Efficient importance sampling for large sums of independent and identically distributed random variables

Nadhir Ben Rached, Abdul-Lateef Haji-Ali, Gerardo Rubino, Raúl Tempone

Efficient Importance Sampling for Large Sums of Independent and Identically Distributed Random Variables

Nadhir Ben Rached *, Abdul-Lateef Haji-Ali †, Gerardo Rubino ‡, and Raúl Tempone §

¶

Abstract

We aim to estimate the probability that the sum of nonnegative independent and identically distributed random variables falls below a given threshold, i.e., $\mathbb{P}(\sum_{i=1}^{N} X_i \leq \gamma)$, via importance sampling (IS). We are particularly interested in the rare event regime when N is large and/or γ is small. The exponential twisting is a popular technique that, in most of the cases, compares favorably to existing estimators. However, it has several limitations: i) it assumes the knowledge of the moment generating function of X_i and ii) sampling under the new measure is not straightforward and might be expensive. The aim of this work is to propose an alternative change of measure that yields, in the rare event regime corresponding to large N and/or small γ, at least the same performance as the exponential twisting technique and, at the same time, does not introduce serious limitations. For distributions whose probability density functions (PDFs) are $\mathcal{O}(x^d)$, as $x \to 0$ and $d > -1$, we prove that the Gamma IS PDF with appropriately chosen parameters retrieves asymptotically, in the rare event regime, the same performance of the estimator based on the use of the exponential twisting technique. Moreover, in the Log-normal setting, where the PDF at zero vanishes faster than any polynomial, we numerically show that a Gamma IS PDF with optimized parameters clearly outperforms the exponential twisting change of measure. Numerical experiments validate the efficiency of the proposed estimator in delivering a highly accurate estimate in the regime of large N and/or small γ.

*Chair of Mathematics for Uncertainty Quantification, Department of Mathematics, RWTH Aachen University, Aachen, Germany (benrached@uq.rwth-aachen.de).

†School of Mathematical & Computer Sciences, Heriot-Watt University, Edinburgh, UK (a.hajiali@hw.ac.uk).

‡INRIA, Rennes - Bretagne Atlantique, France (Gerardo.Rubino@inria.fr).

§Computer, Electrical and Mathematical Sciences & Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia (raul.tempone@kaust.edu.sa). Alexander von Humboldt Professor in Mathematics for Uncertainty Quantification, RWTH Aachen University, Aachen, Germany (tempone@uq.rwth-aachen.de).

¶This work was supported by the KAUST Office of Sponsored Research (OSR) under Award No. URF/1/2584-01-01 and the Alexander von Humboldt Foundation. A-L. Haji-Ali was supported by a Sabbatical Grant from the Royal Society of Edinburgh.

Keywords: Importance sampling, rare event, exponential twisting, Gamma IS PDF.

AMS subject classifications: 65C05, 62P30.

1 Introduction

Efficient estimation of rare event probabilities finds various applications in the performance evaluation/prediction of wireless communication systems operating over fading channels [25]. In particular, the left-tail of the cumulative distribution function (CDF) of sums of nonnegative independent and identically distributed (i.i.d.) random variables (RVs) is an example of a rare event probability that is of practical importance. More specifically, the outage probability at the output of equal gain combining (EGC) and maximum ratio combining (MRC) receivers can be expressed as the CDF of the sum of fading channel envelops (for EGC) and fading channel gains (for MRC) [7].

The accurate estimation of the left-tail of the CDF of sums of RVs requires the use of variance reduction techniques because the naive Monte Carlo (MC) sampler is computationally expensive [4, 19, 23]. Moreover, the existing closed-form approximations [12, 13, 17, 20, 21, 27, 28] fail to be accurate when the tail of the CDF is considered. The literature is rich in works in which variance reduction techniques were developed to efficiently estimate rare event probabilities corresponding to the left-tail of the CDF of sums of RVs, see [1, 3, 5, 7, 9, 10, 15] and the references therein. For instance, the authors in [3] used the concept of exponential twisting, which is a popular importance sampling (IS) technique, to propose a logarithmically efficient estimator of the CDF of the sum of i.i.d. Log-normal RVs. The logarithmic efficiency is a popular property in rare event simulation used to ensure estimators' efficiency; see [7] for a formal definition of this property. In [15], the CDF of the sum of correlated Log-normal RVs was considered. The authors developed an IS estimator based on shifting the mean of the corresponding multivariate Gaussian distribution. Under mild assumptions, they proved that their proposed estimator is logarithmically efficient. Based on [15] and under the assumption that the left-tail sum distribution is determined by only one dominant component, the authors in [1] combined IS with the control variate technique to construct an estimator with the asymptotically vanishing relative error property, which is the most desired property in the field of rare event simulations [19]. In [7], two unified IS approaches were developed using the hazard rate twisting concept [8, 18] to efficiently estimate the CDF of sums of independent RVs. The first estimator is shown to be logarithmically efficient, whereas the second achieves the bounded relative error property (a stronger criterion than the logarithmic efficiency) for i.i.d. sums of RVs and under the given assumption that was shown to hold for most of the practical distributions used to model the amplitude/power of fading channels.

The efficiency of the above mentioned estimators was studied when the number of summand N was kept fixed. More specifically, recall that the objective is to efficiently estimate the probability that the sum of nonnegative i.i.d. RVs falls below a given threshold, i.e., $\mathbb{P}(\sum_{i=1}^{N} X_i \leq \gamma)$. A close look at the above mentioned estimators shows that the efficiency results were proved when the rarity parameter γ decreased whereas N is kept fixed. However, In most cases, the efficiency of the existing estimators was considerably affected when N increases. This represents the main motivation of the present work. We aim to introduce a highly accurate estimator that efficiently estimate $\mathbb{P}(\sum_{i=1}^{N} X_i \leq \gamma)$ in the rare event regime when N is large and/or γ is small. The main contributions of the present work are:

- The exponential twisting technique is a popular IS change of measure that, in most of the cases, compares favorably to existing estimators. It is the optimal IS probability density function (PDF) in the sense that it minimizes the Kullback-Leibler (KL) divergence with respect to the underlying PDF under certain constraints [22]. However, it has some limitations. First, it requires the knowledge of the moment generating function of X_i, $i = 1, 2, \cdots, N$. Second, sampling according to the new IS PDF is not straightforward and might be expensive. Moreover, the twisting parameter is not available in a closed-form expression and needs to be estimated numerically. The contribution of this work is to propose an alternative IS estimator that asymptotically yields at least the same efficiency as the one given by the estimator based on exponential twisting and at the same time does not introduce the above limitations.

- We prove that for distributions whose PDF vanishes at zero polynomially, the Gamma IS PDF with appropriately chosen parameters retrieves, in the regime of rare events, the same performances as the exponential twisting PDF.

- The above result does not apply to the Log-normal setting as the corresponding PDF approaches zero faster than any polynomial. Interestingly, we show that in this setting, the Gamma change of measure with optimized parameters achieves a substantial amount of variance reduction compared to the one given by exponential twisting.

- In addition to yielding at least the same performance as the estimator based on exponential twisting, the estimator based on the use of the Gamma PDF as an IS PDF is easy to implement. This is compared to the estimator based on exponential twisting whose implementation is possible only under restrictive assumptions.

- Numerical comparisons with some of the existing estimators validate that the proposed estimator can deliver highly accurate estimates with low computational cost in the rare event regime corresponding to large N and/or small γ.

The paper is organized as follows. In section 2, we define the problem setting and motivate the work. In section 3, we introduce the exponential twisting change of measure and present its limitations. The main contribution of this work is presented in section 4, where we show, under different scenarios, that the Gamma IS PDF with optimized parameters retrieves at least the same performance as the exponential twisting technique. Finally, numerical experiments are shown in section 5 to compare the proposed estimator with various existing estimators.

2 Problem Setting and Motivation

Let $X_1, X_2, \cdots, X_N$ be i.i.d. nonnegative RVs with common PDF $f_X(\cdot)$ and CDF $F_X(\cdot)$. Let $\boldsymbol{x} = (x_1, \cdots, x_n)^t$ and $h_{\mathbf{X}}(\boldsymbol{x}) = \prod_{i=1}^N f_X(x_i)$ be the joint PDF of the random vector $(X_1, \cdots, X_N)^t$. We consider the estimation of

$$\alpha(\gamma, N) = \mathbb{P}_{h_{\mathbf{X}}}\left(\sum_{i=1}^N X_i \leq \gamma\right),$$

where $\mathbb{P}_{h_{\mathbf{X}}}(\cdot)$ is the probability under which the random vector $\mathbf{X} = (X_1, \cdots, X_N)^t$ is distributed according to $h_{\mathbf{X}}(\cdot)$. We focus on the estimation of $\alpha(\gamma, N)$ when N is large and/or γ is small. Before delving into the core of the paper, we illustrate via a simple example that the efficiency of an IS estimator, that performs well when γ decreases and N is not sufficiently large, can deteriorate when we increase the values of N. We first write the quantity of interest as

$$\mathbb{P}_{h_{\mathbf{X}}}\left(\sum_{i=1}^N X_i \leq \gamma\right) = \mathbb{P}_{h_{\boldsymbol{w}}}\left(\sum_{i=1}^N w_i \leq 1\right)\left(F_X(\gamma)\right)^N, \quad (1)$$

where $w_i = \frac{X_i}{\gamma} | \{\frac{X_i}{\gamma} \leq 1\}$, $i = 1, 2, \cdots, N$. The estimator is then given by estimating the first factor on the right-hand side of (1) by the naive MC method. Note that this estimator can be understood as applying IS with IS PDF being the truncation of the underlying PDF over the hypercube $[0, \gamma]^N$. It can be easily proved that for fixed N, this estimator achieves the desired bounded relative error property with respect to the rarity parameter γ for distributions that satisfy $f_w(x) \sim bx^d$ as x approaches zero and for $d > -1$ and $b > 0$, see [9]. This property means that coefficient of variation, defined as the ratio between the standard deviation of an estimator and its mean, remains bounded as $\gamma \to 0$, see [19]. More precisely, when this property holds, the number of required samples to meet a fixed accuracy requirement remains bounded independently of how small $\alpha(\gamma, N)$ is. The question now is what happens when N is large. Using the Chernoff bound, we obtain

$$\mathbb{P}_{h_{\boldsymbol{w}}}(\sum_{i=1}^N w_i \leq 1) \leq \min_{\theta > 0} \exp\left(\theta + N \log\left(\mathbb{E}_{f_w}[\exp(-\theta w)]\right)\right),$$

where $E_{f_w}[\cdot]$ denotes the expectation under $f_w(\cdot)$. In particular, when $\theta = 1$, the squared coefficient of variation $1/\mathbb{P}_{h_{\boldsymbol{w}}}(\sum_{i=1}^N w_i \leq 1)$ is lower bounded by

$\exp\left(-1 - N\log\left(\mathbb{E}_{f_w}[\exp(-w)]\right)\right)$. This shows that the squared coefficient of variation increases at least exponentially, which proves that the efficiency of the estimator deteriorates when N is large.

3 Exponential Twisting

In this section, we review a popular IS approach, which is exponential twisting, and enumerate its limitations in estimating the quantity of interest. When applicable, it is well acknowledged that the exponential twisting technique is expected to produce a substantial amount of variance reduction and to compare favorably, in most of the cases, to existing estimators. For distributions with light right tails and under the i.i.d. assumption, the estimator based on exponential twisting can be proved, under some regularity assumptions, to be logarithmically efficient when the probability of interest is either $\mathbb{P}(\sum_{i=1}^N X_i > \gamma)$ and $\gamma \to +\infty$ or $\mathbb{P}(\sum_{i=1}^N X_i > \gamma N)$ and with $N \to +\infty$ [4]. In the left tail setting, which is the region of interest in the present work, the exponential twisting was shown in [3] to achieve the logarithmic efficiency property in the case of i.i.d. Log-normal RVs when the probability of interest is $\mathbb{P}(\sum_{i=1}^N X_i < N\gamma)$ and either $N \to +\infty$ or $\gamma \to 0$.

In [22], the exponential twisting technique was also shown to be optimal in the sense that it minimizes the KL divergence with respect to the underlying PDF under the constraint that the rare set $\{\boldsymbol{x}$, such that $\sum_{i=1}^N x_i \leq \gamma\}$ is no longer rare. The IS PDF is selected to be the solution of the following optimization problem, see [22]:

$$
\begin{aligned}
&\inf_{h^*_{\mathbf{X}} \geq 0} \int h^*_{\mathbf{X}}(\boldsymbol{x}) \log\left(\frac{h^*_{\mathbf{X}}(\boldsymbol{x})}{h_{\mathbf{X}}(\boldsymbol{x})}\right) d\boldsymbol{x} \\
&s.t \quad \int h^*_{\mathbf{X}}(\boldsymbol{x}) d\boldsymbol{x} = 1 \\
&\qquad \mathbb{E}_{h^*_{\mathbf{X}}}\left[\sum_{i=1}^N X_i\right] = \gamma \\
&\qquad h^*_{\mathbf{X}}(\boldsymbol{x}) \geq 0, \quad x_i \geq 0 \text{ for all } i.
\end{aligned}
\tag{2}
$$

The solution of this problem is given as (see [22] for a more general setting)

$$
h^*_{\mathbf{X}}(\boldsymbol{x}) = \frac{h_{\mathbf{X}}(\boldsymbol{x}) \exp\left(\theta \sum_{i=1}^N x_i\right)}{\mathbb{E}_{h_{\mathbf{X}}}\left[\exp\left(\theta \sum_{i=1}^N X_i\right)\right]}, \tag{3}
$$

and θ solves

$$
\frac{\mathbb{E}_{h_{\mathbf{X}}}\left[\sum_{i=1}^N X_i \exp\left(\theta \sum_{i=1}^N X_i\right)\right]}{\mathbb{E}_{h_{\mathbf{X}}}\left[\exp\left(\theta \sum_{i=1}^N X_i\right)\right]} = \gamma.
$$

Hence, we clearly observe that the optimal density is given by exponentially twisting each univariate PDF $f_X(\cdot)$

$$f_X^*(x) = \frac{f_X(x)\exp(\theta x)}{M(\theta)}, \quad x \geq 0,$$

with $M(\theta) = E_{f_X}[\exp(\theta X)]$ and θ satisfies

$$\frac{M^{'}(\theta)}{M(\theta)} = \gamma/N.$$

Observe, however, that the exponential twisting technique has various restrictive limitations. The main one is that sampling according to $f_X^*(\cdot)$ is not straightforward. One generally needs the use of an acceptance-rejection technique, the complexity of which can be dramatic when the probability of acceptance is relatively small. In such a case, the computational complexity of the algorithm can be huge and even worse than the naive MC method. There are other less critical drawbacks. First, computations are much simpler if the moment generating function $M(\theta)$ is known in closed-form. Such a requirement does not hold in general. Also, the twisting parameter θ does not have, in general, a closed-form expression, and hence, it should be approximated numerically.

4 Gamma Family as IS PDF

The objective of this paper is to propose an alternative change of measure that approximately yields at least the same performance as the exponential twisting technique and at the same time does not introduce serious limitations. We distinguish three scenarios depending on how the PDF $f_X(\cdot)$ approaches zero.

4.1 $f_X(x) \sim b$ as x goes to 0 and $b > 0$ is a constant

Recall that the exponential twisting change of measure satisfies

$$f_X^*(x) \propto f_X(x)\exp(\theta x), \quad x \geq 0,$$

with $\theta \to -\infty$ as $\gamma \to 0$ and/or $N \to +\infty$. Therefore, as $f(x) \sim b$ and $b > 0$, and by letting $\tilde{M}(\theta) = -\frac{1}{\theta}$, we instead consider the following change of measure

$$\tilde{f}_X(x) = \frac{\exp(\theta x)}{\tilde{M}(\theta)}, \quad x \geq 0.$$

The value of θ satisfies $\frac{\tilde{M}^{'}(\theta)}{\tilde{M}(\theta)} = \frac{\gamma}{N}$. Through simple computation, we obtain $\theta = -\frac{N}{\gamma}$. To conclude, when $f(x) \sim b$ and $b > 0$, we propose a change of measure given by the exponential distribution with rate $\frac{N}{\gamma}$. In the regime of rare events, we expect the proposed change of measure to yield approximately the same performance as the one given by exponential twisting.

Table I: Some PDF asymptotics around zero [a]

Distribution	PDF	Proportional to
Exponential $k > 0$	$k\exp(-kx)$	1
Gamma $k, \beta > 0$	$\frac{1}{\beta^k \Gamma(k)} x^{k-1} \exp(-\frac{x}{\beta})$	x^{k-1}
Weibull $k, \lambda > 0$	$\frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} \exp(-(\frac{x}{\lambda})^k)$	x^{k-1}
Nakagami-m $m, \Omega > 0$	$\frac{2m^m}{\Gamma(m)\Omega^m} x^{2m-1} \exp(-\frac{m}{\Omega}x^2)$	x^{2m-1}
Generalized Gamma $a, d, p > 0$	$\frac{p/a^d}{\Gamma(d/p)} x^{d-1} \exp(-(\frac{x}{a})^p)$	x^{d-1}
Rice $\sigma > 0, \nu \geq 0 > 0$	$\frac{x}{\sigma^2} \exp(-\frac{x^2+\nu^2}{2\sigma^2}) I_0((\frac{x\nu}{\sigma^2}))$	x
Gamma-Gamma $\Omega > 0, m > k > 0, m - k \notin \mathbb{N}$	$\frac{2(km)^{\frac{k+m}{2}}}{\Gamma(k)\Gamma(m)\Omega}(\frac{x}{\Omega})^{\frac{k+m}{2}-1} K_{k-m}\left(2\sqrt{\frac{kmx}{\Omega}}\right)$	x^{k-1}
$\kappa - \mu$ distribution $\kappa, \mu > 0$	$\frac{2\mu(1+\kappa)^{\frac{\mu+1}{2}} x^{\mu}}{\Omega^{\frac{\mu+1}{2}} \kappa^{\frac{\mu-1}{2}} \exp(\mu\kappa)} \exp(-\frac{(1+\kappa)\mu x^2}{\Omega}) I_{\mu-1}\left(2\mu\sqrt{\frac{\kappa(\kappa+1)}{\Omega}} x\right)$	$x^{2\mu-1}$

[a]Functions $I_\xi(\cdot)$, and $K_\xi(\cdot)$ are respectively the modified Bessel functions of the first kind and order ξ and the second kind and order ξ [14].

4.2 $f_X(x) = x^d g(x)$ with $g(x) \sim b$ as x goes to 0, $d > -1$, and $b > 0$

Using the same methodology as in section 4.1, the change of measure that we consider is

$$\tilde{f}_X(x) = \frac{x^d \exp(\theta x)}{\tilde{M}(\theta)}, \;\; x \geq 0.$$

Therefore, the new measure corresponds to the Gamma distribution with shape parameter $d+1$ and scale parameter $-1/\theta$. The normalizing constant is $\tilde{M}(\theta) = \frac{\Gamma(d+1)}{(-\theta)^{d+1}}$. Hence, the value of θ satisfying $\frac{\tilde{M}'(\theta)}{\tilde{M}(\theta)} = \frac{\gamma}{N}$ is given by

$$\theta = -\frac{N}{\gamma}(d+1).$$

In Table I, we provide a non-exhaustive list of distributions that belong to section 4.2 (note that distributions in section 4.2 include those in section 4.1).

These distributions are among the most practical distributions used to model the amplitudes and powers of wireless communications fading channels. For these distributions, the Gamma IS PDF is expected to asymptotically deliver the same performance as the one given by the exponential twisting IS PDF without introducing serious limitations.

4.3 Distributions that do not approach 0 polynomially

Distributions in this class are much more difficult to handle and need to be tackled on a case-by-case basis. We consider the sum of i.i.d. standard Log-normal RVs. The density decreases to 0 at a faster rate than any polynomials and thus the Gamma distribution with fixed shape parameter will not recover the results given by the use of the exponential twisting technique. Note that in [3], the exponential twisting technique was applied to the sum of i.i.d. standard Log-normals by i) providing an unbiased estimator of the MGF, ii) approximating the value of θ, and iii) using acceptance-rejection to sample from the IS PDF.

The main difficulty is that the PDF of the Log-normal distribution does not have a Taylor expansion at $x = 0$. The first estimator we propose is based on truncating the support $[0, +\infty]$ and only work on $[a, +\infty]$ with $a = \delta\gamma/N$. This allows to use a Taylor expansion at $x = a$. This procedure, however, creates a bias that needs to be controlled. We show numerically that this estimator exhibits better performances than the one based on exponential twisting. Moreover, we observe that, in the regime of rare events, the proposed estimator achieves approximately the same performances as the Gamma IS PDF with shape parameter equal to 2. This is the main motivation behind introducing a second estimator whose IS PDF is a Gamma PDF with optimized parameters. The nnumerical results show that the second estimator achieves substantial variance reduction with respect to the first estimator.

4.3.1 Biased estimator

We rewrite the quantity of interest as

$$\mathbb{P}_{h_{\mathbf{X}}}\left(\sum_{i=1}^{N} X_i \leq \gamma\right) \approx \left(1 - F_X(\frac{\delta\gamma}{N})\right)^N \times \mathbb{P}_{h_{\mathbf{X}}}\left(\sum_{i=1}^{N} X_i \leq \gamma \Big| X_i > \frac{\delta\gamma}{N}\right), \quad (4)$$

where δ is a fixed value belonging to $[0, 1)$. The first factor on the right-hand side has a known closed-form expression. Let $\bar{f}_X(\cdot)$ be the PDF of $X_i|\{X_i > \frac{\delta\gamma}{N}\}$, $i = 1, 2, \cdots, N$, whose expression is given as follows:

$$\bar{f}_X(x) = \frac{1}{x\sqrt{2\pi}} \frac{\exp\left(-\frac{(\log(x))^2}{2}\right)}{P(X_i > \frac{\delta\gamma}{N})}, \quad x \geq \frac{\delta\gamma}{N}.$$

Next, we write the second right hand side of (4) as follows:

$$\mathbb{P}_{h_{\mathbf{X}}}\left(\sum_{i=1}^{N} X_i \leq \gamma \Big| X_i > \frac{\delta\gamma}{N}\right) = \mathbb{P}_{\bar{h}_{\mathbf{X}}}\left(\sum_{i=1}^{N} X_i \leq \gamma\right),$$

with $\bar{h}_{\mathbf{X}}(\boldsymbol{x}) = \prod_{i=1}^{N} \bar{f}_X(x_i)$. The exponential twisting change of measure is then given by

$$\bar{f}_X^*(x) \propto \bar{f}_X(x)\exp(\theta x),\ \ x \geq \frac{\delta\gamma}{N}.$$

Now, by using the Taylor expansion of $\bar{f}_X(\cdot)$ at the point $x = \delta\gamma/N$, we write

$$\bar{f}_X(x) = \bar{f}_X(\frac{\delta\gamma}{N}) + (x - \frac{\delta\gamma}{N})\bar{f}_X^{'}(\frac{\delta\gamma}{N}) + \frac{(x-\frac{\delta\gamma}{N})^2}{2}\bar{f}_X^{''}(\xi_{x,\delta,N}),$$

where $\xi_{x,\delta,N}$ is between $\frac{\delta\gamma}{N}$ and x. Hence, the approximate exponential twisting change of measure is given by

$$\tilde{f}_X(x) = \frac{\bar{f}_X\exp(\theta x) + (x-\frac{\delta\gamma}{N})\bar{f}_X^{'}\exp(\theta x)}{\tilde{M}(\theta)},\ \ x \geq \frac{\delta\gamma}{N}, \tag{5}$$

with the notation $\bar{f}_X = \bar{f}_X(\frac{\delta\gamma}{N})$ and $\bar{f}_X^{'} = \bar{f}_X^{'}(\frac{\delta\gamma}{N})$. We assume that $\frac{\delta\gamma}{N}$ is strictly less than $\exp(-1)$ to ensure that $\bar{f}_X^{'} > 0$. This assumption is not restrictive, as we are interested in the rare event regime corresponding to N large and/or γ small. Through a simple computation, we get

$$\tilde{M}(\theta) = -\frac{\exp(\theta\delta\gamma/N)}{\theta}\bar{f}_X + \frac{\exp(\theta\delta\gamma/N)}{\theta^2}\bar{f}_X^{'}.$$

The value of θ that solves $\frac{\tilde{M}^{'}(\theta)}{\tilde{M}(\theta)} = \frac{\gamma}{N}$ is given by

$$\theta = -\frac{\bar{f}_X - c\bar{f}_X^{'} + \sqrt{(\bar{f}_X - c\bar{f}_X^{'})^2 + 8\bar{f}_X\bar{f}_X^{'}c}}{2c\bar{f}_X},$$

with $c = \frac{\gamma}{N}(1-\delta)$. The remaining part is to sample from $\tilde{f}_X(\cdot)$. To do this, we write

$$\tilde{f}_X(x) = -\frac{\bar{f}_X\exp(\theta\delta\gamma/N)}{\tilde{M}_X(\theta)\theta}\tilde{f}_1(x) + \frac{\bar{f}_X^{'}\exp(\theta\delta\gamma/N)}{\tilde{M}_X(\theta)\theta^2}\tilde{f}_2(x),$$

where $\tilde{f}_1(x) = -\frac{\theta\exp(\theta x)}{\exp(\theta\delta\gamma/N)}$ and $\tilde{f}_2(x) = \frac{\theta^2(x-\delta\gamma/N)\exp(\theta x)}{\exp(\theta\delta\gamma/N)}$ are two valid PDFs for $x > \delta\gamma/N$.

The question that remains is related to controlling the bias through a proper choice of the parameter δ. Let $\alpha_1(\gamma,N) = \left(1 - F_X(\frac{\delta\gamma}{N})\right)^N \mathbb{P}_{h_{\mathbf{X}}}\left(\sum_{i=1}^{N} X_i \leq \gamma \Big| X_i > \frac{\delta\gamma}{N}\right)$. Then, the global relative error can be upper bounded as follows:

$$\left|\frac{\alpha(\gamma,N) - \hat{\alpha}_{IS}}{\alpha(\gamma,N)}\right| \leq \frac{\alpha(\gamma,N) - \alpha_1(\gamma,N)}{\alpha(\gamma,N)} + \left|\frac{\alpha_1(\gamma,N) - \hat{\alpha}_{IS}}{\alpha_1(\gamma,N)}\right|, \tag{6}$$

wihere $\hat{\alpha}_{IS}$ is the IS estimator of $\alpha_1(\gamma, N)$ based on M i.i.d. realizations sampled according to the PDF in (5). The parameter δ is then chosen to control the bias term in the above equation, that is the first right-hand side of the above inequality. The second term on the right-hand side is the statistical relative error of estimating $\alpha_1(\gamma, N)$ by $\hat{\alpha}_{IS}$. From the Central Limit Theorem (CLT), this error term is approximately proportional to the coefficient of variation of $\hat{\alpha}_{IS}$.

To achieve a global relative error of order ϵ, it is sufficient to bound the two error terms, i.e., the statistical relative error and the relative bias, by $\epsilon/2$. Hence, the value of δ is selected such that the following inequality holds

$$0 \leq \frac{\alpha(\gamma, N) - \alpha_1(\gamma, N)}{\alpha(\gamma, N)} \leq \epsilon/2. \tag{7}$$

The following lemma provides the relation between δ and ϵ such that (7) is fulfilled.

Lemma 1. The following expression of $\delta(\epsilon, N, \gamma)$

$$\delta(\epsilon, N, \gamma) = \frac{N}{\gamma} \exp\left(\Phi^{-1}\left(\frac{\epsilon}{2N} \frac{(\Phi(\log(\gamma/N)))^N}{(\Phi(\log(\gamma)))^{N-1}}\right)\right), \tag{8}$$

where $\Phi(\cdot)$ is the CDF of the standard Normal distribution, ensures that (7) holds.

Proof. We first write that

$$\begin{aligned}
&\alpha(\gamma, N) - \alpha_1(\gamma, N) \\
&= \mathbb{P}_{h_{\mathbf{X}}}\left(\{\sum_{i=1}^{N} X_i \leq \gamma\} \cap \cup_{i=1}^{N}\{X_i \leq \delta\gamma/N\}\right) \\
&\leq \mathbb{P}_{h_{\mathbf{X}}}(\cup_{i=1}^{N}\{X_i \leq \delta\gamma/N\} \cap \cap_{i=1}^{N}\{X_i \leq \gamma\}) \\
&\leq \sum_{i=1}^{N} \mathbb{P}_{h_{\mathbf{X}}}\left(\{X_i \leq \delta\gamma/N\} \cap \cap_{j\neq i}\{X_i \leq \gamma\}\right) \\
&= N\mathbb{P}_{h_{\mathbf{X}}}\left(X_1 \leq \delta\gamma/N, X_2 \leq \gamma, \cdots, X_N \leq \gamma\right) \\
&= N\Phi(\log(\delta\gamma/N))\left(\Phi(\log(\gamma))\right)^{N-1}.
\end{aligned} \tag{9}$$

On the other hand, we have

$$\alpha(\gamma, N) \geq \left(\Phi(\log(\gamma/N))\right)^N.$$

Therefore, we get

$$\frac{\alpha(\gamma, N) - \alpha_1(\gamma, N)}{\alpha(\gamma, N)} \leq N \frac{\Phi\left(\log(\delta\gamma/N)\right)\left(\Phi(\log(\gamma))\right)^{N-1}}{\left(\Phi(\log(\gamma/N))\right)^N}.$$

By equating the right-hand side of the above inequality with $\epsilon/2$, we obtain

$$\delta(\epsilon, N, \gamma) = \frac{N}{\gamma} \exp\left(\Phi^{-1}\left(\frac{\epsilon}{2N} \frac{(\Phi(\log(\gamma/N)))^N}{(\Phi(\log(\gamma)))^{N-1}}\right)\right),$$

and hence the proof is concluded. □

4.3.2 The Gamma family as an IS PDF

when we consider a sufficiently small value of δ in the above analysis, we observe from the expression of the IS PDF in (5) that the proposed estimator with the change of measure in (5) achieves approximately the same performance as the Gamma IS PDF with shape parameter equal to 2. This suggests investigating whether the Gamma family can achieve further variance reduction with respect to the approach in the previous subsection. Note that the advantage of using the Gamma family as IS PDFs compared to the approach in the previous subsection is that the estimator is unbiased. Recall that the Gamma PDF is given by

$$\tilde{f}_X(x) = \frac{x^{k-1}\exp(-x/\theta)}{\Gamma(k)\theta^k},\ x > 0, \tag{10}$$

wihere $\theta > 0$ and $k > 0$ are the scale and shape parameters. The value of θ is chosen to be equal to $\theta = \frac{\gamma}{Nk}$ to ensure that the expected value of each of the X_i's, $i = 1, 2, \cdots, N$, under the PDF $\tilde{f}_X(\cdot)$ is equal to $\frac{\gamma}{N}$. The likelihood ratio is then given by

$$\begin{aligned}&\mathcal{L}(x_1, x_2, \cdots, x_N)\\ &= \frac{(\Gamma(k)\theta^k)^N \exp(\frac{\sum_{i=1}^N x_i}{\theta} - \frac{1}{2}\sum_{i=1}^N (\log(x_i))^2)}{\prod_{i=1}^N x_i^k (\sqrt{2\pi})^N}.\end{aligned}$$

The second moment of the IS estimator is bounded by

$$\begin{aligned}&\mathbb{E}_{\tilde{h}_{\mathbf{X}}}\left[\mathcal{L}^2(X_1, X_2, \cdots, X_N)\mathbf{1}_{(\sum_{i=1}^N X_i \leq \gamma)}\right]\\ &\leq (\frac{\Gamma(k)\theta^k}{\sqrt{2\pi}})^{2N} \exp(\frac{2\gamma}{\theta})\\ &\times \mathbb{E}_{\tilde{h}_{\mathbf{X}}}\left[\exp(-\sum_{i=1}^N (\log(x_i))^2 - 2k\sum_{i=1}^N \log(x_i))\right]\\ &\leq (\frac{\Gamma(k)(\frac{\gamma}{Nk})^k}{\sqrt{2\pi}})^{2N} \exp(2kN + k^2 N).\end{aligned}$$

The last upper bound is found by maximizing the function $x \to -(\log(x))^2 - 2k\log(x)$ for $x > 0$. Next, using Stirling's formula for the gamma function

$\Gamma(k) = \sqrt{2\pi} k^{k-\frac{1}{2}} \exp(-k)(1 + \mathcal{O}(\frac{1}{k}))$, we get

$$\begin{aligned}
&\mathbb{E}_{\tilde{h}_{\mathbf{x}}}\left[\mathcal{L}^2(X_1, X_2, \cdots, X_N)\mathbf{1}_{(\sum_{i=1}^N X_i \leq \gamma)}\right] \\
&\lesssim C k^{-N}\left(\frac{\gamma}{N}\right)^{2Nk} \exp(k^2 N) \\
&= C \exp(N(k^2 - 2k\log(N/\gamma) - \log(k)))
\end{aligned}$$

where C is a constant. Next, the value of k is chosen such that it minimizes the above right-hand side term. The solution of this minimization problem is given as follows:

$$k^* = \frac{1}{2}\left(\log(\frac{N}{\gamma}) + \sqrt{(\log(\frac{N}{\gamma}))^2 + 2}\right). \tag{11}$$

Note that when N is large and/or γ is small, the value of k^* satisfies $k^* \sim \log(\frac{N}{\gamma})$.

5 Numerical results

In this section, we show some selected numerical results to compare the performance of the proposed estimators compared to some of the existing estimators. We consider three scenarios depending on the distribution of X_i, $i = 1, 2, \cdots, N$: the Weibull, the Gamma-Gamma, and the Log-normal distributions. Note that the proposed approach is not restricted to these three distributions (see Table I for a non-exhaustive list of distributions that can be handled).

We define the squared coefficient of variation of an unbiased estimator $\hat{\alpha}(\gamma, N)$ of $\alpha(\gamma, N)$ as

$$SCV(\hat{\alpha}(\gamma, N)) = \frac{\mathrm{var}\left[\hat{\alpha}(\gamma, N)\right]}{\alpha^2(\gamma, N)}. \tag{12}$$

Note that, from the CLT, the number of required samples to meet ϵ statistical relative error with 95% confidence is equal to $(1.96)^2 SCV(\hat{\alpha}(\gamma, N))/\epsilon^2$. Therefore, when we compare two estimators, the one with the smaller squared coefficient of variation exhibits better performance than the other.

5.1 Weibull Case

In this section, we assume that X_i, $i = 1, 2, \cdots, N$, are distributed according to the Weibull distribution whose PDF is given in Table I. The comparison is made with respect to the second IS approach of [7] that is based on using the hazard rate twisting (HRT). In Figure 1 and Figure 2, we plot the squared coefficient of variations given by the HRT technique and the proposed approach for two different values of the shape parameter: $k = 1.5$ and $k = 0.5$, respectively. The value of $\alpha(\gamma, N)$ ranges approximately from 10^{-20} to 10^{-6} (respectively from

10^{-16} to 10^{-6}) using the system's parameters of Figure 1 (respectively of Figure 2). These figures show that the proposed approach clearly outperforms the one based on HRT. For instance, when $k = 1.5$, $\lambda = 1$, $\gamma = 0.5$, and $N = 12$, the proposed approach is approximately 270 times more efficient than the one based on HRT. More specifically, to meet the same accuracy, the number of samples needed by the approach based on HRT should be approximately 270 times the number of samples needed by the proposed approach.

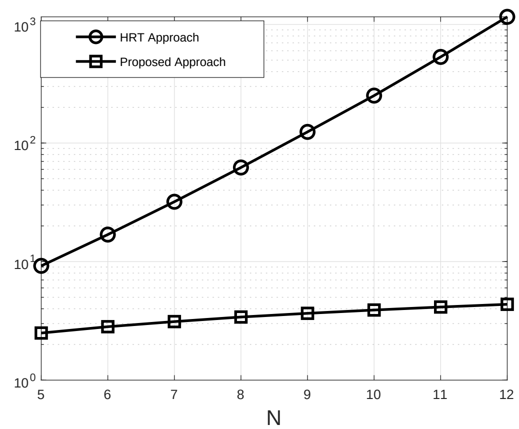

Figure 1: Squared coefficient of variation as a function of N where X_i are i.i.d. Weibull RVs with rate $\lambda = 1$, $k = 1.5$, and $\gamma = 0.5$.

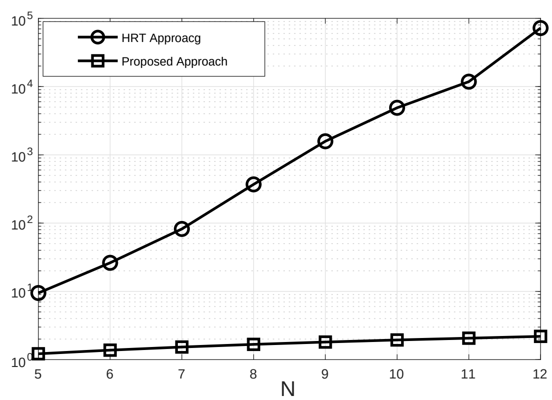

Figure 2: Squared coefficient of variation as a function of N where X_i are i.i.d. Weibull RVs with rate $\lambda = 1$, $k = 0.5$, and $\gamma = 0.01$.

In the next experiment, we aim to compare the proposed approach with the HRT one when N is fixed and γ decreases. In Figure 3, we compare the efficiency of both approaches in terms of squared coefficient of variations plotted as a function of γ for two scenarios depending on the value of N ($N = 8$ and $N = 10$). In this case, the value of $\alpha(\gamma, N)$ ranges approximately from 10^{-16} to 10^{-6} for $N = 8$ and from 10^{-22} to 10^{-8} for $N = 10$. We observe a clear outperformance of the proposed approach compared to the one based on using HRT for both values of N. While the HRT approach was proved in [7] to achieve the bounded relative error property with respect to γ and for a fixed value of N, it is clear from Figure 3 that the asymptotic bound increases substantially with respect to N, and hence the performance of the HRT approach is dramatically affected by increasing N. On the other hand, we observe that increasing the value of N has a minor effect on the efficiency of the proposed approach, i.e., the squared coefficient of variation is approximately unchanged for both values of N and for the considered range of γ. This numerical observation suggests to conclude

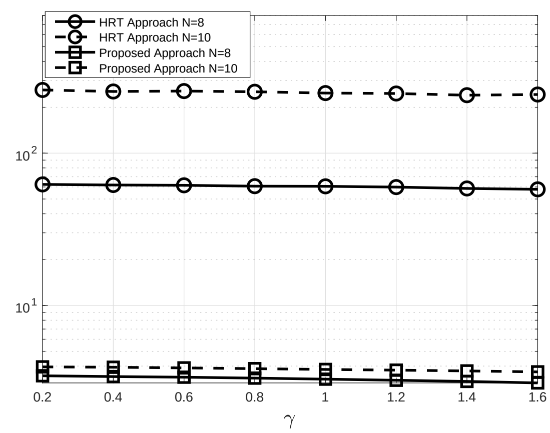

Figure 3: Squared coefficient of variation as a function of γ where X_i are i.i.d. Weibull RVs with rate $\lambda = 1$, $k = 1.5$.

that the proposed approach satisfies the bounded relative error property with an asymptotic bound that increases with a very slow rate, compared to the one given by the HRT approach, as we increase N. For illustration, the proposed approach is approximately 18 (respectively 64) times more efficient than the HRT one when $N = 8$ (respectively $N = 10$) and $\gamma = 0.2$. Note that the previous observations are valid independently of the value of $\alpha(\gamma, N)$ (see Figure 3, where the squared coefficient of variation is approximately constant for a fixed value of N and for the considered range of γ). This experiment and the numerical results in Figures 1 and 2 validate the ability of the proposed approach to deliver a very accurate and efficient estimate of $\alpha(\gamma, N)$ when N increases and/or γ decreases.

5.2 Gamma-Gamma Case

The Gamma-Gamma distribution is used for various challenging applications in wireless communications. For instance, it exhibited a good fit to experimental data and was used to model wireless radio-frequency channels [24] and to model atmospheric turbulences in free-space optical communication systems [16]. The PDF of X_i is given in Table I. In Figure 4, we compare the proposed approach

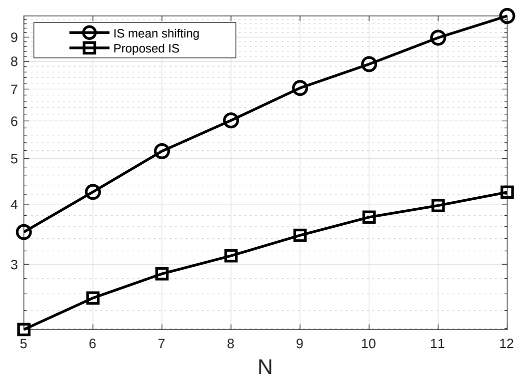

Figure 4: Squared coefficient of variation as a function of N where X_i are i.i.d. Gamma-Gammal RVs with $m = 4$, $k = 1.7$, $\Omega = 1$, and $\gamma = 0.5$.

with the one in [6] by plotting the corresponding squared coefficient of variations as a function of N and for a fixed value of γ. Note that in [6], the proposed IS PDF is simply another Gamma-Gamma PDF with shifted mean. We call this method the IS-based mean-shifted approach. The range of the quantity of interest $\alpha(\gamma, N)$ is approximately from 10^{-18} to 10^{-5}. We observe that the proposed estimator outperforms the one in [6]. Also, we observe that the outperformance of the proposed estimator compared to the one based on mean shifting increases as we increase N. Moreover, we should note here that the cost per sample (in terms of CPU time) of the approach in [6] is twice the cost of the proposed approach. This is because a Gamma-Gamma RV is generated by the product of two independent Gamma RVs, see [11]. For illustration, we observe from Figure 4 that when $N = 12$, the proposed approach is approximately 2.5 times (five times if we include the computing time in the comparison) more efficient than the one of [6].

5.3 Log-normal Case

The Log-normal distribution can be used to model several types of attenuation including shadowing [26], and weak-to-moderate turbulence channels in free-space optical communications [2]. The standard Log-normal PDF (the associated Gaussian RV has zero mean and unit variance) is given by

$$f_X(x) = \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{(\log(x))^2}{2}\right),\ x > 0.$$

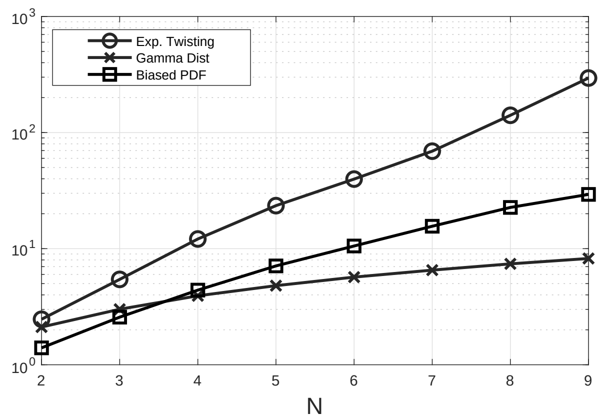

Figure 5: Squared coefficient of variation as a function of N where X_i are i.i.d. standard Log-normal RVs with $\gamma = 0.5$, and $\epsilon = 0.05$.

Figure 5 shows the squared coefficient of variation given by the exponential twisting [3], and the two proposed approaches, i.e., the one based on the biased estimator and the other based on using the Gamma distribution as an IS PDF. The value of $\alpha(\gamma, N)$ ranges approximately from 10^{-20} to 10^{-2}. For the considered range of N, we observe that out of these three approaches, it is the one using the Gamma distribution as an IS PDF that outperforms the others. When $N = 9$, it is approximately 30 times more efficient than the one based on exponential twisting. In addition to the efficiency in terms of number of samples, it is worth recalling that the exponential twisting technique developed in [3] is computationally expensive in terms of computing time compared to the proposed approaches. Moreover, Figure 5 also shows that the approach based on the biased estimator achieves better performances than the one based on exponential twisting. It is important to mention here that, for the comparison to be fair, the squared coefficient of variation of the biased estimator should be multiplied by 4. This follows from the error analysis in (6), in which the statistical relative error should be bounded by $\epsilon/2$, where ϵ is the required global relative error.

In Figure 6, we plot the squared coefficient of variations given by the three approaches as a function of γ and for two different values of N ($N = 8$ and $N = 10$). The quantity of interest $\alpha(\gamma, N)$ ranges approximately from 10^{-15} to 10^{-6} for $N = 8$ and from 10^{-21} to 10^{-9} for $N = 10$. We observe that the approach based on using the Gamma distribution as an IS PDF clearly asymptotically outperforms the two other approaches. For both values of N, the outperformance increases as we decrease γ. Moreover, the biased estimator exhibits better performances than the exponential twisting one for both values of N and for the considered range of γ values. Furthermore, increasing N has a considerable negative effect on the performances of the exponential twisting and the biased IS-based approaches. On the other hand, Figure 6 shows that increasing N does not largely effect the performance of the IS estimator based on the use of the Gamma distribution as an IS PDF. For illustration, the approach based on using the Gamma distribution as an IS PDF is approximately 15 times (respectively 35) more efficient that the exponential twisting one when $N = 8$ (respectively $N = 10$) and $\gamma = 0.6$.

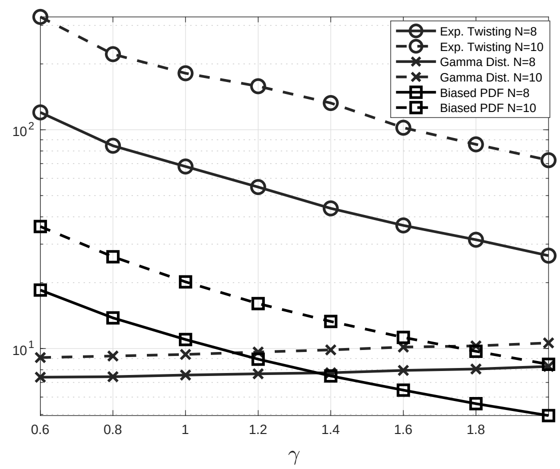

Figure 6: Squared coefficient of variation as a function of γ where X_i are i.i.d. standard Log-normal RVs with $\epsilon = 0.05$.

It is important to mention that the outperformance of the estimator based on using the Gamma distribution as an IS PDF over the one based on using the biased estimator is expected. As it was mentioned above, the latter approach gives approximately the same performance as the Gamma distribution with shape parameter equal to 2 while the former one uses the Gamma distribution as an IS PDF with an optimized shape parameter (the shape parameter was chosen to minimize an upper bound of the second moment of the proposed estimator, see the expression of k^* in (11)).

All of the above comparisons have been carried out in terms of the number of sampled needed to meet a fixed accuracy requirement. In order to include the computing time in our comparison, we define the Work Normalized Relative Variance (WNRV) metric of an unbiased estimator $\hat{\alpha}(\gamma, N)$ of $\alpha(\gamma, N)$ as follows (see [9]):

$$\begin{aligned}&\text{WNRV}(\hat{\alpha}(\gamma, N))\\&= \frac{SCV(\hat{\alpha}(\gamma, N))}{M} \times \text{computing time in seconds.}\end{aligned} \quad (13)$$

The computing time is the time in seconds needed to get an estimator of $\alpha(\gamma, N)$ using M i.i.d. samples of $\hat{\alpha}(\gamma, N)$. When comparing two estimators, the one that exhibits less WNRV is more efficient than the other estimator. More precisely, an estimator is efficient in terms of WNRV than another estimator means that it achieves less relative error for a given computational budget, or equivalently it needs less computing time to achieve a fixed relative error. Using the same setting as in Figure 6, we plot in Figure 7 the WNRV metric as a function of γ for two scenarios depending on the value of N ($N = 8$ and $N = 10$). We

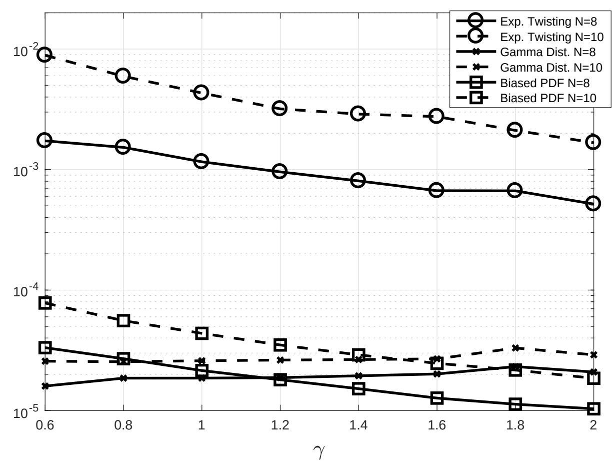

Figure 7: WNRV as a function of γ where X_i are i.i.d. standard Log-normal RVs with $\epsilon = 0.05$.

observe that as $\alpha(\gamma, N)$ is getting smaller, it is the approach based on using the Gamma PDF as an IS PDF that outperforms the two other approaches in terms of WNRV (the efficiency increases as the event becomes rarer). It is worth recalling that the WNRV of the approach based on biased PDF should be multiplied by 4 in order for the analysis to be fair (this follows from the error analysis that was performed in section 4.3.1). Moreover, Figure 7 shows that, in addition to reducing the variance, as shown in Figure 6, the approach based on using the Gamma IS PDF also reduces the computing time compared to the one using the exponential twisting technique. To see that, for $N = 10$ and $\gamma = 0.6$, the approach based on using the Gamma IS PDF is approximately 35 times (respectively 340 times) more efficient than the one based on exponential twisting when using the squared coefficient of variation metric (respectively the

WNRV metric). More specifically, the Gamma based IS approach approximately reduces the computing time by a factor of 10 with respect to the exponential twisting approach.

6 Conclusion

We developed efficient importance sampling estimators to estimate the rare event probabilities corresponding to the left-tail of the cumulative distribution function of large sums of nonnegative independent and identically distributed random variables. The proposed estimators achieve at least the same performance as the exponential twisting technique. The main conclusion is that the Gamma PDF with suitably chosen parameters achieves for most of the well-practical distributions substantial amount of variance reduction, and at the same time avoids the restrictive limitations when using the exponential twisting technique. The numerical results validate the efficiency of the proposed approach in being able to accurately and efficiently estimate the quantity of interest in the rare event regime corresponding to large N and/or small γ.

References

[1] M.-S. Alouini, N. Ben Rached, A. Kammoun, and R. Tempone. On the efficient simulation of the left-tail of the sum of correlated Log-normal variates. *Monte Carlo Methods and Applications*, Mar. 2018.

[2] I. S. Ansari, M.-S. Alouini, and J. Cheng. On the capacity of FSO links under Lognormal and Rician-Lognormal turbulences. In *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, pages 1–6, 2014.

[3] Søren Asmussen, Jens Ledet Jensen, and Leonardo Rojas-Nandayapa. Exponential family techniques for the lognormal left tail. *Scandinavian Journal of Statistics*, 43(3):774–787, Sep. 2016.

[4] Søren Asmussen and Peter W. Glynn. *Stochastic simulation : Algorithms and Analysis*. Stochastic modelling and applied probability. Springer, New York, 2007.

[5] N. C. Beaulieu and G. Luan. Improving simulation of lognormal sum distributions with hyperspace replication. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–7, 2019.

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

[7] N. Ben Rached, A. Kammoun, M.-S. Alouini, and R. Tempone. Unified importance sampling schemes for efficient simulation of outage capacity over generalized fading channels. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):376–388, Mar. 2016.

[8] Nadhir Ben Rached, Fatma Benkhelifa, Abla Kammoun, Mohamed-Slim Alouini, and Raul Tempone. On the generalization of the hazard rate twisting-based simulation approach. *Statistics and Computing*, Nov. 2016.

[9] Nadhir Ben Rached, Zdravko I. Botev, Abla Kammoun, Mohamed-Slim Alouini, and Raul Tempone. On the sum of order statistics and applications to wireless communication systems performances. *IEEE Transactions on Wireless Communications*, 17(11):7801–7813, November 2018.

[10] Zdravko I. Botev, Robert Salomone, and Daniel Mackinlay. Fast and accurate computation of the distribution of sums of dependent log-normals. *Annals of Operations Research*, pages 1–28.

[11] N. D. Chatzidiamantis, G. K. Karagiannidis, and D. S. Michalopoulos. On the distribution of the sum of Gamma-Gamma variates and application in MIMO optical wireless systems. In *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, pages 1–6, 2009.

[12] D. B. Da Costa and M. D. Yacoub. Accurate approximations to the sum of generalized random variables and applications in the performance analysis of diversity systems. *IEEE Transactions on Communications*, 57(5):1271–1274, May. 2009.

[13] N. Y. Ermolova. Moment generating functions of the generalized $\eta - \mu$ and $\kappa - \mu$ distributions and their applications to performance evaluations of communication systems. *IEEE Communications Letters*, 12(7):502–504, Jul. 2008.

[14] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products.* Elsevier/Academic Press, Amsterdam, seventh edition, 2007.

[15] Archil Gulisashvili and Peter Tankov. Tail behavior of sums and differences of Log-normal random variables. *Bernoulli*, 22(1):444–493, Feb. 2016.

[16] M. Hajji and F. E. Bouanani. Performance analysis of mixed Weibull and Gamma-Gamma dual-hop RF/FSO transmission systems. In *2017 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, pages 1–5, 2017.

[17] J. Hu and N. C. Beaulieu. Accurate closed-form approximations to Ricean sum distributions and densities. *IEEE Communications Letters*, 9(2):133–135, Feb. 2005.

[18] Sandeep Juneja and Perwez Shahabuddin. Simulating heavy tailed processes using delayed hazard rate twisting. *ACM Trans. Model. Comput. Simul.*, 12(2):94–118, Apr. 2002.

[19] Dirk P. Kroese, Thomas Taimre, and Zdravko I. Botev. *Handbook of Monte Carlo methods*. Wiley, N.J, 2011.

[20] J. A. Lopez-Salcedo. Simple closed-form approximation to Ricean sum distributions. *IEEE Signal Processing Letters*, 16(3):153–155, Mar. 2009.

[21] M. D. Renzo, F. Graziosi, and F. Santucci. Further results on the approximation of Log-normal power sum via Pearson type IV distribution: a general formula for log-moments computation. *IEEE Transactions on Communications*, 57(4):893–898, Apr. 2009.

[22] Ad Ridder and Reuven Rubinstein. Minimum cross-entropy methods for rare-event simulation. *SIMULATION*, 83(11):769–784, 2007.

[23] G. Rubino and B. Tuffin. *Rare Event Simulation using Monte Carlo Methods*. Wiley, 2009.

[24] P. M. Shankar. Error rates in generalized shadowed fading channels. *Wireless Personal Communications*, 28:233–238, Feb. 2004.

[25] Marvin Kenneth Simon and Mohamed-Slim Alouini. *Digital Communication over Fading Channels*. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, Hoboken, N.J, 2nd ed edition, 2005.

[26] Tjeng Thiang Tjhung, Chin Choy Chai, and Xiaodai Dong. Outage probability for Lognormal-shadowed Rician channels. *IEEE Transactions on Vehicular Technology*, 46(2):400–407, 1997.

[27] Z. Xiao, B. Zhu, J. Cheng, and Y. Wang. Outage probability bounds of EGC over dual-branch non-identically distributed independent Lognormal fading channels with optimized parameters. *IEEE Transactions on Vehicular Technology*, 68(8):8232–8237, Aug. 2019.

[28] B. Zhu and J. Cheng. Asymptotic outage analysis on dual-branch diversity receptions over non-identically distributed correlated Lognormal channels. *IEEE Transactions on Communications*, 67(10):7126–7138, Oct. 2019.