my journal manuscript No.
(will be inserted by the editor)

On the Generalization of the Hazard Rate Twisting-Based Simulation Approach

Nadhir Ben Rached · Fatma Benkhelifa · Abla Kammoun · Mohamed-Slim Alouini · Raul Tempone.

Received: date / Accepted: date

Abstract Estimating the probability that a sum of random variables (RVs) exceeds a given threshold is a well-known challenging problem. A naive 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. An alternative approach 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 generalization of the well-known hazard rate twisting Importance Sampling based approach that presents the advantage of being logarithmic efficient for arbitrary sums of 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 the proposed method with some existing techniques.

Keywords Naive Monte Carlo · rare events · Importance Sampling · hazard rate twisting · logarithmic efficient · twisting parameter

---

Part of this work has been accepted at the IEEE International Conference on Communications (ICC'2015), London, UK, Jun. 2015, (Ben Rached et al. 2015a)

The authors are in Computer, Electrical and Mathematical Science and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia, and also members of the KAUST Strategic Research Initiative on Uncertainty Quantification in Science and Engineering (SRI-UQ) (e-mail: nadhir.benrached, fatma.benkhelifa, abla.kammoun, slim.alouini, raul.tempone@kaust.edu.sa).

1 Introduction

The problem of estimating the probability that a sum of random variables (RVs) exceeds a certain threshold is often encountered in various fields such as in the performance analysis of wireless communication systems (Simon and Alouini 2004), in queuing systems and insurance risk (Asmussen and Kroese 2006). For instance, a sum of RVs might represent the total co-channel interference power from all the transmissions in neighboring cells (Stüber 2001). Evaluating the probability that this sum exceeds a sufficiently large threshold is a question of major interest and can help in predicting the occurrence of outage events. A second application of our problem can be motivated by the field of insurance risk and concerns the case of an insurance company aiming to evaluate the probability that the total number of claims, modeled by the sums of independent RVs, exceed a large threshold (Asmussen and Glynn 2007). This question is very important as measuring this probability tells about the risk of undergoing a large loss. Unfortunately, closed-form expressions of most of the challenging sum distributions 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 (Stüber 2001; Ghavami et al. 2004; Navidpour et al. 2007; Sagias and Karagiannidis 2005; Babich and Lombardi 2000; Healey et al. 2000). In order to tackle this issue, several analytical approaches, which consists in determining accurate closed-form approximations, approaching the distribution of the sum of these RVs were proposed (Fenton 1960; Schwartz and Yeh 1982; Beaulieu and Xie 2004; Beaulieu and Rajwani 2004; Filho and Yacoub 2006; Hu and Beaulieu 2005; Yilmaz and Alouini 2009). However, these analytical approaches present the inconvenience 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 (MC) methods.

The naive 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, variance reduction techniques constitute an alternative approach that helps improve the computational efficiency of the naive MC simulation (Bucklew 2004). Many research efforts have been carried out to propose efficient variance reduction algorithms to efficiently estimate the probability of interest. The exponential twisting technique, derived from the large deviation theory, is among the well-known Importance Sampling (IS) approach that was shown to exhibit a good efficiency for problems involving a sum of light-tailed distributions (Sadowsky and Bucklew 1990; Sadowsky 1993). For instance, it was applied to estimate the bit error rate of direct-detection optical systems employing avalanche photodiode receivers in Ben Letaief (1995).

However, the scope of applicability 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. A lot of research efforts have been devoted to develop efficient algorithms, when the underlying distributions are heavy-tailed. Of valuable interest are for instance the works developed in Asmussen and Kroese (2006); Dupuis et al. (2007); Juneja (2007); Juneja and Shahabuddin (2002); Blanchet and Liu (2008); Kroese and Rubinstein (2004); Rubinstein and Kroese (2004); Asmussen and Kortschak (2015). In effect, the work of Asmussen and Kroese (2006) 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 independent and identically distributed (i.i.d) RVs. The authors in Hartinger and Kortschak (2009) have then extended the result that the estimator of Asmussen and Kroese (2006) has bounded relative error (or even a stronger criterion, namely the asymptotically vanishing relative error property) to a boarder class of sums distributions such as the sum of standard Log-normal and the sum of Weibull (with an assumption on the shape parameter) RVs. The work of Asmussen and Kroese (2006) was also generalized in Chan and Kroese (2011) to the independent and not identically distributed case but its efficiency was only proven under the setting of Pareto-distributed RVs. An alternative extension of the work of Asmussen and Kroese (2006) to the independent and not identically distributed setting was developed in Nandayapa (2008) where the proposed estimator was shown to possess the bounded relative error property for Log-normal and regularly varying distributions. In addition to methods based on the artifice of conditional MC, we distinguish the dynamic IS scheme of Dupuis et al. (2007) whose efficiency was proven only for regularly varying distributions and that of Juneja (2007) 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 constitutes the major motivation behind our work. In fact, we propose in this paper to generalize the hazard rate twisting IS-based approach of Juneja and Shahabuddin (2002), which was originally developed to estimate the probability that a sum of i.i.d RVs with subexponential decay exceeds a certain threshold, to the problem of estimating the probability that a sum of independent but not necessarily identically distributed arbitrary RVs exceeds a given threshold. 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. This lies behind the main motivation of our work where we consider the problem of arbitrary sums of RVs regardless their signs or the nature of their tails.

– We establish a powerful result, in that the logarithmic efficiency of the proposed method holds for any arbitrary RVs. This includes, for instance, the interesting cases of summations 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.
– The wide scope of applicability of the proposed approach lies principally to our particular way of selecting the twisting parameter. In fact, we propose a minmax approach yielding a closed-form expression of this parameter that guarantee the logarithmic efficiency for problems involving arbitrary sums of variates. Moreover, a detailed study of the minmax formula is also conducted for sums of positive distributions having eventually concave hazard functions. A non exhaustive list of these distributions includes for instance the Log-normal and the Weibull (with shape parameter less than 1) variates.

Note that a similar approach has been developed in Ben Rached et al. (2015b) for a specific class of distributions. However, this does not affect the contribution made in the present work as we tackle the most general framework involving a sum of arbitrary RVs without any restriction on either their signs or their tails.

The rest of the paper is organized as follows. In section 2, we state the problem setting. In Section 3, the generalization of the hazard rate twisting technique is presented and the main result proving the logarithmic efficiency criterion is stated in Theorem 1. In the same section, a case study is analysed with details. Finally, some selected simulation results are shown in Section 4 to assess the performance of the proposed IS scheme.

2 Problem Setting

Let $X_1, X_2, ..., X_N$ be a sequence of independent but not necessarily identically distributed continuous 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:

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

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 $X_1(\omega_j), ..., X_N(\omega_j)$ 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 expensive for the estimation of rare events. In fact, from the Central Limit Theorem, it can be shown that the naive MC estimation with 10% relative error 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 90%. This has triggered the need for alternative methods to naive MC simulations with improved computational efficiency.

IS is a variance reduction technique which aims to increase the computational efficiency of the naive MC simulation (Bucklew 2004). 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 $g_i(\cdot)$, $i = 1, 2, ..., N$, 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 of 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 $j = 1, ..., M$, 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 (Asmussen and Kroese 2006; Juneja 2007) and the logarithmic efficiency property (Asmussen and Kroese 2006; Juneja and Shahabuddin 2002). In practice, it is difficult to achieve the bounded relative error property. This has lead researchers to often settle for estimators satisfying weaker properties such as the logarithmic efficiency criterion. Let us consider the RV $T_{\gamma_{th}}$ 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

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

Applying the Logarithm on both side and using the fact that $\log(\alpha) < 0$, we conclude that, for all p^*, we have

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

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

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

It is worth mentioning that the naive MC simulation is not logarithmically efficient for the estimation of α since, in this case, the limit 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 (a RV X is said to have a light-tailed distribution if its MGF $M_X(\theta)$ is finite for some $\theta > 0$, see Kroese et al. (2011)). The exponential twisting by an amount $\theta \geq 0$ 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 MGF of the RV X_i, $i = 1, 2, ..., N$. In most of the cases, this technique achieves optimal efficiency results (Asmussen and Glynn 2007; Sadowsky and Bucklew 1990; Sadowsky 1993).

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 Juneja and Shahabuddin (2002), 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. We define the hazard rate $\lambda_i(\cdot)$ associated to the RV X_i as:

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

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

$$\Lambda_i(x) = -\log\left(1 - F_i(x)\right). \quad (12)$$

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

$$f_i(x) = \lambda_i(x) \exp\left(-\Lambda_i(x)\right). \quad (13)$$

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

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

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

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

The above hazard rate twisting change of measure was shown in Juneja and Shahabuddin (2002) to achieve the logarithmic efficiency property with $\theta = 1 - b/\Lambda(\gamma_{th})$, where b is any positive constant, for problems involving i.i.d sums of subexponential non-negative variates. In this work, we propose to extend the result that the hazard rate twisting technique possess the logarithmic efficiency criterion to the general framework of sums involving independent but not necessarily identically distributed arbitrary RVs. This generalized result is mainly due to our particular choice of the twisting parameter θ via a minmax approach which will be described in the next section. It is worth recalling that our main objective is to propose a generic IS approach that could be applicable to arbitrary sums of RVs. We do not claim that our approach would outperform any other existing method in the literature. For instance, the exponential twisting technique is likely to outperform our method in the light-tail setting. However this does not call into question the worthiness of our technique, as it is capable to address a large scope of scenarios, not necessarily covered by the exponential twisting technique.

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.

3 Proposed Hazard Rate Twisting Approach

3.1 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):

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

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'):

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

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

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

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

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

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

$$\sum_{i=1}^{N} X_i^*(\gamma_{th}) = \gamma_{th}. \tag{21}$$

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 logarithmic efficiency criterion holds for arbitrary sums of independent and not necessarily identically distributed RVs.

3.2 Logarithmic Efficiency Criterion

This section is devoted to the proof of our main result. In particular, we prove that by using the twisting parameter θ^* given in (20), the logarithmic efficiency of the corresponding estimator holds for any arbitrary sums of 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* $A(\gamma_{th}) = \sum_{i=1}^{N} \Lambda_i(X_i^*(\gamma_{th}))$*. Then, we have*

$$\lim_{\gamma_{th} \to +\infty} A(\gamma_{th}) = +\infty. \tag{22}$$

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

$$\cap_{i=1}^{N} \{X_i \geq X_i^*(\gamma_{th})\} \subset \left\{ \sum_{i=1}^{N} X_i \geq \gamma_{th} \right\}. \tag{23}$$

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

$$A\left(\gamma_{th}\right) \geq -\log\left(\alpha\right). \tag{24}$$

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 logarithmically efficiently estimated using the 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

$$\mathbb{E}_{\theta^*}\left[T_{\gamma_{th}}^2\right] \leq \left(\frac{A(\gamma_{th})}{N}\right)^{2N} \exp\left(2N - 2A(\gamma_{th})\right). \tag{25}$$

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

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

Now, combining (24) and (26) and using the fact that the right-hand side of (26) is negative for a sufficiently large γ_{th} (this follows from Lemma 1), we get

$$\frac{\log\left(\mathbb{E}_{\theta^*}\left[T^2_{\gamma_{th}}\right]\right)}{\log(\alpha)} \geq \frac{2N\left(1+\log\left(\frac{A(\gamma_{th})}{N}\right)\right)-2A(\gamma_{th})}{-A(\gamma_{th})}. \tag{27}$$

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

$$\lim_{\gamma_{th}\to+\infty} \frac{\log\left(\mathbb{E}_{\theta^*}\left[T^2_{\gamma_{th}}\right]\right)}{\log(\alpha)} \geq 2. \tag{28}$$

Hence, from (8), the logarithmic efficiency (9) holds thereby ending the proof.

3.3 Case Study

Theorem 1 establishes the logarithmic efficiency criterion of the proposed IS estimator which uses θ^* as the twisting parameter. While the logarithmic efficiency 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 functions including Weibull RVs with shape parameter greater than 1 can be handled using convex optimization algorithms (Boyd and Vandenberghe 2004). If the convexity of the hazard 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 with continuous PDFs belonging to the same family of distributions (for instance a sum of Weibull RVs with different shape and scale parameters) with hazard 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 (Jelenkovic and Momcilovic 2002). 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 (Albert W. Marshall 2007). 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 a global minimizer of (P') satisfies 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 $e_1, ..., e_N$ such that $e_i = (\eta_1, ..., \eta_{i-1}, \gamma_{th} - \sum_{j \neq i} \eta_j, \eta_{i+1}, ..., \eta_N)$. Therefore the minimum of (P') over $S(N, \gamma_{th})$ is either achieved in one of the extreme points 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

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

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 $\{1, 2, ..., N\}$. 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*

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

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

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

The remaining work is to prove that

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

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 $x \geq \eta_{i_0}$. Note that in the case where $\eta_{i_0} = 0$, $\lambda_{i_0}(\eta_{i_0})$ may be infinite. In this case, we use the concavity of $\Lambda_{i_0}(\cdot)$ in $[a, +\infty]$ for a fixed $a > 0$ so that $\lambda_{i_0}(x)$ is upper bounded by $\lambda_{i_0}(a)$ for all $x \geq a$. In both cases, there exists $a > 0$ such that $\Lambda_{i_0}(\cdot)$ is Lipschitz in the interval $[a, +\infty)$, that is for all x and y in the interval $[a, +\infty)$, we have

$$\left|\Lambda_{i_0}(x) - \Lambda_{i_0}(y)\right| \leq \lambda_{i_0}(a)\left|x - y\right|. \tag{49}$$

By taking $x = \gamma_{th}$ and $y = X_{i_0}^*(\gamma_{th})$, it follows that as $\gamma_{th} \to +\infty$:

$$\Lambda_{i_0}(\gamma_{th}) - \Lambda_{i_0}(X_{i_0}^*(\gamma_{th})) = \mathcal{O}(\gamma_{th} - X_{i_0}^*(\gamma_{th})). \tag{50}$$

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

$$\Lambda_{i_0}(\gamma_{th}) - \Lambda_{i_0}(X_{i_0}^*(\gamma_{th})) = o(\Lambda_{i_0}(\gamma_{th})), \tag{51}$$

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

Remark 2 *Distributions satisfying (29) were considered in Juneja and Shahabuddin (2002) 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 Juneja and Shahabuddin (2002), as γ_{th} increases. Hence, we deduce that our proposed approach recovers the method of Juneja and Shahabuddin (2002) in the particular problems involving i.i.d sums of subexponential distributions with eventually concave hazard functions.*

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, for a sufficiently large γ_{th}, the following*

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

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

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

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

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

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 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:

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

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

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

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

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

3.4 Algorithm

A pseudo-code describing all steps to estimate α by the 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 $\{X_i\}_{i=1}^N$ according to the twisted PDFs $\{f_{i,\theta^*}(\cdot)\}_{i=1}^N$. To this end, several methods can be used, among them, we distinguish the acceptance rejection technique, the Markov Chain Monte Carlo algorithm (Kroese et al. 2011), or the inverse CDF sampling method (Devroye 1986). The inverse CDF sampling method is merely based on the observation that, for a given a CDF $F(\cdot)$, the

RV $F^{-1}(U)$ where U is uniformly distributed RV over $[0,1]$ has a CDF given by $F(\cdot)$. For this method to be applicable, an analytical expression for the CDF inverse is required. In the sequel, we show that the CDF inverse of the twisted RV is related to that of the non-twisted RV. 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 $F(\cdot)$. 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 have 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 with the same shape parameter k but with a different scale parameter β' as follows*

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

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

4 Simulation Results

This section presents some selected simulations results in order to illustrate the performance of the proposed IS scheme. First of all, we illustrate the wide scope of applicability of the proposed estimator through various comparisons with some existing methods. Then, we analyze in a second 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 $T_{\gamma_{th}}$).

4.1 Efficiency of the Proposed IS Algorithm

The main objective of this part is to emphasize the wide scope of applicability of our proposed approach. In fact, contrary to previous derived estimators which were shown to be efficient only under some specific classes of distributions, our approach has the feature of being logarithmic efficient for problems involving arbitrary sums of RVs. To illustrate the latter statement, we perform two comparisons with first the approach of Chan and Kroese (2011) and second with that of Nandayapa (2008) and we aim to identify interesting scenarios in which our proposed approach exhibits better performances. Note that these two algorithms are extensions of the conditional Monte Carlo approach proposed in Asmussen and Kroese (2006), originally derived to deal with i.i.d sums of RVs, to independent and not identically distributed sums of variates. In the remaining part of the paper, we denote the approaches of Chan and Kroese (2011) and Nandayapa (2008) by CMC_1 and CMC_2 respectively. Our choice is mainly motivated by the good performances of these algorithms. In fact, when all RVs are i.i.d, the algorithm of Asmussen and Kroese (2006) is known to achieve a bounded relative error property in the case where the RVs are drawn from regularly varying distributions and to satisfy the logarithmic efficiency criterion when the RVs follow a Weibull-like distributions with shape parameter k less than $\log(3/2)/\log(2)$. The authors in Hartinger and Kortschak (2009) have extended the result of Asmussen and Kroese (2006) by showing that the estimator achieves a stronger criterion, namely; the asymptotically vanishing relative error property, for i.i.d sum of regularly varying distributions, the standard Log-normal variate, and the Weibull RV for $k < \log(3/2)/\log(3)$. The work of Chan and Kroese (2011) has extended the estimator of Asmussen and Kroese (2006) to sums involving independent but not identically distributed RVs and show that the bounded relative error property holds under the Pareto distribution. Moreover, it was proven numerically in Chan and Kroese (2011) that for the Weibull distribution the proposed algorithm performs much better when the shape parameters are small than when they are large. An alternative extension of the estimator of Asmussen and Kroese (2006) to independent and not identically distributed sums of variates has been proposed in Nandayapa (2008) where it was shown that the proposed estimator achieves the bounded relative error for Log-normal and regularly varying RVs.

In our opinion, CMC_1 and CMC_2 might exhibit better performances than the proposed IS approach for problems where their proposed estimators were shown to be efficient. But, in the general case, there is no guarantee that this always occurs. In particular, scenarios which do not fall within the scope of efficiency of these two algorithms, constitute potential situations in which our method might achieve better performances. This is the aim of this part where we identified various interesting scenarios in which our approach exhibits better performances. It is important to mention that the latter statement is general and valid regardless of the algorithm we are comparing with. More precisely, this statement is not restricted to CMC_1 and CMC_2 since we can end up with the same conclusions if we perform the comparison with any existing method, i.e. we can always identify some scenarios in which our approach outperforms the method we are comparing with. This is because, as it was mentioned before, our proposed approach is logarithmic efficient for arbitrary sums of variates whereas, to the best of our knowledge, all the existing estimators share the common dominator of being only efficient under specific classes of distributions.

4.1.1 Comparison with CMC_1

In this subsection, we compare our proposed IS scheme to the algorithm of Chan and Kroese (2011) . The proposed estimator in Chan and Kroese (2011) writes as:

$$\hat{\alpha}_{CMC_1} = \frac{1}{M}\sum_{k=1}^{M} T'_{\gamma_{th}}(\omega_k), \tag{67}$$

where $T'_{\gamma_{th}} = \sum_{i=1}^{N} \bar{F}_i\left(\max(\gamma_{th} - \sum_{j\neq i} X_j, M_{-i})\right)$, $\bar{F}_i(\cdot) = 1 - F_i(\cdot)$, and $M_{-i} = \max(X_1, ..., X_{i-1}, X_{i+1}, ..., X_N)$, $i = 1, 2, ..., N$.

In order to perform the comparison, we need to define some performance measures. First, the variance reduction metric of our proposed IS scheme compared to naive MC simulations is defined as:

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

Similarly, the amount of variance reduction of CMC_1 with respect to the naive MC simulations is defined as

$$\xi_{CMC_1} = \frac{\alpha(1-\alpha)}{\mathrm{var}\left[T'_{\gamma_{th}}\right]}. \tag{69}$$

The amount of variance ξ_{IS} (respectively ξ_{CMC_1}) measures the gain achieved by the proposed IS scheme (respectively the CMC_1 technique) over naive MC simulations in terms of necessary number of simulation runs to meet a fixed accuracy requirement.

In order to be able to include the computational time in our efficiency study, we define another metric, which serves to compare the proposed IS approach with the CMC_1 algorithm, as follows (Asmussen et al. 2011):

$$\text{eff} = \frac{\text{var}_{\theta^*}\left[T_{\gamma_{th}}\right]}{\text{var}\left[T'_{\gamma_{th}}\right]} \frac{\text{time}_{IS}}{\text{time}_{CMC_1}}. \tag{70}$$

This metric is a measure of efficiency that includes not only the computational gain, i.e. the gain in terms of number of simulation runs for a fixed accuracy requirement, but also the computational time needed to get the estimators. Note that the smaller is eff the bigger is the efficiency of the proposed IS approach compared to the CMC_1 one. In the following simulation results, the computational time is provided in seconds.

In the Weibull setting, the CMC_1 approach might outperform our approach for small values of the shape parameters. However, there is no guarantee that this will always happen. In fact, we identify in the following results three different settings depending on the nature of the tail of the underlying RVs and we show the outperformance of the proposed IS scheme over the CMC_1 one. 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 1 provides the performance results for the CMC_1 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 simulations. Moreover, the proposed IS technique achieves better performances than that of the CMC_1. The gain in efficiency becomes even higher as the threshold increases, i.e. the efficiency metric eff decreases as we increase the threshold. Moreover, in terms of number of simulation runs, our proposed IS scheme offers a computational gain over the CMC_1 approach. 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_1 method.

Table 1 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 $M = 10^7$.

	Proposed Approach			CMC_1 Approach			
γ_{th}	$\hat{\alpha}_{IS}$	ξ_{IS}	time_{IS}	$\hat{\alpha}_{CMC_1}$	ξ_{CMC_1}	time_{CMC_1}	eff
35	1.34e-4	200.30	6.66	1.34e-4	113.56	22.69	0.1664
40	1.74e-5	1.05e3	6.62	1.74e-5	282.01	22.56	0.0788
45	2.18e-6	5.42e3	6.59	2.18e-6	694.27	22.66	0.0373
50	2.76e-7	2.44e4	6.68	2.76e-7	1.49e3	22.66	0.0180
55	3.44e-8	1.08e5	6.69	3.40e-8	4.03e3	22.62	0.0110

In a second experiment, we consider the case where the underlying sum involves a mixture of heavy and light tailed RVs. Table 2 presents the obtained results when the sum of $N = 10$ independent Weibull distributions with shape parameter being selected from $\{0.8, 1\}$.

Table 2 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 $M = 10^7$.

	Proposed Approach			CMC_1 Approach			
γ_{th}	$\hat{\alpha}_{IS}$	ξ_{IS}	time_{IS}	$\hat{\alpha}_{CMC_1}$	ξ_{CMC_1}	time_{CMC_1}	eff
30	8.26e-5	565.75	4.06	8.22e-5	99.52	19.62	0.0364
35	4.88e-6	5.67e3	4.03	4.91e-6	292.22	19.89	0.0104
40	2.64e-7	6.01e4	4.07	2.65e-7	956.52	19.46	0.0033
45	1.36e-8	6.21e5	4.08	1.40e-8	1.99e3	19.79	6.61e-4

From this table, it becomes clear that the proposed IS approach can achieve better performances than that of the CMC_1 algorithm. The gain in performance is higher than the one shown by Table 1. For example, when $\gamma_{th} = 45$, our IS approach is 1513 times more efficient than the CMC_1 algorithm. This result is quite expected, the efficiency of the CMC_1 algorithm being shown in Chan and Kroese (2011) for small shape parameters.

Finally, we compare the performance of the proposed scheme where the sum includes only light-tailed RVs. 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 3 Sum of $N = 10$ independent Weibull Distribution with $\beta_i = 0.5 + i/10$, $k_i = 2$, $i = 1, 2, ..., 10$, and $M = 10^7$.

	Proposed Approach			CMC_1 Approach			
γ_{th}	$\hat{\alpha}_{IS}$	ξ_{IS}	time_{IS}	$\hat{\alpha}_{CMC_1}$	ξ_{CMC_1}	time_{CMC_1}	eff
15	5.65e-4	92.47	3.45	5.64e-4	17.12	18.61	0.0343
16	8.03e-5	429.41	3.45	8.05e-5	28.88	18.58	0.0125
17	9.17e-6	2.47e3	3.50	9.18e-6	53.01	18.67	0.0040
18	8.55e-7	1.76e4	3.51	8.74e-7	78.44	18.41	8.4972e-4
19	6.42e-8	1.55e5	3.42	6.94e-8	158.52	18.57	1.8835e-4

Table 3 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 3, the gain of our method over the CMC_1 method is evidently clear. It is important to note that for the above three scenarios, the computational time required by our method is less than that needed by the CMC_1 approach. This fact shows that in these three cases the resolution of (P') is not time-consuming.

4.1.2 Comparison with CMC_2

In this part, we aim to confirm the conclusions deduced from the previous comparisons. In fact, we perform a second comparison of our proposed IS approach with that of Nandayapa (2008) and we show that we can again identify scenarios wherein our proposed IS approach exhibits better performances. The estimator of Nandayapa (2008) is as follows:

$$\hat{\alpha}_{CMC_2} = \frac{1}{M} \sum_{k=1}^{M} T^{''}_{\gamma_{th}}(\omega_k), \tag{71}$$

where $T^{''}_{\gamma_{th}} = \frac{1}{p(J)} \bar{F}\left(\max\left(\gamma_{th} - \sum_{i \neq J} X_i, M_{-J}\right)\right)$, and J is a discrete RV with probability mass function $P(J = j) = \frac{\bar{F}_j(\gamma_{th})}{\sum_{i=1}^{N} \bar{F}_i(\gamma_{th})}$. It was shown in Nandayapa (2008) that the previous estimator achieves the bounded relative error property under Log-normal or regularly varying distributions. In these two settings, we expect the CMC_2 estimator to achieve better performances than our proposed IS estimator (this is because the bounded relative error property is stronger than the logarithmic efficiency criterion that our estimator possess). However, given that our estimator is general, being logarithmic efficient for arbitrary sums of variates, we can always identify interesting scenarios (other than the sum of Log-normal or regularly varying distributions) wherein our IS approach could achieve better performances than the CMC_2 algorithm. To illustrate the previous statement, we consider, as in the previous part, the problem of evaluating the probability that a sum of Weibull RVs exceeds a certain threshold. Numerical results shows, similarly to the previous comparison, that for small values of shape parameters of the Weibull distributions, the CMC_2 might exhibit better performances than the proposed IS approach. This results is expected since in the i.i.d setting for instance the CMC_2 algorithm, which coincides with that of Asmussen and Kroese (2006), has the asymptotic vanishing relative error property for $k < \log(3/2)/\log(3)$. However, as we increase these parameters, the outperformances of our method becomes evidently clear. These statements will validated in the following tables.

For the comparison, we use the same metrics as in the previous part. In a first experiment, we consider the sum of $N = 5$ Weibull variates with shape parameters equals to either 0.7 or 0.75. Table 1 shows the computational gain over naive MC simulation and the computational time given by the proposed IS scheme as well as the CMC_2 algorithm. The efficiency metric eff defined in (70) is also provided in this table.

Table 4 Sum of $N = 5$ independent Weibull Distribution with $\beta_i = 0.5 + i/10$, $i = 1, 2, ..., 5$, $k_i = 0.7$, $i = 1, 2$, $k_i = 0.75$, $i = 3, 4, 5$, and $M = 10^7$.

	Proposed Approach			CMC_2 Approach			
γ_{th}	$\hat{\alpha}_{IS}$	ξ_{IS}	time$_{IS}$	$\hat{\alpha}_{CMC_2}$	ξ_{CMC_2}	time$_{CMC_2}$	eff
25	2.42e-4	153.92	3.41	2.42e-4	282.54	6.30	0.9936
30	4.14e-5	539.67	3.40	4.16e-5	813.86	6.29	0.8152
35	7.59e-6	1.75e3	3.35	7.58e-6	2.32e3	6.28	0.7072
40	1.47e-6	5.47e3	3.42	1.47e-6	6.53e3	6.27	0.6512
45	2.99e-7	1.66e4	3.41	2.99e-7	1.86e4	6.28	0.6084

This table shows again that both techniques achieve a considerable gain over naive MC simulations. Moreover, while the CMC_2 algorithm exhibits slightly better performances over our proposed IS approach in terms of required number of simulation runs, i.e. ξ_{CMC_2} is bigger than ξ_{IS}, Table 4 shows that our IS approach is slightly more efficient than the CMC_2 method, i.e. the efficiency metric eff is less than 1 for all values of γ_{th} presented in Table 4. The latter statement is mainly due to the fact that the computational time required by our estimator is less than that required by the CMC_2 estimator.

In a second experiment, we aim to show that the efficiency of our approach over the CMC_2 algorithm can be made larger if we increase the shape parameters of the Weibull distributions. To this end, we consider in Table 5 the problem involving the sum of $N = 10$ i.i.d Weibull RVs with shape parameter $k = 0.8$.

Table 5 Sum of $N = 10$ independent Weibull Distribution with $\beta_i = 1$, $k_i = 0.8$, $i = 1, 2, ..., 10$ and $M = 10^7$.

	Proposed Approach			CMC_2 Approach			
γ_{th}	$\hat{\alpha}_{IS}$	ξ_{IS}	time$_{IS}$	$\hat{\alpha}_{CMC_2}$	ξ_{CMC_2}	time$_{CMC_2}$	eff
40	4.22e-5	898.41	6.46	4.24e-5	133.54	8.86	0.1084
45	6.60e-6	4.26e3	6.56	6.60e-6	279.76	8.87	0.0486
50	1.02e-6	1.92e4	6.54	1.02e-6	609.22	8.85	0.0234
55	1.58e-7	8.51e4	6.51	1.57e-7	1.27e3	8.88	0.0109
60	2.47e-8	3.62e5	6.56	2.49e-8	2.82e3	8.92	0.0057

From this table, we see clearly the outperformance of our approach compared to CMC_2. Moreover, the efficiency is increasing as we increase the threshold, i.e. the efficiency metric eff is decreasing as γ_{th} increases. For instance, for $\gamma_{th} = 60$, our proposed IS scheme is approximately 175 times more efficient than the CMC_2 algorithm.

In a last experiment, we aim to study the impact of increasing N on our approach as well as on the CMC_2 one. To this end, we consider the problem of the sum of $N = 15$ i.i.d Weibull variates with the same parameters as the previous experiment. Table 6 shows the efficiency results given by our IS approach and the CMC_2 method.

Table 6 Sum of $N = 15$ independent Weibull Distribution with $\beta_i = 1$, $k_i = 0.8$, $i = 1, 2, ..., 15$ and $M = 10^7$.

	Proposed Approach			CMC_2 Approach			
γ_{th}	$\hat{\alpha}_{IS}$	ξ_{IS}	time$_{IS}$	$\hat{\alpha}_{CMC_2}$	ξ_{CMC_2}	time$_{CMC_2}$	eff
50	3.55e-5	432.62	9.76	3.57e-5	72.07	9.27	0.1754
55	6.16e-6	1.98e3	9.79	6.28e-6	135.12	9.19	0.0727
60	1.06e-6	9.03e3	9.80	1.05e-6	241.03	9.27	0.0282
65	1.78e-7	3.94e4	9.72	1.75e-7	457.20	9.20	0.0123
70	2.97e-8	1.70e5	9.76	2.96e-8	804.35	9.30	0.0050

This table reveals that as N increases, the amount of variance reduction with respect to naive MC simulations, of our approach and the CMC_2 method decreases. For sake of illustration, for a probability of the order of 10^{-6}, the amount of variance reduction of the proposed IS scheme is approximately 1.92e3 when $N = 10$ and 9e3 when $N = 15$, showing that a greater N results in less amount of variance reduction. Moreover, this table confirms again the high gains of our proposed method compared to the CMC_2 approach. As a matter of fact, our approach is 200 times more efficient than the CMC_2 method for a threshold value equal to 70. Furthermore, we observe from Table 5 and Table 6 that the efficiency of our method compared to the CMC_2 algorithm remains unchanged as N increases, showing that the degradation in performance of both techniques occurs with the same rate. Finally, from the above three tables, it is important to mention that, similarly to the conclusion made in the comparison with the CMC_1 approach, the computational time needed by our proposed algorithm is less or comparable to that needed by the CMC_2 method. This shows again that the computational complexity of solving (P') was not time-demanding.

4.2 Sensitivity Analysis of the Minmax Approach

Obviously, the optimal choice of the twisting parameter $\hat{\theta}$ 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 proposed 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 logarithmic efficiency, 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. 1, $\mathbb{E}_{\theta^*}\left[T^2_{\gamma_{th}}\right]$ 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. Note that the optimal twisting parameter $\hat{\theta}$ is approximated by successive dichotomy. 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

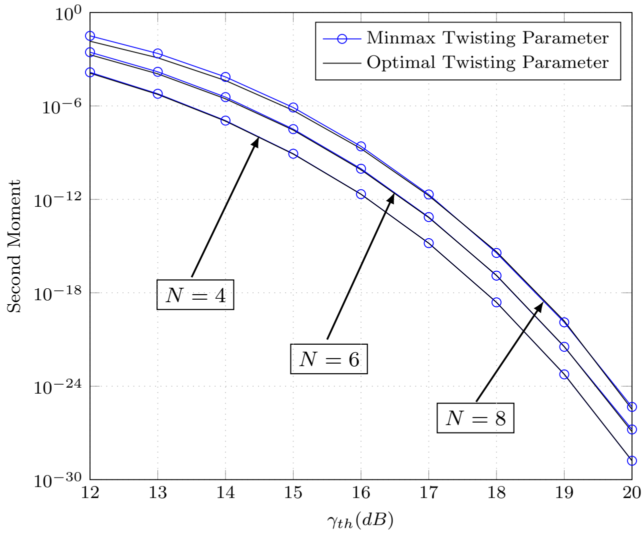

Fig. 1 Second moment of $T_{\gamma_{th}}$ with the minmax and the optimal twisting parameter for the sum of N Weibull RVs with shape parameters $k_i = 0.8$, scale parameters $\beta_i = 1$, $i = 1, 2, ..., N$, and $M = 10^7$.

parameter θ^* in retrieving approximately the same performances obtained by using the twisting parameter $\hat{\theta}$.

5 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 logarithmic efficiency 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 two algorithms based on a conditional Monte Carlo technique were presented and illustrated the efficiency of our technique in handling a large range of scenarios.

References

I.O. Albert W. Marshall, *Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Familie*. Science & Business Media (Springer, New York, 2007)

S. Asmussen, D.P. Kroese, Improved algorithms for rare event simulation with heavy tails. Advances in Applied Probability **38**(2), 545–558 (2006)

S. Asmussen, D. Kortschak, Error rates and improved algorithms for rare event simulation with heavy weibull tails. Methodology and Computing in Applied Probability **17**(2), 441–461 (2015)

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

S. Asmussen, P.W. Glynn, *Stochastic simulation : algorithms and analysis*. Stochastic modelling and applied probability (Springer, New York, 2007)

F. Babich, G. Lombardi, Statistical analysis and characterization of the indoor propagation channel. IEEE Transactions on Communications **48**(3), 455–464 (2000)

N.C. Beaulieu, F. Rajwani, Highly accurate simple closed-form approximations to Lognormal sum distributions and densities. IEEE Communications Letters **8**(12), 709–711 (2004)

N.C. Beaulieu, Q. Xie, An optimal Lognormal approximation to Lognormal sum distributions. IEEE Transactions on Vehicular Technology **53**(2), 479–489 (2004)

K. Ben Letaief, Performance analysis of digital lightwave systems using efficient computer simulation techniques. IEEE Transactions on Communications **43**(234), 240–251 (1995)

N. Ben Rached, F. Benkhelifa, M.-S. Alouini, R. Tempone, A fast simulation method for the Log-normal sum distribution using a hazard rate twisting technique, in *Proc. of the IEEE International Conference on Communications (ICC'2015)*, London, UK, 2015a

N. Ben Rached, A. Kammoun, M.-S. Alouini, R. Tempone, An improved hazard rate twisting approach for the statistic of the sum of subexponential variates. IEEE Communications Letters **19**(1), 14–17 (2015b)

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

S.P. Boyd, L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, 2004)

J.A. Bucklew, *Introduction to Rare Event Simulation*. Springer series in statistics (Springer, New York, 2004)

J.C. Chan, D. Kroese, Rare-event probability estimation with conditional Monte Carlo. Annals of Operations Research **189**(1), 43–61 (2011)

L. Devroye, *Non-Uniform Random Variate Generation* (Springer, New York, 1986)

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

L. Fenton, The sum of Log-normal probability distributions in scatter transmission systems. IRE Transactions on Communications Systems **8**(1), 57–67 (1960)

J.C.S.S. Filho, M.D. Yacoub, Simple precise approximations to Weibull sums. IEEE Communications Letters **10**(8), 614–616 (2006)

M. Ghavami, R. Kohno, L. Michael, *Ultra wideband signals and systems in communication engineering* (Wiley, Chichester, 2004)

J. Hartinger, D. Kortschak, On the efficiency of the asmussen-kroese-estimator and its application to stop-loss transforms. Blätter der DGVFM **30**(2), 363–377 (2009)

A. Healey, C.H. Bianchi, K. Sivaprasad, Wideband outdoor channel sounding at 2.4 GHz, in *Proc. of the IEEE Conference on Antennas and Propagation for Wireless Communications*, Waltham, MA, US, 2000

J. Hu, N.C. Beaulieu, Accurate simple closed-form approximations to Rayleigh sum distributions and densities. IEEE Communications Letters **9**(2), 109–111 (2005)

P. Jelenkovic, P. Momcilovic, Resource sharing with subexponential distributions, in *Proc. of the IEEE 21st Annual Joint Conference of the Computer and Communications (INFOCOM' 2002)*, New York, USA, 2002

S. Juneja, Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error. Queueing Systems **57**(2-3), 115–127 (2007)

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

D.P. Kroese, T. Taimre, Z.I. Botev, *Handbook of Monte Carlo methods* (Wiley, N.J, 2011)

D.P. Kroese, R.Y. Rubinstein, The transform likelihood ratio method for rare event simulation with heavy tails. Queueing Systems **46**(3), 317–351 (2004)

L.R. Nandayapa, Risk probabilities: asymptotics and simulation, phd thesis, university of aarhus (2008)

S.M. Navidpour, M. Uysal, M. Kavehrad, BER performance of free-space optical transmission with spatial diversity. IEEE Transactions on Wireless Communications **6**(8), 2813–2819 (2007)

R.Y. Rubinstein, D.P. Kroese, *The Cross Entropy Method: A Unified Approach To Combinatorial Optimization, Monte-carlo Simulation (Information Science and Statistics)* (Springer, Secaucus, NJ, USA, 2004)

J.S. Sadowsky, On the optimality and stability of exponential twisting in Monte Carlo estimation. IEEE Transactions on Information Theory **39**(1), 119–128 (1993)

J.S. Sadowsky, J.A. Bucklew, On large deviations theory and asymptotically efficient Monte Carlo estimation. IEEE Transactions on Information Theory **36**(3), 579–588 (1990)

N.C. Sagias, G.K. Karagiannidis, Gaussian class multivariate weibull distributions: Theory and applications in fading channels. IEEE Transactions on Information Theory **51**(10), 3608–3619 (2005)

S.C. Schwartz, Y.S. Yeh, On the distribution function and moments of power sums with Lognormal component. The Bell Systems Technical Journal (1982)

M.K. Simon, M.-S. Alouini, *Digital Communication over Fading Channels, 2nd Edition* (Wiley, New York, USA, 2004)

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

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