24th International Congress on Modelling and Simulation, Sydney, NSW, Australia, 5 to 10 December 2021
mssanz.org.au/modsim2021

Estimating tail probabilities of random sums of scale mixture of phase-type random variables

H. Yao [a] **and T. Taimre** [a]

[a] *School of Mathematics and Physics, The University of Queensland, Brisbane QLD 4072, AUSTRALIA*

Email: hui.yao@uq.net.au

Abstract: We consider the problem of estimating tail probabilities of random sums of scale mixture of phase-type distributions — a class of distributions corresponding to random variables which can be represented as a product of a non-negative but otherwise arbitrary random variable with a classical phase-type random variable. Our motivation arises from applications in risk, queueing problems for estimating ruin probabilities, and waiting time distributions respectively. Classical rare-event simulation algorithms cannot be implemented in this setting because these methods typically rely on the availability of the cumulative distribution function or the moment generating function, but these are difficult to compute or not even available for the class of scale mixture of phase-type distributions. In this paper, we address these issues and propose alternative simulation methods for estimating tail probabilities of random sums of scale mixture of phase-type distributions; our algorithm combines importance sampling and conditional Monte Carlo methods. The empirical performance of the method suggested is explored via numerical experimentation.

Keywords: *Asmussen–Kroese Estimator, Conditional Monte Carlo, importance sampling, scale mixture of phase-type distribution*

1 INTRODUCTION

Tail probabilities of random sums have attracted the interest of researchers for decades, as they are important quantities in many fields; for instance, in insurance risk theory, these quantities correspond to the ruin probabilities associated with certain initial capital, and the random sum is a geometric sum of ladder heights (integrated tails of the claim sizes) in a claim surplus process, see for example Asmussen and Albrecher [2010]. In queueing theory, stable single server Markovian queues with service times have a geometric length in equilibrium, and these quantities correspond to the probabilities that an arriving customer must wait longer than a certain time, see for example Asmussen [2003].

Estimating tail probabilities of random sums has been extensively investigated both for light- and heavy-tailed summands. Classical methods, including large deviations, saddlepoint approximations, and exponential change of measure, are most commonly used. However, the application of these methods requires the existence of the moment generating function (MGF) of the summand, so are limited to light-tailed cases. In the heavy-tailed setting, subexponential theory provides asymptotic approximations but these offer poor accuracy for moderately large values (Foss et al. [2013]). From a rare-event simulation perspective, there are a few efficient estimators proposed in the literature (Asmussen et al. [2000]; Juneja and Shahabuddin [2002]; Asmussen and Kroese [2006]). Simulation methods for the heavy-tailed case often require the cumulative distribution function (CDF) or probability density function (PDF) of the summand so these are frequently and easily implemented.

In this paper we consider the problem of efficiently estimating the quantity

$$\ell(u) = \mathbb{P}(S_N > u)\,, \tag{1}$$

where $S_N = Z_1 + \cdots + Z_N$, for large u. Specifically, N is a discrete light-tailed random variable supported over positive integers and $\{Z_i\}_{i\in\mathbb{N}}$ is independent of N and forms a sequence of independent, non-negative and identically distributed random variables having stochastic representation $Z_i = W_i\,X_i$, where the random variables W_i and X_i are non-negative and independent of each other.

In our setting, the sequence $\{Z_i\}_{i\in\mathbb{N}}$ belongs to the class of W-mixture of PH-distributions, which can be represented as a product $W_i\,X_i$, where each W_i is a non-negative but otherwise arbitrary random variable and X_i is a classical PH-distributed random variable. The concept of W-mixture of PH-distributions can be found in Keilson and Steutel [1974]. Bladt et al. [2015] proposed a subclass of W-mixture of PH-distributions to approximate any heavy-tailed distribution. Such a class is very attractive in stochastic modeling because it inherits many important properties of PH-distributions — including being dense in the class of non-negative distributions and closed under finite convolutions — while it also circumvents the problem that individual PH-distributions are light-tailed, implying that the tail behavior of a heavy-tailed distribution cannot be captured correctly by classical PH-distributions alone. Rojas-Nandayapa and Xie [2018] showed that a distribution in the W-mixture of PH-distributions class is heavy-tailed if and only if the random variable W_i has unbounded support.

The CDF and PDF of a distribution in the W-mixture of PH-distributions class are both available in closed form but these are given in terms of infinite dimensional matrices. The problem of approximating tail probabilities of scale mixture of PH distributed random variables is not easily tractable from a computational perspective. Bladt et al. [2015] addressed this issue and proposed a methodology which can be easily adapted to compute geometric random sums of scale mixture of PH-distributions. Their approach is based on an infinite series representation of the tail probability of the geometric sum which can be computed to any desired precision at the cost of increased computational effort. In this paper we explore an obvious alternative approach: rare-event estimation.

More precisely, we propose and analyze simulation methodologies to approximate the tail probabilities of a random sum of scale mixture of PH distributed random variables. We remark that since the CDF and PDF of a distribution in the W-mixture of PH-distributions class is effectively not available, most algorithms for the heavy-tailed setting discussed above cannot be implemented. Our approach is to use conditioning arguments and adapt the Asmussen–Kroese estimator proposed in Asmussen and Kroese [2006]. The Asmussen–Kroese approach is usually not directly implementable in our setting because the CDF of the product $W_i\,X_i$ is typically not available. We address this issue using a simple conditioning argument on the PH random variable X_i in this paper, thereby simply simulating the PH-distribution and using the CDF of W_i. Moreover, when the scaling random variables are light-tailed, we explore the use of importance sampling (IS) on the maximum of the random variables that are conditioned on.

The remainder of the paper is organized as follows. In Section 2 we provide background knowledge on PH distributions and scale mixture of PH distributions. In Section 3 we introduce the proposed algorithm. In Section 4 we present the empirical results for several examples. Section 5 provides some concluding remarks and an outlook to future work.

2 PRELIMINARIES

In this section we provide a general overview of classical PH distributions and scale mixture of PH distributions.

2.1 Phase-type Distributions and Properties

PH-distributions have been used in stochastic modeling since being introduced in Neuts [1975]. Apart from being mathematically tractable, PH-distributions have the additional appealing feature of being dense in the class of non-negative distributions. That is, for any distribution on the positive real axis there exists a sequence of PH-distributions which converges weakly to the target distribution (see Asmussen [2003] for details). In other words, PH-distributions may approximate arbitrarily closely any distribution with support on $[0, \infty)$.

In order to define a PH-distribution, we first consider a continuous-time Markov chain (CTMC) $\{Y(t),\ t \geq 0\}$ on the finite state space $E = \{1, 2, \ldots, p\} \cup \{\triangle\}$, where states $1, 2, \ldots, p$ are transient and state $\triangle$ is absorbing. Further, let the process have an initial probability of starting in any of the p transient phases given by the $1 \times p$ probability vector $\boldsymbol{\alpha}$, with $\alpha_i \geq 0$ and $\sum_{i=1}^{p} \alpha_i \leq 1$. Hence, the process $\{Y(t) : t \geq 0\}$ has an intensity matrix (or infinitesimal generator) $\mathbf{Q}$ of the form:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix},$$

where $\mathbf{T}$ is a $p \times p$ sub-intensity matrix of transition rates between the transient states, $\mathbf{t}$ is a $p \times 1$ vector of transition rates to the absorbing state, and $\mathbf{0}$ is a $1 \times p$ zero row vector.

The (continuous) PH-distribution is the distribution of time until absorption of $\{Y(t) : t \geq 0\}$. The 2-tuple $(\boldsymbol{\alpha}, \mathbf{T})$ completely specifies the PH-distribution, and is called a PH-representation. The CDF is given by $F(y) = 1 - \boldsymbol{\alpha} \exp(\mathbf{T}\, y)\mathbf{1},\ y \geq 0$, where $\mathbf{1}$ is a column vector with all ones.

Besides being dense in the non-negative distributions, the class of continuous PH-distributions forms the smallest family of distributions on $\mathbb{R}_+$ which contains the point mass at zero and all exponential distributions, and is closed under finite mixtures, convolutions, and infinite mixtures (among other interesting properties) (see Neuts [1975]).

2.2 Scale Mixture of Phase-type Distributions

A random variable Z of the form $Z := W \cdot X$ is scale mixture of PH distributed if $X \sim F$, where F is a PH-distribution, and $W \sim H$, where H is an arbitrary non-negative distribution. We call W the scaling random variable and H the scaling distribution. It follows that the CDF of Z can be written as the Mellin–Stieltjes convolution of the two non-negative distributions F and H:

$$B(z) = \int_0^\infty F(z/w)\,\mathrm{d}\,H(w) = \int_0^\infty H(z/x)\,\mathrm{d}\,F(x), \quad z \geq 0\,.$$

The integral expression above is available in closed form in very few isolated cases. Thus, we should rely on numerical integration or simulation methods for its computation.

Recall that a non-negative random variable X has a heavy-tailed distribution if and only if $\mathbb{E}(\mathrm{e}^{\theta X}) = \infty,\ \forall\theta > 0$, equivalently, if $\limsup_{x\to\infty} \mathbb{P}(X > x)\mathrm{e}^{\theta\, x} = \infty,\ \forall\theta > 0$. Otherwise, we say X is light-tailed.

The key motivation for considering the class of scale mixture of PH distributions over the class of PH is that the latter class forms a subclass of light-tailed distributions while a distribution in the former class with scaling random variable having unbounded support is heavy-tailed (Rojas-Nandayapa and Xie [2018]). Hence, the class of scale mixture of PH distributions turns out to be an appealing tractable class for approximating heavy-tailed distributions.

3 SIMULATION METHODS

In this section, we introduce our rare-event simulation estimators for $\ell(u)$. The key approach is combining an Asmussen–Kroese-type Algorithm with conditional Monte Carlo when scaling random variables are heavy-tailed distributed, and further with importance sampling when scaling random variables are light-tailed distributed. The method is inspired on using the tower property of expectation $\mathbb{E}\widehat{\ell}(u) = \mathbb{E}\mathbb{E}[\widehat{\ell}(u)|T]$, and so in practical terms, one should be able to simulate T and compute $\mathbb{E}[\widehat{\ell}(u)|T]$.

Asmussen and Kroese [2006] is efficient for estimating tail probabilities of random sums of heavy-tailed summands. Further study on the efficiency of Asmussen–Kroese estimator can be found in Hartinger and Kortschak [2009]. Such an algorithm is based on the "principle of the single large jump" — a property held by a subclass of heavy-tailed distributions called subexponential and which includes practically all common heavy-tailed distributions. The principle of the single large jump indicates that the most likely scenario in which a sum of subexponential random variables becomes large is because of a single summand taking a large value as opposed to the scenario where two or more summands take large values. The key idea is based on the following symmetry argument:

$$\mathbb{P}(S_N > u|N = n) = n\mathbb{P}(S_n > u, \max\{Z_i, i = 1, \ldots, n\} = Z_n) = n\mathbb{E}\overline{B}(Z^\star_{n-1} \vee (u - S_{n-1}))\,,$$

where $\overline{B}(\cdot)$ is the complementary CDF of Z_i, $Z^\star_{n-1} = \max\{Z_i, i = 1, \ldots, n-1\}$, $S_{n-1} = \sum_{i=1}^{n-1} Z_i$, and $x \vee y = \max(x, y)$.

In our setting, the summands $Z_i = W_i X_i$ are heavy-tailed whenever W_i has unbounded support (Rojas-Nandayapa and Xie [2018]) and so it is natural to consider this estimator here. Unfortunately, the Asmussen–Kroese approach is usually not directly implementable in our setting because the CDF of Z_i is typically unavailable. Instead, we consider a simple modification, by conditioning on a single random variable W or X. We further consider applying a change of measure to this single random variable when necessary. Conditioning on $N = n$, the Asmussen–Kroese estimator (without the control variate correction) for $\ell(u)$ takes the form

$$\widehat{\ell}_{\text{AK}}(u) = n\,\overline{B}\left(Z^\star_{n-1} \vee (u - S_{n-1})\right)\,. \tag{2}$$

In this paper, we simply condition on X (conditioning on W has been studied in Yao et al. [2016]), to arrive at

$$\widehat{\ell}_{\text{ConAK}}(u) = n\,\overline{H}_X\left(\frac{Z^\star_{n-1} \vee (u - S_{n-1})}{X}\right)\,, \tag{3}$$

where $\overline{H}_X(\cdot)$ is the complementary CDF of a scaling random variable.

We then consider improving the efficiency of our algorithm by implementing importance sampling over the distribution H when necessary, ensuring that the random sum after conditioning is equal in expectation to u. Inspired by popular methodologies drawn from light-tailed problems, we suggest the following algorithm.

3.1 Exponential Twisting

The exponential twisting method is asymptotically efficient for tail probabilities of random sums with light-tailed summands (Siegmund [1976]). The method is specified as follows:

Define an exponential family of PDFs $\{f_\theta, \theta \in \Theta\}$ based on the original PDF of the random variable X, f, via $f_\theta(x_i) = \frac{\mathrm{e}^{\theta x_i}}{M_{X_i}(\theta)} f(x_i) = \mathrm{e}^{\theta x_i - \ln M_{X_i}(\theta)} f(x_i)$, where $M_{X_i}(\theta) = \int \mathrm{e}^{\theta x_i} f(x_i)\mathrm{d}x_i$ is the MGF of X_i. The likelihood ratio of a single element associated with this change of measure is $\frac{f(X_i)}{f_\theta(X_i)} = \mathrm{e}^{-\theta X_i + \zeta(\theta)}$, where $\zeta(\theta) = \ln M_{X_i}(\theta)$ is the cumulant function of X_i. Then the twisted mean is $\mu_\theta = \mathbb{E}_\theta(X_i) = \zeta'(\theta)$ (Kroese et al. [2011]).

The selection of a proper twisting parameter θ is a key aspect in the implementation of this method. For dealing with the random sum (1), we select the twisting parameter θ such that the changed mean of the random sum is equal to the threshold, that is, $\mathbb{E}_\theta(S_n|N = n) = u$. Note that $\mathbb{E}_\theta(\sum_{i=1}^n Z_i|N = n) = n\,\mathbb{E}_\theta[W_i]\,\mathbb{E}[X_i]$. We only perform exponential twisting on the $\max\{Z_i, i = 1, \ldots, n\}$, so we choose the value θ which solves $\zeta'(\theta) = u/\mathbb{E}(W_i)$.

Combining the estimator (3) with the above importance sampling method for both light- and heavy-tailed scaling distributions, we then arrive at the following estimator for a single replicate:

$$\widehat{\ell}_{\text{ConAK+IS}}(u) = N\,\overline{H}_X\left(\frac{Z^\star_{N-1} \vee (u - S_{N-1})}{X}\right) \times \frac{f(X)}{f_\theta(X)}\,. \tag{4}$$

We further combine the estimators (3) and (4) with a control variate for N to improve the efficiency:

$$\widehat{\ell}_{\text{ConAK+CV}}(u) = N\,\overline{H}_X\left(\frac{Z^\star_{N-1} \vee (u - S_{N-1})}{X}\right) - (N - \mathbb{E}(N))\,\overline{H}_X\left(\frac{u}{X}\right), \tag{5}$$

$$\widehat{\ell}_{\text{ConAK+IS+CV}}(u) = N\overline{H}_X\left(\frac{Z^\star_{N-1} \vee (u - S_{N-1})}{X}\right) \times \frac{f(X)}{f_\theta(X)} - (N - \mathbb{E}(N))\,\overline{H}_X\left(\frac{u}{X}\right) \times \frac{f(X)}{f_\theta(X)}. \tag{6}$$

Combining conditional Asmussen–Kroese, importance sampling, and control variate is summarized in the following algorithm used to generate a single replicate:

Algorithm 1. *(Conditional Asmussen–Kroese, Importance Sampling and Control Variate Algorithm)*

1. *Generate* N.
2. *Generate* $Z_1, \ldots, Z_{N-1}$ *as* $Z_i = W_i\,X_i$ *with* $W_1, \ldots, W_{N-1} \overset{i.i.d}{\sim} H$ *and* $X_1, \ldots, X_{N-1} \overset{i.i.d}{\sim} F$.
3. *Compute* $S_{N-1} = \sum_{i=1}^{N-1} Z_i$*, find* $Z^\star_{N-1} \vee (u - S_{N-1})$.
4. *Generate* $X \sim F_\theta$*, with parameter* θ *solving* $\zeta'(\theta) = u/\mathbb{E}(W_i)$.
5. *Return* $\widehat{\ell}_{\mathit{ConAK+IS+CV}}(u) = N\overline{H}_X\left(\frac{Z^\star_{N-1} \vee (u-S_{N-1})}{X}\right) \times \frac{f(X)}{f_\theta(X)} - (N - \mathbb{E}(N))\,\overline{H}_X\left(\frac{u}{X}\right) \times \frac{f(X)}{f_\theta(X)}$.

4 NUMERICAL EXPERIMENTS

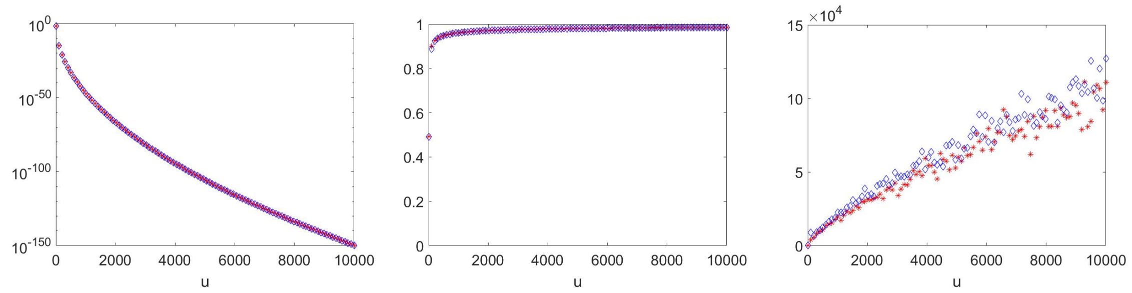
In this section, we provide three simple sets of numerical experiments to illustrate the proposed simulation methods. In all illustrative cases, we consider each of the W_i to have support on all of $[0, \infty)$ and each of the X_i to be Erlang distributed with shape parameter being α and rate parameter being β, as Erlang distributions play an important role in the class of PH-distributions (O'Cinneide [1991]). The class of generalized Erlang distributions is dense in the set of all probability distributions on the non-negative half-line. We remind the reader that in all cases, the resulting product $Z_i = W_i\,X_i$ is heavy-tailed.

Below we will explore the estimate $\widehat{\ell}_{\text{ConAK+IS+CV}}$ which is the conditional Asmussen–Kroese estimator with importance sampling on the maximum of summands and control variate for the number of summands, and $\widehat{\ell}_{\text{ConAK+CV}}$ which is the conditional Asmussen–Kroese estimator with control variate for the number of summands. We will include the estimate $\widehat{\ell}_{\text{ConAK+IS+CV}}$ (conditioning on scaling random variables W) from Yao et al. [2016] shown as blue diamonds in all the figures below for comparison.

Example 1 (Light-tailed Scaling: Exponential)**.** *Taking a sample size of* 10^5*,* $W_i \overset{i.i.d}{\sim}$ *Exp*(μ)*,* $X_i \overset{i.i.d}{\sim}$ *Erlang*(α, β) *and* $N \sim$ *Geo*(p) *with* $\mu = 1$*,* $\alpha = 1$*,* $\beta = 3$*, and* $p = 0.8$*. The corresponding estimate* $\widehat{\ell}_{\mathit{ConAK+IS+CV}}(u)$ *of* $\ell(u)$ *for* u *in the range* $[0, 10^4)$*, empirical logarithmic rates (by estimating* $\log(\mathbb{V}ar(\widehat{\ell}(u)))/2\log(\ell(u))$*) are shown in Figure 1b, and work-normalized relative variance (by estimating* $\tau\mathbb{V}ar(\widehat{\ell}(u))/\ell(u)^2$*, where* τ *is the simulation time) are shown in Figure 1c.*

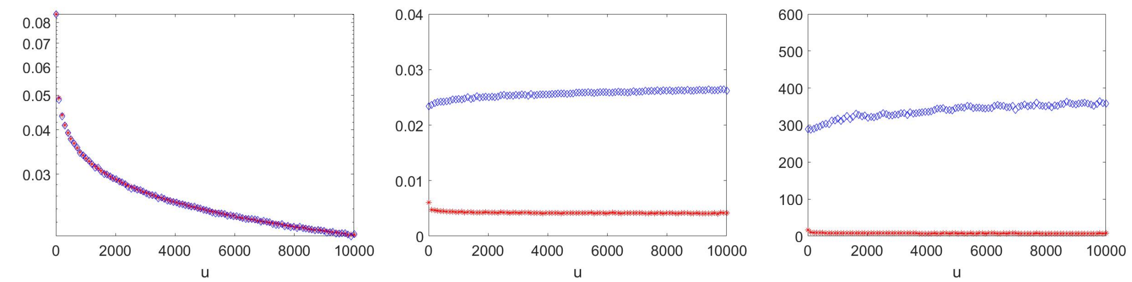
Example 2 (Heavy-tailed Scaling: Weibull)**.** *Taking a sample size of* 10^5*,* $W_i \overset{i.i.d}{\sim}$ *Weibull*(a, b)*,* $X_i \overset{i.i.d}{\sim}$ *Erlang*(α, β)*, and* $N \sim$ *Geo*(p) *with scale parameter* $a = 2$*, shape parameter* $b = 0.1$*,* $\alpha = 2$*,* $\beta = 3$*, and* $p = 0.8$*. The corresponding estimate* $\widehat{\ell}_{\mathit{ConAK+CV}}(u)$ *of* $\ell(u)$ *(conditioning on Erlang random variable) for* u *in the range* $[0, 10^4)$*, empirical relative error (by estimating* $\sqrt{\mathbb{V}ar(\widehat{\ell}(u))}/\ell(u)$*) are shown in Figure 2b, and work-normalized relative variance (by estimating* $\tau\mathbb{V}ar(\widehat{\ell}(u))/\ell(u)^2$*, where* τ *is the simulation time) are shown in Figure 2c.*

Example 3 (Heavy-tailed Scaling: Lognormal)**.** *Taking a sample size of* 10^5*,* $W_i \overset{i.i.d}{\sim}$ *LogN*(μ, σ^2)*,* $X_i \overset{i.i.d}{\sim}$ *Erlang*(α, β)*, and* $N \sim$ *Geo*(p) *with* $\mu = 2$*,* $\sigma = 3$*,* $\alpha = 2$*,* $\beta = 3$*, and* $p = 0.8$*. The corresponding estimate* $\widehat{\ell}_{\mathit{ConAK+CV}}(u)$ *of* $\ell(u)$ *(conditioning on Erlang random variable) for* u *in the range* $[0, 10^4)$*, empirical relative error (by estimating* $\sqrt{\mathbb{V}ar(\widehat{\ell}(u))}/\ell(u)$*) are shown in Figure 3b, and work-normalized relative variance (by estimating* $\tau\mathbb{V}ar(\widehat{\ell}(u))/\ell(u)^2$*, where* τ *is the simulation time) are shown in Figure 3c.*



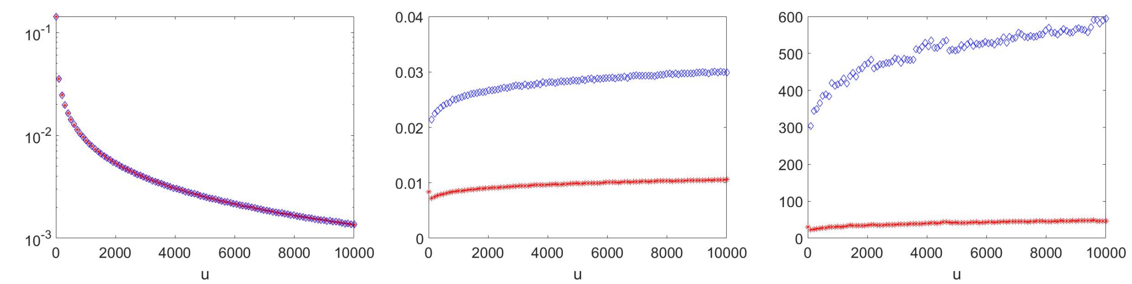
(a) Estimate of $\ell(u)$ as a function of u on a logarithmic scale.
(b) Empirical logarithmic rate as a function of u.
(c) Empirical work-normalized relative variance as a function of u.

Figure 1: The random variable N is geometrically distributed with success probability $p = 0.8$, $W_i \overset{\text{i.i.d}}{\sim} \text{Exp}(1)$, $X_i \overset{\text{i.i.d}}{\sim} \text{Erlang}(1, 3)$. Red Stars: ConAK+IS+CV (Conditioning and IS on Erlang). Blue Diamonds: ConAK+IS+CV (Conditioning and IS on Exponential).



(a) Estimate of $\ell(u)$ as a function of u on a logarithmic scale.
(b) Empirical relative error as a function of u.
(c) Empirical work-normalized relative variance as a function of u.

Figure 2: The random variable N is geometrically distributed with success probability $p = 0.8$, $W_i \overset{\text{i.i.d}}{\sim} \text{Weibull}(2, 0.1)$, $X_i \overset{\text{i.i.d}}{\sim} \text{Erlang}(2, 3)$. Red Stars: ConAK+CV (Conditioning on Erlang). Blue Diamonds: ConAK+IS+CV (Conditioning and IS on Weibull).



(a) Estimate of $\ell(u)$ as a function of u on a logarithmic scale.
(b) Empirical relative error as a function of u.
(c) Empirical work-normalized relative variance as a function of u.

Figure 3: The random variable N is geometrically distributed with success probability $p = 0.8$, $W_i \overset{\text{i.i.d}}{\sim} \text{LogN}(2, 9)$, $X_i \overset{\text{i.i.d}}{\sim} \text{Erlang}(2, 3)$. Red Stars: ConAK+CV (Conditioning on Erlang). Blue Diamonds: ConAK+IS+CV (Conditioning and IS on Lognormal).

5 DISCUSSION AND OUTLOOK

In this paper, we proposed straightforward simulation methods for estimating tail probabilities of random sums of scale mixture of PH distributed summands. When combining Asmussen–Kroese estimation with the conditional Monte carlo method, we can either condition on the scaling random variables or PH random variables. The given examples show that conditioning on the PH random variable performs better when the scaling random variables are heavy-tailed. However, when the scaling random variables are light-tailed, we found the estimates are not accurate for large u. Thus, we further exploit importance sampling applied to the maximum of the random variables that are conditioned on.

In the given examples, we take the distribution of N to be geometrically distributed. Since the method can perform poorly when the number of summands is large, we implemented an additional control variate for N to improve the accuracy of this estimator. We observe that in Example 1 the ConAK+IS+CV method appears to attain logarithmic efficiency, as the empirical logarithmic rates tend to one when u tends to infinity (note: the estimator is said to be logarithmically efficient if $\liminf_{u\to\infty} \frac{\log(\mathbb{V}\mathrm{ar}(\widehat{\ell}(u)))}{2\log(\ell(u))} \geq 1$), and conditioning on the scaling random variables and PH random variables perform similar. In Examples 2 and 3, the ConAK+CV and ConAK+IS+CV methods appear to attain bounded relative error, and the ConAK+CV method conditioning on the PH random variables attain lower relative error and work-normalized relative variance.

In addition to establishing theoretical guarantees around efficiency, an interesting avenue for further work is to develop effective simulation methods for this problem when there is structured dependence between the random variables.

ACKNOWLEDGMENTS

The authors acknowledge the valuable input of Dr Leonardo Rojas-Nandayapa on an earlier version of this manuscript.

REFERENCES

Asmussen, S. [2003], *Applied Probabilities and Queues*, 2nd edn, Springer-Verlag, New York.

Asmussen, S. and Albrecher, H. [2010], *Ruin Probabilities*, 2nd edn, World Scientific Publishing.

Asmussen, S., Binswanger, K. and Højgaard, B. [2000], 'Rare events simulation for heavy-tailed distributions', *Bernoulli* **6**(2), 303–322.

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

Bladt, M., Nielsen, B. F. and Samorodnitsky, G. [2015], 'Calculation of ruin probabilities for a dense class of heavy tailed distributions', *Scandinavian Actuarial Journal* **2015**(7), 573–591.

Foss, S., Korshunov, D. and Zachary, S. [2013], *An Introduction to Heavy-Tailed and Subexponential Distributions*, 2nd edn, Springer, New York.

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

Juneja, S. and Shahabuddin, P. [2002], 'Simulating heavy tailed processes using delayed hazard rate twisting', *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **12**(2), 94–118.

Keilson, J. and Steutel, F. [1974], 'Mixtures of distributions, moment inequalities and measures of exponentiality and normality', *The Annals of Probability* pp. 112–130.

Kroese, D. P., Taimre, T. and Botev, Z. I. [2011], *Handbook of Monte Carlo Methods*, 1st edn, John Wiley and Sons, Hoboken, New Jersey.

Neuts, M. F. [1975], *Probability Distributions of Phase Type*, Liber Amicorum Professor Emeritus H. Florin, Department of Mathematics, University of Louvain, Belgium, pp. 173–206.

O'Cinneide, C. A. [1991], 'Phase-type distributions and majorization', *The Annals of Applied Probability* **1**, 219–227.

Rojas-Nandayapa, L. and Xie, W. [2018], 'Asymptotic tail behaviour of phase-type scale mixture distributions', *Annals of Actuarial Science* **12**(2), 412–432.

Siegmund, D. [1976], 'Importance sampling in the Monte Carlo study of sequential tests', *The Annals of Statistics* **4**, 673–684.

Yao, H., Rojas-Nandayapa, L. and Taimre, T. [2016], Estimating tail probabilities of random sums of infinite mixtures of phase-type distributions, *in* '2016 Winter Simulation Conference (WSC)', pp. 347–358.