} \to 0 \text{ as } \gamma_{th} \to +\infty, \text{ for all } i \neq i_0. \quad (47)$$

The remaining work is to prove that

$$\frac{\Lambda_{i_0}(X^*_{i_0}(\gamma_{th}))}{\Lambda_{i_0}(\gamma_{th})} \underset{+\infty}{\sim} 1, \text{ as } \gamma_{th} \to +\infty. \quad (48)$$

Using the fact that $\Lambda_{i_0}(\cdot)$ is a concave function in the interval $[\eta_{i_0}, +\infty]$, then its derivative which is the hazard rate $\lambda_{i_0}(\cdot)$ is a decreasing function in $[\eta_{i_0}, +\infty]$. Hence, $\lambda_{i_0}(x)$ is upper bounded by $\lambda_{i_0}(\eta_{i_0})$ for all $x \geq \eta_{i_0}$. Consequently, $\Lambda_{i_0}(\cdot)$ is Lipschitz in the interval $[\eta_{i_0}, +\infty)$, that is for all x and y in the interval $[\eta_{i_0}, +\infty)$, we have

$$\left|\Lambda_{i_0}(x) - \Lambda_{i_0}(y)\right| \leq \lambda_{i_0}(\eta_{i_0})\left|x - y\right| \quad (49)$$

By taking $x = \gamma_{th}$ and $y = X^*_{i_0}(\gamma_{th})$, it follows that

$$\Lambda_{i_0}(\gamma_{th}) - \Lambda_{i_0}(X^*_{i_0}(\gamma_{th})) = \mathcal{O}(\gamma_{th} - X^*_{i_0}(\gamma_{th})), \text{ as } \gamma_{th} \to +\infty. \quad (50)$$

Using Lemma 2, we have that $\gamma_{th} - X^*_{i_0}(\gamma_{th}) = \mathcal{O}(1)$. Thus, it follows that

$$\Lambda_{i_0}(\gamma_{th}) - \Lambda_{i_0}(X^*_{i_0}(\gamma_{th})) = o(\Lambda_{i_0}(\gamma_{th})), \quad (51)$$

which leads to (48) and then the proof is concluded. ■

Remark 2. Distributions satisfying (29) were considered in [28] for the particular i.i.d case. In this particular i.i.d setting and from the result of Lemma 3, we can observe that the minmax parameter θ^* in (20) tends to the same value of θ derived in [28], as γ_{th} increases.

Remark 3. To fully characterize the solution of (P') under (29), we need to specify how to determine the index i_0 appearing in Lemma 2 and Lemma 3. In fact, this index satisfies the following

$$\Lambda_{i_0}(\gamma_{th}) \leq \Lambda_i(\gamma_{th}), \forall i \neq i_0. \quad (52)$$

For instance, for the sum of Log-normal RVs with mean μ_i and standard deviation σ_i, $i = 1, 2, ..., N$, the index i_0 satisfies

$$\left(\log(\gamma_{th}) - \mu_{i_0}\right)/\sigma_{i_0} \leq \left(\log(\gamma_{th}) - \mu_i\right)/\sigma_i, \forall i \neq i_0. \quad (53)$$

Thus, for γ_{th} large enough, the index i_0 is independent of γ_{th} and corresponds to

$$i_0 = \arg\max_{i \in \{1,2,...,N\}} \sigma_i. \quad (54)$$

Moreover, if there exists another index with a maximum standard deviation, i_0 corresponds to the RV with a maximum mean.

Remark 4. The results of Lemma 2 and (52) can help provide an initial guess of the solution to problem (P'). This guess can be fed to numerical optimization methods used to solve (P') thereby ensuring their convergence to close-to-optimal solutions.

Distributions with Concavity Property: As we mentioned earlier, for distributions with concave hazard rate functions, an analytic characterization of the optimum solution to (P') can be obtained. For sake of illustration, we treat in particular, the case of Weibull distribution with shape parameter less than 1. The PDF of X_i, $i = 1, 2, ..., N$ is:

$$f_i(x) = \frac{k_i}{\beta_i}\left(\frac{x}{\beta_i}\right)^{k_i - 1} \exp\left(-\left(\frac{x}{\beta_i}\right)^{k_i}\right), \;\; x > 0. \quad (55)$$

where $0 < k_i < 1$ and $\beta_i > 0$, $i = 1, 2, ..., N$, denote respectively the shape and the scale parameters. The hazard rate and the hazard function for each X_i, $i = 1, 2, ..., N$, are as follows

$$\lambda_i(x) = \frac{k_i}{\beta_i}\left(\frac{x}{\beta_i}\right)^{k_i - 1}, \ \ x \geq 0. \tag{56}$$

$$\Lambda_i(x) = \left(\frac{x}{\beta_i}\right)^{k_i}, \ \ x \geq 0.. \tag{57}$$

We can prove through a simple computation that the objective function of (P') is concave and hence (29) is satisfied. In fact, the Hessian $H(X_1, X_2, ..., X_N)$, which is the squared matrix composed of second-order partial derivative of the objective function $\sum_{i=1}^{N} \Lambda_i(X_i)$ at any point $X = (X_1, X_2, ..., X_N) \in (\mathbb{R}^+)^N$, is a diagonal matrix with diagonal elements

$$[H(X_1, X_2, ..., X_N)]_{ii} = \frac{k_i(k_i - 1)}{\beta_i^2}\left(\frac{X_i}{\beta_i}\right)^{k_i - 2}, \tag{58}$$

which are strictly negative for $k_i < 1$, $i = 1, 2, ..., N$. In particular, the objective function is also concave on the convex set $S(N, \gamma_{th}) = \{X = (X_1, X_2, ..., X_N) \in (\mathbb{R}^+)^N$, such that $\sum_{i=1}^{N} X_i = \gamma_{th}\}$. Therefore, the solution of (P') is obtained in one of the extreme points of $S(N, \gamma_{th})$. In other words, the minimum is achieved when

$$X_{i_0}^*(\gamma_{th}) = \gamma_{th}, \text{ and } X_i^*(\gamma_{th}) = 0 \ \forall i \neq i_0, \tag{59}$$

where i_0 satisfying

$$\left(\frac{\gamma_{th}}{\beta_{i_0}}\right)^{k_{i_0}} \leq \left(\frac{\gamma_{th}}{\beta_i}\right)^{k_i}, \ \forall i \neq i_0. \tag{60}$$

It is worth mentioning that for large values of γ_{th}, the index i_0 depends only on the shape and scale parameters and independent of γ_{th}. More precisely, for γ_{th} large enough, it is characterized by

$$i_0 = \arg\min_i k_i. \tag{61}$$

Moreover, if there are more than one RV with minimum shape parameter, the index i_0 corresponds to the one with maximum scale parameter. Note that (59) holds for any distribution with concavity property. For instance, an equivalent result can be obtained for the Gamma distribution with shape parameter less than 1 and for the Pareto distribution.

D. Algorithm

A pseudo-code describing all steps to estimate α by our proposed hazard rate twisting approach is described in Algorithm 1.

Algorithm 1 Hazard rate twisting approach for the estimation of α

Inputs: M, γ_{th}.
Outputs: $\hat{\alpha}_{IS}$.
Find the minmax value θ^* as in (20) by solving the minimization problem (P').
for $i = 1, ..., M$ **do**
  Generate independent realizations of $\{X_j(\omega_i)\}_{j=1}^N$ under the twisted PDF $\{f_{j,\theta^*}(\cdot)\}_{j=1}^N$
  Evaluate $T_{\gamma_{th}}(\omega_i)$ as in (15).
end for
Compute the IS estimator as $\hat{\alpha}_{IS} = \frac{1}{M}\sum_{i=1}^{M} T_{\gamma_{th}}(\omega_i)$.

In the implementation of Algorithm 1, we need to generate samples of $X_1, X_2, ..., X_N$ according to the twisted PDFs $f_{1,\theta^*}, f_{2,\theta^*}, ..., f_{N,\theta^*}$, respectively. To this end, several methods can be used, among them, we distinguish the acceptance rejection technique, the Markov Chain Monte Carlo algorithm [34], or the inverse CDF sampling method [35]. The inverse CDF sampling method is merely based on the observation that, for a given a CDF F, the random variable $F^{-1}(U)$ where U is uniformly distributed RV over $[0, 1]$ has a CDF given by F. For this method to be applicable, an analytical expression for the CDF inverse is required. In the sequel, we show that the inverse CDF of the twisted RVs is related to that of the non-twisted RVs. If the inverse of the CDF of the non-twisted RV $F(\cdot)^{-1}$ admits an analytical expression, so does $F_\theta(\cdot)^{-1}$. To see that, let us consider a RV X with an underlying PDF $f(\cdot)$ and CDF $F(.)$. From (14), the PDF $f_\theta(\cdot)$ associated to X with hazard rate $\lambda(\cdot)$ and hazard function $\Lambda(\cdot)$ is

$$\begin{aligned} f_\theta(x) &= (1 - \theta)\lambda(x)\exp(-(1 - \theta)\Lambda(x)) \\ &= (1 - \theta)f(x)\exp(\theta\Lambda(x)). \end{aligned} \tag{62}$$

Replacing $\lambda(\cdot)$ and $\Lambda(\cdot)$ by their definitions, we get

$$f_\theta(x) = \frac{(1 - \theta)f(x)}{(1 - F(x))^\theta}. \tag{63}$$

By a simple integration, the corresponding CDF is given by

$$F_\theta(x) = -\frac{1}{(1 - F(x))^{\theta - 1}} + 1. \tag{64}$$

Finally, a simple computation leads to an exact expression of the CDF inverse of the RV X under the hazard rate twisting technique

$$F_\theta^{-1}(y) = F^{-1}(1 - (1 - y)^{-\frac{1}{\theta - 1}}), \tag{65}$$

where $F^{-1}(\cdot)$ is the CDF inverse of X under the original PDF $f(\cdot)$. It is worth observing that many of the most frequently encountered distributions has an inverse CDF that possesses an analytical expression. A non-comprehensive list includes the Log-normal and the Weibull distribution, often used for modeling random wireless channels. This argues in favor of the efficiency of the inverse CDF method to handle many practical situations.

Remark 5. We have described in the previous section a method based on the inverse CDF sampling method $F_\theta^{-1}(\cdot)$ to generate samples of a RV X under the twisted PDF $f_\theta(\cdot)$.

For the particular Weibull distribution with parameters k and β, the PDF $f_\theta(\cdot)$ remains a Weibull distribution but with the same shape parameter k and a different scale parameter β' as follows

$$\begin{aligned} f_\theta(x) &= (1-\theta)\lambda(x)\exp\left(-(1-\theta)\Lambda(x)\right) \\ &= (1-\theta)\frac{k}{\beta}\left(\frac{x}{\beta}\right)^{k-1}\exp\left(-(1-\theta)(\frac{x}{\beta})^k\right) \\ &= \frac{k}{\beta'}\left(\frac{x}{\beta'}\right)^{k-1}\exp\left(-(\frac{x}{\beta'})^k\right). \end{aligned} \quad (66)$$

where $\beta' = \frac{\beta}{(1-\theta)^{1/k}}$.

IV. Simulation Results

This section presents some selected simulations results in order to illustrate the performance of the proposed IS scheme. First, we analyze the efficiency of the hazard rate twisting change of measure in increasing the number of occurrences of the considered rate event. Second, we compare the proposed IS scheme to an algorithm based on Conditional MC (CMC) [31] and we show that, in some settings, the proposed method achieves a computational gain over the CMC one. Finally, we analyze in the third subsection the near-optimality of the minmax twisting parameter (20) compared to the unknown optimal twisting parameter (the one that minimizes the actual variance of $T_{\gamma_{th}}$).

A. *Frequency of Occurrence*

As it was mentioned before, a key ingredient of IS techniques is to emphasize the sampling of important values, i.e realizations satisfying $\{S_N \geq \gamma_{th}\}$. We define the frequency of occurrence as the number of samples which satisfy $\{S_N \geq \gamma_{th}\}$. In our first simulation results, we consider the sum of two i.i.d. Log-normal RVs with mean $\mu_{dB} = 0\ dB$ and standard deviation $\sigma_{dB} = 6\ dB$. In Table I, we have computed the frequency of occurrence using the naive MC simulation and the proposed IS technique. Table I exhibits an important feature of the IS change of measure where the frequency of realizations belonging to the rare set $S_N \geq \gamma_{th}$ is almost constant as we increase the threshold. On the other hand, the failure of sampling under the original sum of Log-normal distribution is obvious, being inefficient in acquiring a sufficient number of realizations in the rare sets.

TABLE I
FREQUENCY OF OCCURRENCE FOR THE SUM OF TWO I.I.D. LOG-NORMAL WITH $\mu_{dB} = 0\ dB$, $\sigma_{dB} = 6\ dB$, AND $= M = 10^5$.

Threshold (dB)	$\hat{\alpha}_{IS}$	IS frequency	MC frequency
15	1.47×10^{-2}	28603	1427
20	9.55×10^{-4}	27631	99
25	3.17×10^{-5}	26484	3
30	5.8×10^{-7}	26253	0
35	0.55×10^{-8}	25982	0

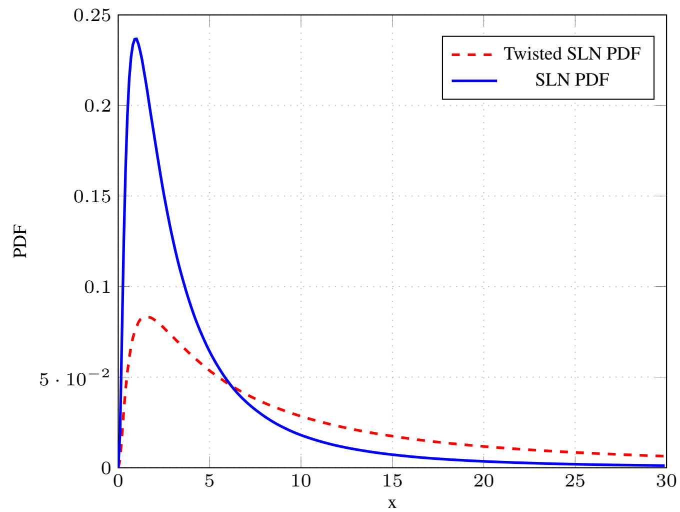

Fig. 1. Twisted and original PDFs of the sum of two i.i.d Log-normal RVs with $\gamma_{th} = 20$, $\mu_{dB} = 0\ dB$, and $\sigma_{dB} = 6\ dB$.

TABLE II
FREQUENCY OF OCCURRENCE FOR THE SUM OF TWO I.I.D. WEIBULL DISTRIBUTION WITH $k = 0.5$, $\beta = 1$, AND $M = 10^5$.

Threshold (dB)	$\hat{\alpha}_{IS}$	IS frequency	MC frequency
10	1.01×10^{-1}	29273	10097
15	1.67×10^{-2}	29270	852
20	1.06×10^{-4}	29244	6
25	4.15×10^{-8}	29143	0
30	3.88×10^{-14}	29049	0

In Table II, we show the same computation using the sum of two i.i.d Weibull distribution with shape parameter $k = 0.5$ and scale parameter $\beta = 1$. Again, important values are sampled more frequently using the IS technique and their frequencies remains almost constant as we increase the threshold. To illustrate this statement, we plotted in Fig. 1 the twisted against the original sum of Log-normal distributions for a fixed threshold $\gamma_{th} = 20$. Clearly, we see that twisting the hazard rate of each component in the summation leads to a heavier twisted PDF. As a consequence, the event $\{S_N > \gamma_{th}\}$ are more likely to occur under the twisted PDF than under the original one.

B. *Efficiency of the Proposed IS Algorithm*

In this subsection, we compare our proposed IS scheme to the algorithm based on CMC [31] which is an extension of [25] to the non i.i.d case. The proposed estimator in [31] writes as:

$$\hat{\alpha}_{CMC} = \frac{1}{M}\sum_{k=1}^{M} T'_{\gamma_{th}}(\omega_k) \quad (67)$$

where $T'_{\gamma_{th}} = \sum_{i=1}^{N} \bar{F}_i\left(\max(\gamma_{th} - \sum_{j\neq i} X_j, M_{-i})\right)$, $\bar{F}_i = 1 - F_i$, and $M_{-i} = \max(X_1, ..., X_{i-1}, X_{i+1}, ..., X_N)$, $i = 1, 2, ..., N$. Our choice is mainly motivated by the good performances of CMC based algorithms. In particular, when all RV are i.i.d, the algorithm of [25] is known to achieve a bounded relative error in the case where RVs are drawn from regularly varying distributions and to satisfy the asymptotic optimality criterion when RVs are i.i.d and follow a Weibull-like distributions with shape parameter k less than $\log(3/2)/\log(2)$. As for the non i.i.d case, it was proven numerically in [31] that for the Weibull distribution the CMC algorithm performs much better when the shape parameters are small than when they are large. In our opinion, the CMC might exhibit better performances than the proposed hazard rate twisting approach for small values of the shape parameter of the Weibull distribution. But, in the general case, there is no guarantee that this always occurs. In particular, scenarios which do not fall within the conventional scope of applicability of the CMC algorithm, constitute potential situations in which our method might achieve better performances. In light of this observation, we identified three different settings where the proposed algorithm outperforms the CMC algorithm.

The comparison is performed based on the variance reduction metric which we define for the proposed IS scheme as:

$$\xi_1 = \frac{\alpha(1-\alpha)}{\mathrm{var}_\theta\left[T_{\gamma_{th}}\right]} \tag{68}$$

This metric corresponds to the efficiency of the proposed IS scheme compared to the naive MC simulation. Similarly, the efficiency of the CMC approach with respect to the naive MC simulation is defined as

$$\xi_2 = \frac{\alpha(1-\alpha)}{\mathrm{var}\left[T'_{\gamma_{th}}\right]} \tag{69}$$

We compare the proposed IS scheme to the CMC algorithm under three different settings depending on the nature of the tail of the underlying RVs. We consider in the first case the sum of $N = 10$ components drawn from the Weibull distribution with shape parameter being given by either 0.8 or 0.9, a setting which corresponds to the sum of heavy-tailed distributions. Table III provides the performance results for the CMC method and the proposed IS scheme where the minmax twisting parameter (20) is used. We deduce from this table that both techniques offer good performances compared to the naive MC simulation. Moreover, the proposed IS technique achieves better performances than that of the CMC. The gain in efficiency becomes even higher as the threshold increases. For instance, for $\gamma_{th} = 55$, the amount of variance reduction achieved by our proposed IS algorithm is approximately 26.8 times the amount of variance reduction given by the CMC method.

TABLE III
SUM OF $N = 10$ INDEPENDENT WEIBULL DISTRIBUTION WITH $\beta_i = 0.5 + i/10$, $k_i = 0.8$, $i = 1, 2, ..., 5$, $k_i = 0.9$, $i = 6, 7, ..., 10$, AND $M = 10^7$. THE EFFICIENCY IS COMPUTED WITH RESPECT TO THE NAIVE MC SIMULATION.

	Importance Sampling		Conditional Monte Carlo	
γ_{th}	$\hat{\alpha}_{IS}$	ξ_1	$\hat{\alpha}_{CMC}$	ξ_2
35	1.34×10^{-4}	200.30	1.34×10^{-4}	113.56
40	1.74×10^{-5}	1.05×10^3	1.74×10^{-5}	282.01
45	2.18×10^{-6}	5.42×10^3	2.18×10^{-6}	694.27
50	2.76×10^{-7}	2.44×10^4	2.76×10^{-7}	1.49×10^3
55	3.44×10^{-8}	1.08×10^5	3.40×10^{-8}	4.03×10^3

In a second experiment, we consider the case where the underlying sum involves a mixture of heavy and light tailed RVs. Table IV presents the obtained result when the sum of $N = 10$ independent Weibull distributions with shape parameter being selected from $\{0.8, 1\}$.

TABLE IV
SUM OF $N = 10$ INDEPENDENT WEIBULL DISTRIBUTION WITH $\beta_i = 0.5 + i/10$, $k_i = 0.8$, $i = 1, 2$, $k_i = 1$, $i = 3, 7, ..., 10$, AND $M = 10^7$. THE EFFICIENCY IS COMPUTED WITH RESPECT TO THE NAIVE MC SIMULATION.

	Importance Sampling		Conditional Monte Carlo	
γ_{th}	$\hat{\alpha}_{IS}$	ξ_1	$\hat{\alpha}_{CMC}$	ξ_2
30	8.26×10^{-5}	565.75	8.22×10^{-5}	99.52
35	4.88×10^{-6}	5.67×10^3	4.91×10^{-6}	292.22
40	2.64×10^{-7}	6.01×10^4	2.65×10^{-7}	956.52
45	1.36×10^{-8}	6.21×10^5	1.40×10^{-8}	1.99×10^3

From this table, it becomes clear that the proposed IS approach can achieve better performances than that of the CMC algorithm. The gain in performance is higher than the one shown by Table III. For example, when $\gamma_{th} = 45$, the efficiency obtained from the IS technique is approximately 312 times the one given by the CMC technique. This result is quite expected, the efficiency of the CMC algorithm being shown in [31] for small shape parameters.

Finally, we compare the performance of the proposed scheme where the sum includes only light-tailed random variables. While we are aware that the exponential twisting approach is considered as more appropriate to handle light-tailed settings, its use to the present context is not possible since it requires the MGF to admit a closed form expression, a condition which is not satisfied for Weibull distributed RVs.

TABLE V
SUM OF $N = 10$ INDEPENDENT WEIBULL DISTRIBUTION WITH $\beta_i = 0.5 + i/10$, $k_i = 2$, $i = 1, 2, ..., 10$, AND $M = 10^7$. THE EFFICIENCY IS COMPUTED WITH RESPECT TO THE NAIVE MC SIMULATION.

	Importance Sampling		Conditional Monte Carlo	
γ_{th}	$\hat{\alpha}_{IS}$	ξ_1	$\hat{\alpha}_{CMC}$	ξ_2
15	5.65×10^{-4}	92.47	5.64×10^{-4}	17.12
16	8.03×10^{-5}	429.41	8.05×10^{-5}	28.68
17	9.17×10^{-6}	2.47×10^3	9.18×10^{-6}	53.01
18	8.55×10^{-7}	1.76×10^4	8.74×10^{-7}	78.44
19	6.42×10^{-8}	1.55×10^5	6.94×10^{-8}	158.52

Table V represents the obtained result in the case where a sum of $N = 10$ light-tailed independent Rayleigh RVs is used (The Rayleigh RV is actually a Weibull distribution with shape parameter equal to 2). Again in this setting, as shown in Table V, the gain of our method over the CMC method is evidently clear.

C. Sensitivity Analysis of the Minmax Approach

Obviously, the optimal choice of the twisting parameter θ corresponds to the one minimizing the variance of $T_{\gamma_{th}}$ or equivalently the quantity $\mathbb{E}_\theta\left[T^2_{\gamma_{th}}\right]$. This optimal value is in general unknown. For this reason, we propose to select the twisting parameter θ^* that minimizes an upper bound of $\mathbb{E}_\theta\left[T^2_{\gamma_{th}}\right]$. While we have shown that working with θ^* guarantees the asymptotic optimality, it is not clear how the performance of the proposed technique compares with the optimal approach consisting in twisting the hazard rates by the quantity $\hat{\theta}$. We aim in this subsection to analyze the closeness of $\mathbb{E}_{\theta^*}\left[T^2_{\gamma_{th}}\right]$ to $\mathbb{E}_{\hat{\theta}}\left[T^2_{\gamma_{th}}\right]$ and to investigate whether our minmax choice is efficient, in the sense that it does not worsen the minimum variance considerably. To this end, we plot, in

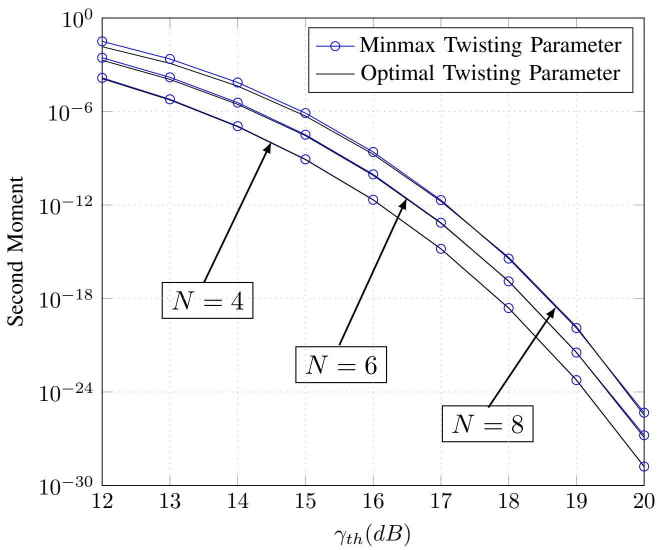

Fig. 2. Second moment of $T_{\gamma_{th}}$ with the minmax and the optimal twisting parameter for the sum of N Weibull RVs with shape parameters $k_i = 0.8$, scale parameters $\beta_i = 1$, $i = 1, 2, ..., N$, and $M = 10^7$.

Fig. 2, $\mathbb{E}_{\theta^*}\left[T^2_{\gamma_{th}}\right]$ and $\mathbb{E}_{\hat{\theta}}\left[T^2_{\gamma_{th}}\right]$ with respect to the threshold γ_{th} in the cases where a sum of $N = 4, 6$, and 8 Weibull distributed RVs are considered. The results in this figure clearly show that both values $\hat{\theta}$ and θ^* achieve almost the same variance reduction. This argues in favor of the efficiency of the minmax parameter θ^* in retrieving approximately the same performances obtained by using the twisting parameter $\hat{\theta}$.

V. Conclusion

This paper provided additional results on the hazard rate twisting IS-based approach. By a proper selection of the twisting parameter, we proved that the asymptotic optimality criterion holds for sums involving independent and not necessarily identically distributed arbitrary RVs. This finding enlarges the framework of hazard rate twisting techniques for the general case of sums involving arbitrary independent RVs. Simulations results comparing our method to an algorithm based on Conditional Monte Carlo was presented and illustrated the efficiency of our technique in handling a large range of scenarios.

References

[1] M.K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels, 2nd Edition*. New York: Wiley, 2004.

[2] G. L. Stüber, *Principles of Mobile Communication, 2nd Edition.* Norwell, MA, USA: Kluwer Academic Publishers, 2001.

[3] M. Ghavami, R. Kohno, and L. Michael, *Ultra wideband signals and systems in communication engineering*. Chichester: Wiley, 2004.

[4] S. M. Navidpour, M. Uysal, and M. Kavehrad, "BER performance of free-space optical transmission with spatial diversity." *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2813–2819, Aug. 2007.

[5] N. Sagias and G. Karagiannidis, "Gaussian class multivariate weibull distributions: Theory and applications in fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3608–3619, Oct. 2005.

[6] F. Babich and G. Lombardi, "Statistical analysis and characterization of the indoor propagation channel," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 455–464, Mar. 2000.

[7] A. Healey, C. Bianchi, and K. Sivaprasad, "Wideband outdoor channel sounding at 2.4 GHz," in *IEEE Conference on Antennas and Propagation for Wireless Communications*, Waltham, MA, US, Nov. 2000.

[8] L. Fenton, "The sum of Log-normal probability distributions in scatter transmission systems." *IRE Transactions on Communications Systems*, vol. 8, no. 1, pp. 57–67, 1960.

[9] S. C. Schwartz and Y. S. Yeh, "On the distribution function and moments of power sums with Lognormal component." *The Bell Systems Technical Journal*, 1982.

[10] N. Beaulieu and Q. Xie, "An optimal Lognormal approximation to Lognormal sum distributions," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 2, pp. 479–489, Mar. 2004.

[11] N. Beaulieu and F. Rajwani, "Highly accurate simple closed-form approximations to Lognormal sum distributions and densities," *IEEE Communications Letters*, vol. 8, no. 12, pp. 709–711, Dec. 2004.

[12] J. Filho and M. Yacoub, "Simple precise approximations to Weibull sums," *IEEE Communications Letters*, vol. 10, no. 8, pp. 614–616, Aug. 2006.

[13] J. Hu and N. Beaulieu, "Accurate simple closed-form approximations to Rayleigh sum distributions and densities," *IEEE Communications Letters*, vol. 9, no. 2, pp. 109–111, Feb. 2005.

[14] F. Yilmaz and M.-S. Alouini, "Sum of Weibull variates and performance of diversity systems," in *International Wireless Communications and Mobile Computing Conference (IWCMC'2009)*, Leipzig, Germany, Jun. 2009.

[15] J. A. Bucklew, *Introduction to Rare Event Simulation*, ser. Springer series in statistics. New York: Springer, 2004.

[16] D. Lu and K. Yao, "Improved importance sampling technique for efficient simulation of digital communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 1, pp. 67–75, Jan. 1988.

[17] C. Jeruchim, P. Hahn, K. Smyntek, and R. Ray, "An experimental investigation of conventional and efficient importance sampling," *IEEE Transactions on Communications*, vol. 37, no. 6, pp. 578–587, Jun. 1989.

[18] N. Beaulieu, "A composite importance sampling technique for digital communication system simulation," *IEEE Transactions on Communications*, vol. 38, no. 4, pp. 393–396, Apr. 1990.

[19] H.-J. Schlebusch, "On the asymptotic efficiency of importance sampling techniques," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 710–715, Mar. 1993.

[20] J. Sadowsky and J. Bucklew, "On large deviations theory and asymptotically efficient Monte Carlo estimation," *IEEE Transactions on Information Theory*, vol. 36, no. 3, pp. 579–588, May. 1990.

[21] J. Sadowsky, "On the optimality and stability of exponential twisting in Monte Carlo estimation," *IEEE Transactions on Information Theory*, vol. 39, no. 1, pp. 119–128, Jan. 1993.

[22] K. Ben Letaief, "Performance analysis of digital lightwave systems using efficient computer simulation techniques," *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 240–251, Feb. 1995.

[23] S. Asmussen, K. Binswanger, and B. Hojgaard, "Rare events simulation for heavy-tailed distributions," *Bernoulli*, vol. 6, no. 2, pp. 303–322, Apr. 2000.

[24] D. Kroese and R. Rubinstein, "The transform likelihood ratio method for rare event simulation with heavy tails," *Queueing Systems*, vol. 46, no. 3-4, pp. 317–351, Mar. 2004.

[25] S. Asmussen and D. P. Kroese., "Improved algorithms for rare event simulation with heavy tails," *Advances in Applied Probability*, pp. 545–558, 2006.

[26] P. Dupuis, K. Leder, and H. Wang, "Importance sampling for sums of random variables with regularly varying tails," *ACM Trans. Model. Comput. Simul.*, vol. 17, no. 3, Jul. 2007. [Online]. Available: http://doi.acm.org/10.1145/1243991.1243995

[27] S. Juneja, "Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error," *Queueing Systems*, vol. 57, no. 2-3, pp. 115–127, 2007. [Online]. Available: http://dx.doi.org/10.1007/s11134-007-9051-8

[28] S. Juneja and P. Shahabuddin, “Simulating heavy tailed processes using delayed hazard rate twisting,” *ACM Trans. Model. Comput. Simul.*, vol. 12, no. 2, pp. 94–118, Apr. 2002.

[29] Z. Huang and P. Shahabuddin, “A unified approach for finite-dimensional, rare-event Monte Carlo simulation,” in *IEEE Winter Simulation Conference (WSC’2004)*, vol. 2, Washington DC, US, Dec. 2004, pp. 1616–1624 vol.2.

[30] ——, “Rare-event, heavy-tailed simulations using hazard function transformations, with applications to value-at-risk,” in *IEEE Winter Simulation Conference (WSC’2003)*, New Orelans, LA, US, Dec. 2003, pp. 276–284.

[31] J. Chan and D. Kroese, “Rare-event probability estimation with conditional monte carlo,” *Annals of Operations Research*, vol. 189, no. 1, pp. 43–61, 2011. [Online]. Available: http://dx.doi.org/10.1007/s10479-009-0539-y

[32] P. Jelenkovic and P. Momcilovic, “Resource sharing with subexponential distributions,” in *Proc. of IEEE 21st Annual Joint Conference of the Computer and Communications (INFOCOM’ 2002)*, vol. 3, New York, USA, June 2002, pp. 1316–1325 vol.3.

[33] I. O. Albert W. Marshall, *Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Familie*. Springer Science & Business Media, 2007.

[34] D. P. Kroese, T. Taimre, and Z. I. Botev, *Handbook of Monte Carlo methods*. Hoboken, N.J. Wiley, 2011.

[35] L. Devroye, *Non-Uniform Random Variate Generation*. New York: Springer-Verlag, 1986.