Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, and E. Zhou, eds.

ESTIMATING TAIL PROBABILITIES OF GEOMETRIC RANDOM SUMS OF INFINITE MIXTURES OF PHASE-TYPE DISTRIBUTIONS

Hui Yao

Leonardo Rojas-Nandayapa

Thomas Taimre

School of Mathematics and Physics
The University of Queensland
Brisbane QLD 4072, AUSTRALIA

ABSTRACT

We consider the problem of estimating tail probabilities of random sums of infinite mixtures of phase-type (IMPH) distributions — a class of distributions corresponding to random variables which can be represented as a product of an arbitrary random variable with a classical phase-type distribution. Our motivation arises from applications in risk and queueing problems. Classical rare-event simulation algorithms cannot be implemented in this setting because these typically rely on the availability of the cdf or the mgf, but these are difficult to compute or not even available for the class of IMPH distributions. In this paper, we address these issues and propose alternative simulation methodologies for estimating tail probabilities of random sums of IMPH distributions; our algorithms combine IS and conditional Monte Carlo methods. The empirical performance of each method suggested is explored via numerical experimentation.

1 INTRODUCTION

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

$$\ell(u) = \mathbb{P}(S_M > u) \qquad S_M = Z_1 + \cdots + Z_M\,, \tag{1}$$

for large u, where M is a discrete random variable supported over the positive integers and $\{Z_i,\ i = 1,2,\dots\}$ is independent of M and forms a sequence of independent, non-negative and identically distributed random variables having stochastic representation

$$Z_i = W_i X_i\,.$$

In the above, the random variables W_i and X_i are nonnegative and independent of each other. In this paper we will specialize to the case where the distribution of X_i belongs the class of phase-type (PH) distributions and W_i is a nonnegative but otherwise general random variable.

Tail probabilities of random sums are quantities of interest in many areas; for instance, in risk theory $\{Z_i\}$ corresponds to the distributions of the ladder heights (integrated tail of the claim sizes) in a Cramér–Lundberg model while M corresponds to number of ladder heights. Hence, the tail probability of the associated random sum corresponds to the ruin probability with initial capital u, see for example (Asmussen and Albrecher 2010). Stable single server Markovian queues with service times Z_i have a geometric length in equilibrium, so ℓ is the probability that an arriving customer must wait longer than u, see for example (Asmussen 2003).

The problem of approximating tail probabilities of random sums has been extensively explored both for light- and heavy-tailed summands. Large deviations, saddle point approximations, and exponential change of measure are among the classical methods most often used. These methods rely on the existence of the mgf of the claim sizes, so are limited to light-tailed cases. In the heavy-tailed case, subexponential theory provide asymptotic approximations but these offer poor accuracy for moderately large values of u. The later case has been successfully addressed from a rare-event simulation perspective (Embrechts and Veraverbeke 1982). Simulation methodologies for the heavy-tailed case often require just the cdf or pdf of the Z_i's so these are frequently readily implemented.

In our setting, the sequence of $\{Z_i\}$ belongs to the class of IMPH distributions, which can be represented as a product W_iX_i, where each W_i is a nonnegative but otherwise arbitrary random variable and X_i is a classical PH-distributed random variable. Recently, (Bladt, Nielsen, and Samorodnitsky 2014) proposed the class of IMPH distributions to approximate any heavy-tailed distribution. Such a new class is very attractive in stochastic modelling because it inherits many important properties of PH distributions — including being dense in the class of nonnegative 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. A distribution in the IMPH class is heavy-tailed if and only if the random variable W_i has unbounded support (Rojas-Nandayapa and Xie 2015).

The cdf and pdf of a distribution in the IMPH class are both available in closed form but these are given in terms of infinite dimensional matrices so the problem of approximating tail probabilities involving IMPH distributed random variables are not easily tractable from a computational perspective. Bladt *et al.* (Bladt, Nielsen, and Samorodnitsky 2014) addressed this issue and proposed a methodology which can be easily adapted to compute geometric sums of IMPH 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 alternative simulation methodologies to approximate the tail probabilities of a geometric sum of IMPH distributed random variables. We remark that since the cdf and pdf of a distribution in the IMPH class is effectively not available, most algorithms for the heavy-tailed setting discussed above are not implementable. Our approach is to use conditioning arguments and adapt and combine a variety of well-established rare-event simulation methodologies. The proposed algorithms are as follow:

1. Our first algorithm exploits the convolution closure property of PH-distributions. Conditional on the random variable M and the scaling variables W_i, the random sum of IMPH random variables reduces to a convolution of classical PH distributions. As a consequence of the closure property, the latter is just a classical PH random variable which can be easily simulated and whose tail probabilities can be readily obtained at a low computational cost.
2. Our second algorithm 'boosts' the first by using Importance Sampling (IS) over the scaling random variables W_i. While a distribution in the IMPH class is heavy-tailed if the scaling random variable W_i is unbounded, the random variable W_i itself is not necessarily heavy-tailed so we can apply light-tailed IS techniques to it. We explore exponential twisting for light-tailed scaling random variables and adapt hazard rate twisting for heavy-tailed scaling random variables.
3. Our third algorithm adapts the Asmussen–Kroese estimator (Asmussen and Kroese 2006) for scaling random variables with unbounded support. The Asmussen–Kroese approach is usually not directly implementable in our setting because the cdf of the product $W \cdot X$ is typically not available. We address this issue using a simple conditioning argument on the scaling random variable W, thereby simply simulating the phase-type distribution and using the cdf of W.

The remainder of the paper is organized as follows. In Section 2 we provide background knowledge on PH- and IMPH-distributions. In Section 3 we introduce and discuss the proposed algorithms. 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- and IMPH-distributions.

2.1 Phase-type Distributions and Properties

PH-distributions have been used in stochastic modelling 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 $E = \{1,2,\cdots,p\} \cup \{\triangle\}$, 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 α, with $\alpha_i \geq 0$ and $\sum_{i=1}^{p} \alpha_i \leq 1$. Hence, the process $\{Y(t)\}$ has an intensity matrix (or infinitesimal generator) $\mathbf{Q}$ of the form:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix}, \tag{2}$$

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 $(\alpha,\mathbf{T})$ completely specifies the PH-distribution, and is called a phase-type representation. The *cumulative distribution function* (cdf) is given by:

$$F(y) = 1 - \alpha \exp(\mathbf{T}y)\mathbf{1}, \quad y \geq 0,$$

where $\mathbf{1}$ is a column vector with all ones.

Besides being dense in the nonnegative 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, is closed under finite mixtures and convolutions, and is closed under the infinite mixture (among other interesting properties). In particular, we will exploit the property that the class of PH-distributions is closed under convolution in Section 3.

2.2 Infinite Mixtures of Phase-type Distributions

A random variable Z of the form

$$Z := W \cdot X, \tag{3}$$

is an IMPH if $X \sim F$, where F is a PH-distribution, and $W \sim H$, where H is an arbitrary non-negative distribution. In this paper we will be mostly interested in the case where H has unbounded support. 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:

$$B(z) = \int_0^\infty F(z/w)\,\mathrm{d}H(w), \quad z \geq 0. \tag{4}$$

The integral expression above is available in closed form in very few isolated cases. Thus, we should rely on integration or simulation methods for its computation. Nevertheless, we can easily obtain the distribution of Z given $W = w$, namely

$$(Z|W = w) \sim F_w,$$

where $F_w(z) := F(z/w)$, which is the cdf of the scaled random variable expressed as $w \cdot X$. We call F_w a *scaled PH-distribution*. Note that F_w remains PH-distribution (Neuts 1975).

The key motivation for considering the class of IMPH over the class of PH is that the later class forms a subclass of light-tailed distributions while a distribution in the former class having unbounded scaling distribution is heavy-tailed. Recall that a non-negative random variable W has a heavy-tailed distribution if only if $\mathbb{E}(\mathrm{e}^{\theta W}) = \infty$, $\forall \theta > 0$. Equivalently, if $\limsup_{w\to\infty} \mathbb{P}(W > w)\mathrm{e}^{\theta w} = \infty$, $\forall \theta > 0$. Otherwise, we say W is light-tailed.

Hence, the class of IMPH 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 ℓ. The key approach is Conditional Monte Carlo (CMC) — a well known variance reduction technique that reduces the variance of an initial Monte Carlo estimator. The method is based on the Rao–Blackwell theorem which says that the variance of an estimator conditional on a set of sufficient statistics is smaller or equal to the variance of the original estimator; that is

$$\mathrm{Var}(\mathrm{Exp}[\widehat{\ell}|T]) \le \mathrm{Var}[\widehat{\ell}], \tag{5}$$

for some T a sufficient statistic. In practical terms, one should be able to simulate T and compute $\mathrm{Exp}[\widehat{\ell}|T]$.

3.1 Conditional Monte Carlo

First, we will apply CMC to the Crude Monte Carlo estimator $\widehat{\ell} = \mathbb{1}_{\{\sum_{i=1}^{M} Z_i > u\}}$. We will condition on $M, W_1, \ldots, W_M$ so that we obtain

$$\ell = \mathbb{E}[\widehat{\ell}] = \mathbb{E}\mathbb{E}[\widehat{\ell}|M, W_1, \ldots, W_M] = \mathbb{E}\mathbb{P}\left(\sum_{i=1}^{M} Z_i > u|M, W_1, \ldots, W_M\right). \tag{6}$$

The key idea here is that $\mathbb{E}[\widehat{\ell}|M, W_1, \ldots, W_M]$ can be computed explicitly because, conditional on $M, W_1, \ldots, W_M$, the sum $\sum_{i=1}^{M} Z_i$ is just a convolution of scaled phase-type distributions. The closure property under convolution of the class of PH-distributions yields

$$\left(\sum_{i=1}^{M} Z_i \,|\, M, W_1, \cdots, W_M\right) \sim \mathrm{PH}(\gamma, \mathbf{Q}(M, W_1, \ldots, W_M)),$$

where

$$\gamma = (\beta, 0, \cdots, 0), \quad \mathbf{Q} := \mathbf{Q}(M, W_1, \ldots, W_M) \begin{pmatrix} \mathbf{T}_1 & \mathbf{t}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 & \mathbf{t}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_{M-1} & \mathbf{t_{M-1}} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{T}_M \end{pmatrix},$$

with $\mathbf{T}_i = \mathbf{T}/W_i$, and $\mathbf{t}_i = -\mathbf{T}\mathbf{1}\beta/W_i$. Therefore, the CMC estimator takes the form

$$\widehat{\ell}_{\mathrm{CMC}} = \gamma \mathrm{e}^{\mathbf{Q}u}\mathbf{1}.$$

These results are summarized in the following algorithm used to generate a single replicate:

Algorithm 1 (Conditional Monte Carlo)

1. Generate M.
2. Generate $W_1, \cdots, W_M \overset{\text{i.i.d}}{\sim} H$.
3. Compute γ and $\mathbf{Q}$.
4. Return $\gamma \mathrm{e}^{\mathbf{Q}u}\mathbf{1}$.

3.2 Boosting via Importance Sampling

The previous algorithm is still considered *crude* because the random variables $M, W_1, \ldots, W_M$ were simulated from their original distribution. In this section, we considering improving the efficiency of our algorithm by implementing IS over the distribution H. Inspired by popular methodologies drawn from light-tailed and heavy-tailed problems we suggest two alternative algorithms.

3.2.1 Exponential Twisting

We first consider the case of light-tailed scaling distributions. The exponential twisting method is proved to be asymptotically efficient for tail probabilities of geometric sum with light-tailed increments (Siegmund 1976). The method is specified as follows:

Define an exponential family of pdfs $\{h_\theta, \theta \in \Theta\}$ based on the original pdf of the scaling random variable, h, via

$$h_\theta(w) = \frac{\mathrm{e}^{\theta w}}{M(\theta)} h(w) = \mathrm{e}^{\theta w - \ln M(\theta)} h(x), \tag{7}$$

where $M(\theta) = \int \mathrm{e}^{\theta w} h(w) \mathrm{d}w, \theta \in \Theta$ is the moment generating function of W. The likelihood ratio of a single element associated with this change of measure is:

$$\frac{h(W)}{h_\theta(W)} = \mathrm{e}^{-\theta W + \zeta(\theta)}, \tag{8}$$

where $\zeta(\theta) = \ln M(\theta)$ is the cumulant function of W. Then the twisted mean is $\mu_\theta = \mathbb{E}_\theta(W) = \zeta'(\theta)$ (Kroese, Taimre, and Botev 2011).

The selection of a proper twisting parameter θ is a key aspect in the implementation of this algorithm. For dealing with the geometric sum (1), we select the twisting parameter θ such that the changed mean of the random sum is equal to the threshold. That is, $\mathbb{E}_\theta(S_M) = u$. This selection of the value of θ is critical for the performance of the algorithm and it is justified by Cramér's Large Deviation Theorem, which implies that this selection is asymptotically optimal.

However, this approach is not directly applicable to our problem because the moment generating function of a heavy-tailed random variable is not defined for positive values of the argument, so the equation $\mathbb{E}(S) = u$, does not have a solution. Nevertheless, in the case with light-tailed scalings an exponential change of measure can still be applied to the random variables W_i. Inspired by Siegmund's algorithm we propose to select the twisting parameter θ as the solution of the equation $\mathbb{E}_\theta(\sum_{i=1}^M Z_i | M) = u$. Note that $\mathbb{E}_\theta(\sum_{i=1}^M Z_i | M) = M\,\mathbb{E}[W]\,\mathbb{E}_\theta[X]$. Since our estimator is conditional on M, we obtain a twisting parameter which depends on M, so we choose the value $\theta(M)$ which solves $\zeta'(\theta(M)) = u/(M\mathbb{E}(X))$.

These results are summarized in the following algorithm used to generate a single replicate:

Algorithm 2 (Conditional Monte Carlo with Exponential Twisting)

1. Generate M, and determine $\theta(M)$ from $\zeta'(\theta(M)) = u/(M\mathbb{E}(X))$.
2. Generate $W_1, \cdots, W_M \overset{\text{i.i.d}}{\sim} H_{\theta(M)}$.
3. Compute γ, $\mathbf{Q}$, and the likelihood ratio $L := e^{M\zeta(\theta(M)) - \theta(M)\sum_{i=1}^M W_i}$.
4. Return $\gamma \mathrm{e}^{\mathbf{Q}u}\mathbf{1} \times L$.

3.2.2 Hazard Rate Twisting

In this section, we consider the case in which scaling distributions are heavy-tailed. Writing the pdf of the scaling random variable W as h, its hazard rate is denoted by $\lambda(w) = h(w)/\overline{H}(w)$. Let $\Lambda(w) = \int_0^w \lambda(y)\mathrm{d}y = -\ln\overline{H}(w)$ denote the hazard function. In (Huang and Shahabuddin 2004), the authors gave a general hazard rate twisted density with parameter θ, $0 < \theta < 1$:

$$h_{W_i,\theta}(w) = \frac{h_{W_i}(w)\, e^{\theta\Lambda_i(w)}}{M_{\Lambda_i(W_i)}(\theta)}\,, \tag{9}$$

where $M_{\Lambda_i(W_i)}(\theta) \equiv \int_0^\infty h_{W_i}(w)\, e^{\theta\Lambda_i(w)}\mathrm{d}w$, which is a normalization constant. If we simply have $\Lambda_i(w) = \Lambda_{W_i}(w)$, then $M_{\Lambda_i(W_i)}(\theta) = 1/(1-\theta)$. Thus the resulting twisted pdf is $h_{W_i,\theta}(w) = \lambda_i(w)(1-\theta)e^{-\int_0^w (1-\theta)\lambda_i(y)\mathrm{d}y}$, which is the same as that in (Juneja and Shahabuddin 2002).

We now can write the twisted pdf of W as:

$$h_\theta(w) = \lambda(w)(1-\theta)e^{-\int_0^w (1-\theta)\lambda(y)\mathrm{d}y},\ \ w \geq 0\,,$$

where $\theta \equiv 1 - m/\Lambda(u)$. This choice of parameter is proved to be asymptotically optimal in (Huang and Shahabuddin 2004) for any constant m, when considering $X \equiv 1$.

It will later be useful to compute the twisted tail probability:

$$\begin{aligned}\overline{H}_\theta(w) &= \int_w^\infty \lambda(t)(1-\theta)e^{-\int_0^t (1-\theta)\lambda(y)\mathrm{d}y}\mathrm{d}t \\ &= \int_w^\infty (1-\theta)\frac{h(t)}{\overline{H}(t)}e^{-(1-\theta)(-\ln\overline{H}(t))}\mathrm{d}t \\ &= \left[\overline{H}(w)\right]^{1-\theta}.\end{aligned} \tag{10}$$

Thus, conditional on $M = m$, and taking in to account $\mathbb{E}(X)$, the likelihood ratio of a single element associate with this change of measure is:

$$\frac{h(w)}{h_\theta(w)} = \frac{1}{1-\theta(m)}e^{-\theta(m)\Lambda(w)}\,,$$

where $\theta(m) = 1 - m/\Lambda(u/\mathbb{E}(X))$.

These results are summarized in the following algorithm used to generate a single replicate:

Algorithm 3 (Conditional Monte Carlo with Hazard Rate Twisting)

1. Generate M, and determine $\theta(M)$ as $\theta(M) = 1 - M/\Lambda(u/\mathbb{E}(X))$.
2. Generate $W_1, \cdots, W_M \overset{\text{i.i.d}}{\sim} H_{\theta(M)}$.
3. Compute γ, $\mathbf{Q}$, and the likelihood ratio $L := (1-\theta(M))^{-M}e^{-\theta(M)\sum_{i=1}^M \Lambda(W_i)}$.
4. Return $\gamma \mathrm{e}^{\mathbf{Q}u}\mathbf{1} \times L$.

3.3 Asmussen–Kroese-type Algorithm

The Asmussen–Kroese estimator (Asmussen and Kroese 2006) is efficient for estimating tail probabilities of geometric sums of heavy-tailed summands. 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 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:

$$\mathbb{P}(S_M > u | M = m) = m\mathbb{P}(S_m > u, Z_m = \max Z_i, i = 1, \ldots, m) = m\mathbb{E}\overline{B}(Z_{m-1} \vee (u - S_{m-1})) .$$

In our setting, the summands $Z = W \cdot X$ are heavy-tailed whenever W has unbounded support 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 is typically unavailable.

Instead, we consider a simple modification, by conditioning on a single scaling random variable W. We further consider applying a change of measure to this single random variable. Conditional on $M = m$, the Asmussen–Kroese estimator (without the control variate correction) for ℓ takes the form

$$m\overline{B}\left(Z_{m-1} \vee (u - S_{m-1})\right) ,$$

where $B(\cdot)$ is the cdf of Z and $x \vee y = \max(x, y)$. In our context, we simply condition on $W = w$, to arrive at

$$m\overline{F}_w\left(Z_{m-1} \vee (u - S_{m-1})\right) ,$$

where $\overline{F}_w(\cdot)$ is the complementary cdf of a scaled PH random variable, which remains PH. We then apply the same IS idea as in the previous section — ensuring that the random sum after conditioning is equal in expectation to u — to twisting only the scaling random variable of Z_m. We then arrive at the following estimator for a single replicate:

$$M\overline{F}_W\left(Z_{M-1} \vee (u - S_{M-1})\right) \times \frac{h(W)}{h_\theta(W)} .$$

Since M is random, we combine the estimator above with a control variable, and this yields the estimator:

$$M\overline{F}_W\left(Z_{M-1} \vee (u - S_{M-1})\right) \times \frac{h(W)}{h_\theta(W)} - (M - \mathbb{E}(M))\,\overline{F}_W(u) \times \frac{h(W)}{h_\theta(W)} .$$

These results are summarized in the following algorithm used to generate a single replicate:

Algorithm 4 (Conditional Asmussen–Kroese with Importance Sampling)

1. Generate M.
2. Generate $Z_1, \ldots, Z_{M-1}$ as $Z_i = W_i \cdot X_i$ with $W_1, \ldots, W_{M-1} \overset{\text{i.i.d}}{\sim} H$ and $X_1, \ldots, X_{M-1} \overset{\text{i.i.d}}{\sim} F$.
3. Compute $S_{M-1} = \sum_{i=1}^{M-1} Z_i$, and determine θ from $\zeta'(\theta) = (u - S_{M-1})/\mathbb{E}(X)$ (light-tailed W) or $\theta = 1 - 1/\Lambda((u - S_{M-1})/\mathbb{E}(X))$ (heavy-tailed W).
4. Generate $W \sim H_\theta$ and compute the likelihood ratio $L := e^{\zeta(\theta) - \theta W}$ (light-tailed W) or $L := (1 - \theta)^{-1} e^{-\theta \Lambda(W)}$ (heavy-tailed W).
5. Return $M\overline{F}_W\left(Z_{M-1} \vee (u - S_{M-1})\right) \times L - (M - \mathbb{E}(M))\,\overline{F}_W(u) \times L$.

4 NUMERICAL EXPERIMENTS

In this section, we provide three simple sets of numerical experiments to illustrate the proposed simulation methods:

- CMC+IS and AK+IS, the scaling distributions are light-tailed (Exponentially distributed).
- CMC+IS and AK+IS, the scaling distributions are heavy-tailed (Pareto distributed).
- CMC+IS and AK+IS, the scaling distributions are heavy-tailed (Lognormal distributed).

In all illustrative cases, we consider only the simplest PH random variable, namely an Exponentially distributed random variable. We remind the reader that in all cases the resulting product $Z = W \cdot X$ is heavy-tailed. Note that we do not include numerical results for CMC without IS, as the efficiency of this estimator is typically vastly inferior to CMC+IS.

Example 1 (Light-tailed Scaling: Exponential) Let $W \sim \text{Exp}(\lambda)$ being the scaling random variable, with pdf

$$h(w) = \lambda e^{-\lambda w},\ \ w \geq 0\,,$$

and for simplicity consider $X \sim \text{Exp}(\mu)$, with pdf

$$f(x) = \mu e^{-\mu x},\ \ x \geq 0\,.$$

The exponentially-twisted scaling random variable S_θ has pdf

$$h_\theta(w) = \lambda_\theta e^{-\lambda_\theta w},\ \ w \geq 0\,,$$

where $\lambda_\theta < \lambda$. Hence the likelihood ratio for a single scaling element is given by

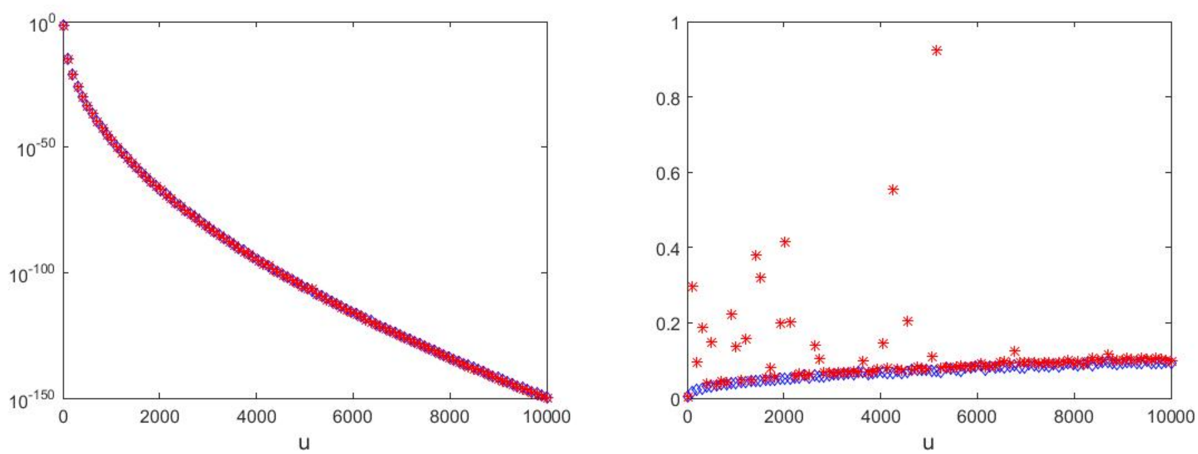
$$\frac{h(w)}{h_\theta(w)} = \frac{\lambda}{\lambda_\theta} e^{-(\lambda - \lambda_\theta)\,w}\,.$$

Thus, conditional on M, we solve $\zeta'(\theta(M)) = u/(M\mathbb{E}(X))$, which in this case reduces to $1/(\lambda - \theta(M)) = u\mu/M$ since $\zeta(\theta) = \ln(\lambda/(\lambda - \theta))$, or in other words $\theta(M) = \lambda - M/(u\mu)$, giving $\lambda_\theta = M/(u\mu)$. This results in likelihood ratio for use in CMC+IS, conditional on M, for all scaling elements given by

$$L := \left(\frac{u\lambda\mu}{M}\right)^M e^{-(\lambda - M/(u\mu))\,\sum_{i=1}^M W_i}\,.$$

Taking a sample size of $N = 10^5$, we apply the CMC+IS algorithm 2 and the AK+IS algorithm 4 to this problem. The corresponding estimates and relative errors are shown in Figure 1.

Figure 1: The parameter for the geometric number of terms in the sum M is $p = 0.2$. The parameters for the IMPH-distributed increments are $\lambda = 1$, $\mu = 3$.



(a) Logarithm of estimated $\ell(u)$ as a function of u. CMC+IS: Red stars. AK+IS: Blue diamonds.

(b) Estimated relative error as a function of u. CMC+IS: Red stars. AK+IS: Blue diamonds.

Example 2 (Heavy-tailed Scaling: Pareto) Let the scaling random variable W be Pareto and with pdf

$$h(w) = \frac{\alpha}{w^{\alpha+1}}\,,\ \ w \geq 1\,,$$

and for simplicity consider $X \sim \text{Exp}(\mu)$. Then W_θ has the hazard-rate twisted pdf:

$$h_\theta(w) = \frac{\alpha(1-\theta)}{w^{1+\alpha(1-\theta)}} \quad w \geq 1\,.$$

Hence the likelihood ratio for a single scaling element is given by

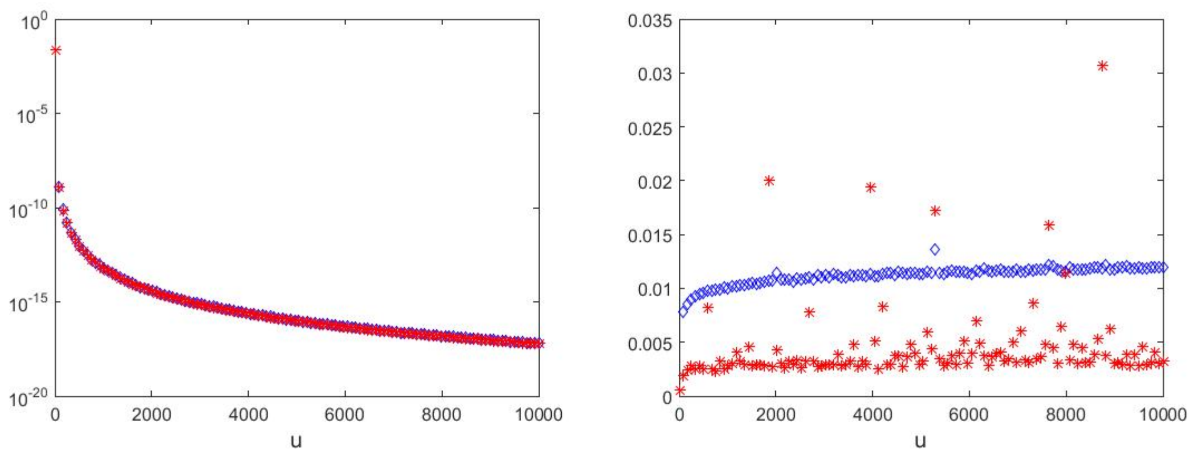
$$\frac{h(w)}{h_\theta(w)} = \frac{1}{1-\theta} w^{-\theta\alpha}\,.$$

Thus, conditional on M, we set $\theta(M) = 1 - M/\Lambda(u/\mathbb{E}(X))$. Since $\Lambda(s) = \alpha \ln(s)$ in this case, this results in likelihood ratio for use in CMC+IS, conditional on M, for all scaling elements given by

$$L := \left(\frac{\alpha}{M}\ln(u\mu)\right)^M \left(\prod_{i=1}^{M} W_i\right)^{-(\alpha - M/\ln(u\mu))}\,.$$

Taking a sample size of $N = 10^5$, we apply the CMC+IS algorithm 3 and the AK+IS algorithm 4 to this problem. The corresponding estimates and relative errors are shown in Figure 2.

Figure 2: The parameter for the geometric number of terms in the sum M is $p = 0.2$. The parameters for the IMPH-distributed increments are $\alpha = 4$, $\mu = 3$.



(a) Logarithm of estimated $\ell(u)$ as a function of u. CMC+IS: Red stars. AK+IS: Blue diamonds.

(b) Estimated relative error as a function of u. CMC+IS: Red stars. AK+IS: Blue diamonds.

Example 3 (Heavy-tailed Scaling: Lognormal) Consider now $X \sim \text{Exp}(\lambda)$, and the scaling random variable $W \sim \text{LogN}(\mu, \sigma^2)$ with pdf

$$h(w) = \frac{1}{w\sigma\sqrt{2\pi}} e^{-\frac{(\ln w - \mu)^2}{2\sigma^2}}, \quad w > 0\,,$$

with corresponding cdf

$$H(w) = \Phi\left(\frac{\ln w - \mu}{\sigma}\right), \quad w > 0\,,$$

where $\Phi(\cdot)$ is the cdf of a standard normal random variate.

The likelihood ratio for a single scaling element is given by

$$\frac{h(w)}{h_\theta(w)} = \frac{1}{1-\theta(m)} \left[\overline{H}(w)\right]^{\theta(m)}\,,$$

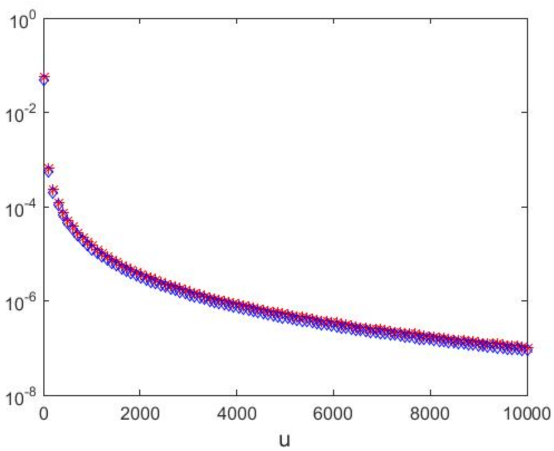
where, conditional on M, $\theta \equiv \theta(M) = 1 - M/\Lambda(u/\mathbb{E}(X))$. This results in likelihood ratio for use in CMC+IS, conditional on M, for all scaling elements given by

$$L := (1-\theta(M))^{-M}\left[\prod_{i=1}^{M}\overline{H}(W_i)\right]^{\theta(M)}.$$

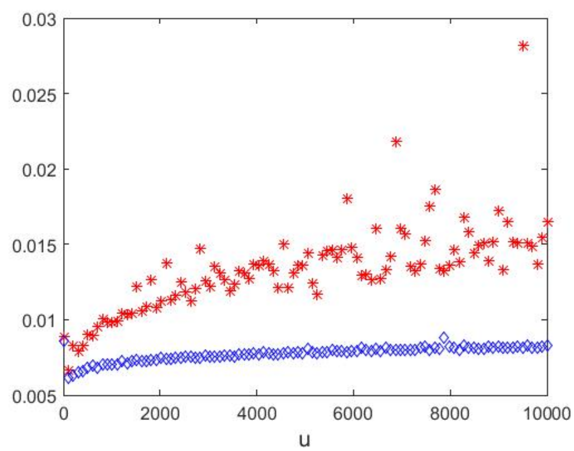
We note that from (10), we have $\overline{H}_\theta(w) = \left[\overline{H}(w)\right]^{1-\theta}$, and so it is straightforward to generate the twisted W by using the inverse transform method.

Taking a sample size of $N = 10^5$, we apply the CMC+IS algorithm 3 and the AK+IS algorithm 4 to this problem. The corresponding estimates and relative errors are shown in Figure 3.

Figure 3: The parameter for the geometric number of terms in the sum M is $p = 0.2$. The parameters for S are $\mu = 0$ and $\sigma = 2$, and X has $\mathrm{Exp}(3)$ distribution.



(a) Logarithm of estimated $\ell(u)$ as a function of u. CMC+IS: Red stars. AK+IS: Blue diamonds.



(b) Estimated relative error as a function of u. CMC+IS: Red stars. AK+IS: Blue diamonds.

5 DISCUSSION AND OUTLOOK

In this paper, we proposed straight-forward simulation methods for estimating tail probabilities of geometric random sums of IMPH distributed summands, exploiting the inherent structure of the summands and properties of the class of PH distributions along the way.

On comparing the proposed CMC+IS and AK+IS approaches in three simple examples, we observe that both methods appear to be producing sharp estimates of the tail probabilities. The AK+IS method has lower estimated relative errors for the exponential and lognormal scaling examples, which suggests that this method is superior to CMC+IS for lighter-tailed scaling random variables. On the other hand, based on the estimated relative errors, CMC+IS appears to yield superior performance than AK+IS for the Pareto example, suggesting it may be preferred for scaling distributions with much heavier tails.

One cautionary point is that the estimates of the relative errors for the CMC+IS method exhibit large variability. This behaviour is often due to inefficient second moment estimates as a result of large skewness in the likelihood ratio terms. In contrast, the estimated relative errors for the AK+IS estimators exhibit stable behaviour. However, it is known that this estimator can under-perform even when the relative error appears stable. Additional stratification strategies such as that proposed in (Juneja and Shahabuddin 2002) can be implemented to improve the accuracy of this estimator.

At present, the parameters of the IS densities are chosen heuristically to ensure that the probability of interest is no longer rare under these changes of measure. One key question which we have not addressed in this work is how to choose this parameters in a principled — and ideally asymptotically optimal — way. Furthermore, in the present work we have assumed for simplicity that all of the random variables in the sum are independent and exponentially distributed. Our work can be easily generalized to include general phase-type distributions. The assumption of independence is not realistic in typical applications. An interesting avenue for further work is to develop effective simulation methods for this problem when there is structured dependence between the random variables.

Finally, a full numerical study comparing the simulation methodologies against the numerical scheme proposed by Bladt *et al.* is still due.

REFERENCES

Asmussen, S. 2003. *Applied Probability and Queues*. 2nd ed. New York: Springer-Verlag.

Asmussen, S., and H. Albrecher. 2010. *Ruin Probabilities*. 2nd ed. Singapore: World Scientific Publishing.

Asmussen, S., and D. Kroese. 2006. "Improved algorithm for rare event simulation with heavy tails". *Advances in Applied Probability* 38:545–558.

Bladt, M., B. Nielsen, and G. Samorodnitsky. 2014. "Calculation of Ruin Probabilities for a Dense Class of Heavy Tailed Distributions". *Scandinavian Actuarial Journal*.

Embrechts, P., and N. Veraverbeke. 1982. "Estimates for the Probability of Ruin with Special Emphasis on the Possibility of Large Claims". *Insurance: Mathematics and Economics* 1:55–72.

Huang, Z., and P. Shahabuddin. 2004. "A Unified Approach for Finite-Dimensional, Rare-Event Monte Carlo Simulation". In *Proceedings of the 2004 Winter Simulation Conference*.

Juneja, S., and P. Shahabuddin. 2002. "Simulating Heavy Tailed Processes Using Delayed Hazard Rate Twisting". *ACM Transactions on Modeling and Computer Simulation* 12.

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

Neuts, M. 1975. "Probability Distributions of Phase Type". *Liber Amicorum Prof. Emeritus H. Florin.*:173–206.

Rojas-Nandayapa, L., and W. Xie. 2015. "Asymptotic tail behavior of phase-type scale mixture distributions". *ArXiv:1502.01811v1 e-prints*.

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

AUTHOR BIOGRAPHIES

Hui Yao is a PhD candidate at the University of Queensland, Australia, where she received her M.Sc. (Statistics) in 2014. Her research interests include Monte Carlo methods, rare event simulation, risk theory, and queueing theory.

Leonardo Rojas-Nandayapa holds a Ph.D. (Science) in 2008 from Aarhus University. He received a B.Sc. (Applied Mathematics, 2002) and M.Sc. (Mathematical Sciences, 2004) from ITAM and IIMAS-UNAM respectively. His research interests are in Applied Probability. He has held lecturer positions at The University of Queensland, Australia and ITAM.

Thomas Taimre is a lecturer of Mathematics and Statistics and received his B.Sc. (Mathematics and Statistics) in 2003, B.Sc. (Hons. I, Statistics) in 2004, and Ph.D. (Mathematics) in 2009, all from The University of Queensland, Australia. He is the co-author of the Handbook of Monte Carlo Methods, which provides a hands-on guide to the theory, algorithms, techniques, and applications of Monte Carlo methods.

His current research is at the interface of probability theory, computer simulation, and mathematical optimisation with biological and other scientific, engineering, and finance disciplines.