

Article

# Estimating Tail Probabilities of Random Sums of Phase-Type Scale Mixture Random Variables <sup>†</sup>

Hui Yao \* and Thomas Taimre

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

\* Correspondence: h.yao@uq.edu.au

<sup>†</sup> This paper is an extended version of our paper published in WSC2016 and MODSIM2021.

**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 phase-type random variable. Our motivation arises from applications in risk, queueing problems for estimating ruin probabilities, and waiting time distributions, respectively. Mixtures of distributions are flexible models and can be exploited in modelling non-life insurance loss amounts. 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 are not even available for the class of scale mixture of phase-type distributions. The contributions of this paper are that we address these issues by proposing alternative simulation methods for estimating tail probabilities of random sums of scale mixture of phase-type distributions which combine importance sampling and conditional Monte Carlo methods, showing the efficiency of the proposed estimators for a wide class of scaling distributions, and validating the empirical performance of the suggested methods via numerical experimentation.

**Keywords:** subexponential distribution; phase-type distribution; rare event simulation; conditional Monte Carlo; importance sampling



**Citation:** Yao, H.; Taimre, T. Estimating Tail Probabilities of Random Sums of Phase-Type Scale Mixture Random Variables.

*Algorithms* **2022**, *15*, 350. <https://doi.org/10.3390/a15100350>

Academic Editors: Georgy Sofronov and Andras Farago

Received: 5 August 2022

Accepted: 23 September 2022

Published: 27 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 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, [1].
- 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, [2].

Estimation from sample data and density approximation with phase-type distributions was considered in [3], and parameter estimation for the class of discrete scaled phase-type distributions was subsequently treated in [4]. Maximum likelihood estimation via an expectation-maximisation (EM) algorithm was exploited in both papers.

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 ([5]). From a rare-event simulation perspective, there are a few efficient estimators proposed in the literature ([6–8]). 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 not easily implementable.

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

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

where  $S_N = Z_1 + \dots + Z_N$ , for large  $u$ . Specifically,  $N$  is a discrete light-tailed random variable supported over the 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 *phase-type distributions* (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 PH-distributed random variable. The concept of  $W$ -mixture of PH distributions can be found in [9]. Bladt et al. [10] proposed a subclass of  $W$ -mixture of PH distributions to approximate any heavy-tailed distribution. Such a class is very attractive in stochastic modelling 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 behaviour of a heavy-tailed distribution cannot be captured correctly by PH distributions alone. Reference [11] 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. [10] 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 analyse 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 [7]. 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 either the PH random variable  $X_i$  or the scaling random variable  $W_i$ , thereby simply simulating the PH distribution and using the CDF of  $W_i$  or simulating the scaling distribution and using the CDF of  $X_i$ . Moreover, we explore the use of importance sampling (IS) on the last term of the summands.

The key contributions of this paper are as follows:

- Developing a number of rare event simulation methodologies;
- Proving the efficiency of proposed estimators under certain conditions;
- Exploring the proposed estimators through various numerical experiments.

The remainder of the paper is organised 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 algorithms. In Section 4, we prove the efficiency of the estimators proposed. In Section 5, we present the empirical results for several examples. Section 6 provides some concluding remarks and an outlook to future work.

## 2. Preliminaries

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

### 2.1. Phase-Type Distributions and Properties

PH distributions have been used in stochastic modelling since being introduced in [12]. 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 [2] for details). In other words, PH distributions may arbitrarily closely approximate 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  $E = \{1, 2, \dots, p\} \cup \{\Delta\}$ , where states  $1, 2, \dots, p$  are transient and state  $\Delta$  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  $\alpha$ , 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)  $Q$  of the form:

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

where  $T$  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  $(\alpha, T)$  completely specifies the PH distribution, and is called a PH-representation. The CDF is given by  $F(y) = 1 - \alpha \exp(Ty)\mathbf{1}$ ,  $y \geq 0$ , 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 [12]).

### 2.2. Scale Mixture of Phase-Type Distributions

A random variable  $Z$  of the form  $Z := W \cdot X$  is a 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  $H$ :

$$B(z) = \int_0^\infty F(z/w) dH(w) = \int_0^\infty H(z/x) dF(x), \quad z \geq 0.$$

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.

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 ([11]). Hence, the class of scale mixture of PH distributions turns out to be an appealing tractable class for approximating heavy-tailed distributions. Recall that a non-negative random variable  $X$  has a heavy-tailed distribution if and only if  $\mathbb{E}[e^{\theta X}] = \infty, \forall \theta > 0$ , equivalently, if  $\limsup_{x \rightarrow \infty} \mathbb{P}(X > x)e^{\theta x} = \infty, \forall \theta > 0$ . Otherwise, we say  $X$  is light-tailed.

### 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 and further with importance sampling when necessary. The method is inspired by using the tower property of expectation  $\mathbb{E}\widehat{\ell}(u) = \mathbb{E}\mathbb{E}[\widehat{\ell}(u)|T]$  where we have conditioned on some statistic  $T$ , and so in practical terms, one should be able to simulate  $T$  and compute  $\mathbb{E}[\widehat{\ell}(u)|T]$ .

#### 3.1. Asmussen–Kroese-Type Algorithm

The Asmussen–Kroese estimator [7] is efficient for estimating tail probabilities of light-tailed random sums of heavy-tailed summands [13].

The key idea is based on the law of total probability and  $\mathbb{P}(S_n > u, \max\{Z_i, i = 1, \dots, n\} = Z_i)$  being the same for  $i = 1, \dots, n$  (as  $\{Z_i, i = 1, \dots, n\}$  are i.i.d. random variables). We then have the following identity:

$$\mathbb{P}(S_N > u|N = n) = n\mathbb{P}(S_n > u, \max\{Z_i, i = 1, \dots, n\} = Z_n) = n\mathbb{E}\bar{B}(Z_{n-1}^* \vee (u - S_{n-1})),$$

where  $\bar{B}(\cdot)$  is the complementary CDF of  $Z_i$ ,  $Z_{n-1}^* = \max\{Z_i, i = 1, \dots, n - 1\}$ ,  $S_{n-1} = \sum_{i=1}^{n-1} Z_i$ , 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 [11] 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 scaling random variable  $W_N$  or on the PH random variable  $X_N$ . We further consider applying a change of measure to this single random variable. Conditioning on  $N$ , the Asmussen–Kroese estimator (without applying the control variate method for  $N$ ) for  $\ell(u)$  takes the form

$$\widehat{\ell}_{AK}(u) = N \bar{B}(Z_{N-1}^* \vee (u - S_{N-1})). \tag{2}$$

In our context, when conditioning on the  $W_N$  associated with  $Z_N$ , we arrive at

$$\widehat{\ell}_{ConAK1}(u) = N \bar{F}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{W_N}\right), \tag{3}$$

where  $\bar{F}(\cdot)$  is the complementary CDF of a PH random variable.

When conditioning on the  $X_N$  associated with  $Z_N$ , we arrive at

$$\widehat{\ell}_{ConAK2}(u) = N \bar{H}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{X_N}\right), \tag{4}$$

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

We then consider improving the efficiency of our algorithm by implementing IS over the distribution  $H$  or  $F$ , ensuring that the random sum after conditioning is equal in expectation to  $u$ . Inspired by popular methodologies drawn from light-tailed and heavy-tailed problems, we suggest two alternative approaches.

##### 3.1.1. Exponential Twisting

The exponential twisting method is asymptotically optimal for tail probabilities of random sums with light-tailed summands ([14]).

**Definition 1.** The estimator  $\widehat{\ell}(u)$  generated using the probability measure  $\mathbb{P}$  is said to be an asymptotically optimal estimator of  $\mathbb{E}_{\mathbb{P}}[\widehat{\ell}(u)]$  if and only if

$$\log\left(\sqrt{\mathbb{E}_{\mathbb{P}}[\widehat{\ell}(u)^2]}\right) / \log\left(\mathbb{E}_{\mathbb{P}}[\widehat{\ell}(u)]\right) \sim 1.$$

The method is specified as follows:

Define an exponential family of PDFs  $\{f_{\theta}, \theta \in \Theta\}$  based on the original PDF of any random variable  $X, f$ , via

$$f_{\theta}(x) = \frac{e^{\theta x}}{M_X(\theta)} f(x) = e^{\theta x - \ln M_X(\theta)} f(x),$$

where  $M_X(\theta) = \int e^{\theta x} f(x) dx$  is the MGF of  $X$ .

The likelihood ratio of a single element associated with this change of measure (i.e., using measure  $f_{\theta}(x) dx$  instead of  $f(x) dx$ ) is

$$\frac{f(X)}{f_{\theta}(X)} = e^{-\theta X + \zeta(\theta)},$$

where  $\zeta(\theta) = \ln M_X(\theta)$  is the cumulant function of  $X$ . Then, the twisted mean is  $\mu_{\theta} = \mathbb{E}_{\theta}[X] = \zeta'(\theta)$  ([15]).

The selection of a proper twisting parameter  $\theta$  is a key aspect in the implementation of this method. For dealing with the random sum (1), we select the twisting parameter  $\theta$  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  $\mathbb{E}_{\theta}[\sum_{i=1}^N Z_i | N = n] = n \mathbb{E}[W] \mathbb{E}_{\theta}[X]$ .

For the case of conditioning on light-tailed scaling random variables, we combine the estimator (3) with the above importance sampling method, and arrive at the following estimator for a single replicate:

$$\widehat{\ell}_{\text{ConAK1+IS}}(u) = N \bar{F}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{W_N}\right) \times \frac{h(W_N)}{h_{\theta}(W_N)}, \tag{5}$$

where  $W_N$  is a generic copy of the scaling random variable, and  $h$  is the PDF of  $W_N$ .

We further combine the estimators (3) and (5) with a control variate for  $N$  to improve the efficiency, which is based on the subexponential asymptotics  $\lim_{u \rightarrow \infty} \frac{\mathbb{P}(S_N > u)}{\mathbb{E}[N] \bar{B}(u)} = 1$ :

$$\widehat{\ell}_{\text{ConAK1+CV}}(u) = N \bar{F}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{W_N}\right) - (N - \mathbb{E}[N]) \bar{F}\left(\frac{u}{W_N}\right), \tag{6}$$

$$\widehat{\ell}_{\text{ConAK1+IS+CV}}(u) = N \bar{F}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{W_N}\right) \times \frac{h(W_N)}{h_{\theta}(W_N)} - (N - \mathbb{E}[N]) \bar{F}\left(\frac{u}{W_N}\right) \times \frac{h(W_N)}{h_{\theta}(W_N)}. \tag{7}$$

For the case of conditioning on the PH random variables, we combine the estimator (4) with the above importance sampling method for both light- and heavy-tailed scaling distributions, and arrive at the following estimator for a single replicate:

$$\widehat{\ell}_{\text{ConAK2+IS}}(u) = N \bar{H}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{X_N}\right) \times \frac{f(X_N)}{f_{\theta}(X_N)}. \tag{8}$$

where  $X_N$  is a generic copy of the PH random variable, and  $f$  is the PDF of  $X_N$ .

We further combine the estimators (4) and (8) with a control variate for  $N$  to improve the efficiency, which is based on the same subexponential asymptotics as in (6):

$$\widehat{\ell}_{\text{ConAK2+CV}}(u) = N \bar{H}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{X_N}\right) - (N - \mathbb{E}[N]) \bar{H}\left(\frac{u}{X_N}\right), \tag{9}$$

$$\widehat{\ell}_{\text{ConAK2+IS+CV}}(u) = N\bar{H}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{X_N}\right) \times \frac{f(X_N)}{f_\theta(X_N)} - (N - \mathbb{E}[N])\bar{H}\left(\frac{u}{X_N}\right) \times \frac{f(X_N)}{f_\theta(X_N)}. \tag{10}$$

### 3.1.2. 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  $\lambda(w) = h(w)/\bar{H}(w)$ . Let  $\Lambda(w) = \int_0^w \lambda(y)dy = -\ln\bar{H}(w)$  denote the hazard function. Reference [8] introduced hazard rate twisting and the hazard rate twisted density with parameter  $\theta, 0 \leq \theta < 1$ :

$$h_\theta(w) = \frac{h(w) e^{\theta\Lambda(w)}}{\check{M}(\theta)},$$

where  $\check{M}(\theta) \equiv \int_0^\infty h(w) e^{\theta\Lambda(w)}dw$ , which is the normalisation constant. The resulting twisted PDF is

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

where  $\theta \equiv 1 - n/\Lambda(u)$ .

The following equation will later be useful to compute the twisted tail probability:

$$\begin{aligned} \bar{H}_\theta(w) &= \int_w^\infty \lambda(t)(1 - \theta) e^{-\int_0^t (1-\theta)\lambda(y)dy} dt \\ &= \int_w^\infty (1 - \theta) \frac{h(t)}{\bar{H}(t)} e^{-(1-\theta)(-\ln\bar{H}(t))} dt \\ &= [\bar{H}(w)]^{1-\theta}. \end{aligned} \tag{11}$$

Thus, conditioning on  $N = n$ , and taking in to account  $\mathbb{E}[X]$ , the likelihood ratio of a single element associated with this change of measure is

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

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

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

$$\widehat{\ell}_{\text{ConAK1+IS}}(u) = N\bar{F}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{W_N}\right) \times \frac{h(W_N)}{h_\theta(W_N)}. \tag{12}$$

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

$$\widehat{\ell}_{\text{ConAK1+IS+CV}}(u) = N\bar{F}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{W_N}\right) \times \frac{h(W_N)}{h_\theta(W_N)} - (N - \mathbb{E}[N])\bar{F}\left(\frac{u}{W_N}\right) \times \frac{h(W_N)}{h_\theta(W_N)}. \tag{13}$$

Note that Equations (12) and (13) have the same form as Equations (6) and (7), respectively. When we combine with the importance sampling method in this section, we exploit the hazard rate twisting method instead of the exponential twisting method.

Conditional (either on  $W_N$  or on  $X_N$ ) Asmussen–Kroese with importance sampling and control variate results are summarised in Algorithms 1 and 2 used to generate a single replicate.

Note that we replace  $u$  by  $Z_{N-1}^* \vee (u - S_{N-1})$  when determining  $\theta$ , since we perform change of measure only on  $Z_N$ .

---

**Algorithm 1** Conditional (on  $W_N$ ) Asmussen–Kroese with importance sampling and control variate algorithm

---

1. Generate  $N$ .
  2. Generate  $Z_1, \dots, Z_{N-1}$  as  $Z_i = W_i X_i$  with  $W_1, \dots, W_{N-1} \stackrel{\text{i.i.d.}}{\sim} H$  and  $X_1, \dots, X_{N-1} \stackrel{\text{i.i.d.}}{\sim} F$ .
  3. Compute  $S_{N-1} = \sum_{i=1}^{N-1} Z_i$ , find  $Z_{N-1}^* \vee (u - S_{N-1})$ , and determine  $\theta$  from  $\zeta'(\theta) = (Z_{N-1}^* \vee (u - S_{N-1}) / \mathbb{E}[X_N])$  (light-tailed  $W_N$ ) or  $\theta = 1 - 1/\Lambda((Z_{N-1}^* \vee (u - S_{N-1}) / \mathbb{E}[X_N]))$  (heavy-tailed  $W_N$ ).
  4. Generate  $W_N \sim H_\theta$  and compute the likelihood ratio  $L := e^{\zeta(\theta) - \theta W_N}$  (light-tailed  $W_N$ ) or  $L := (1 - \theta)^{-1} e^{-\theta \Lambda(W_N)}$  (heavy-tailed  $W_N$ ).
  5. Return  $\hat{\ell}_{\text{ConAK1+IS+CV}}(u) = N \bar{F}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{W_N}\right) \times L - (N - \mathbb{E}[N]) \bar{F}\left(\frac{u}{W_N}\right) \times L$ .
- 

**Algorithm 2** Conditional (on  $X_N$ ) Asmussen–Kroese with importance sampling and control variate algorithm

---

1. Generate  $N$ .
  2. Generate  $Z_1, \dots, Z_{N-1}$  as  $Z_i = W_i X_i$  with  $W_1, \dots, W_{N-1} \stackrel{\text{i.i.d.}}{\sim} H$  and  $X_1, \dots, X_{N-1} \stackrel{\text{i.i.d.}}{\sim} F$ .
  3. Compute  $S_{N-1} = \sum_{i=1}^{N-1} Z_i$ , find  $Z_{N-1}^* \vee (u - S_{N-1})$ .
  4. Generate  $X_N \sim F_\theta$ , with parameter  $\theta$  solving  $\zeta'(\theta) = (Z_{N-1}^* \vee (u - S_{N-1}) / \mathbb{E}[W_N])$ .
  5. Return  $\hat{\ell}_{\text{ConAK2+IS+CV}}(u) = N \bar{H}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{X_N}\right) \times \frac{f(X_N)}{f_\theta(X_N)} - (N - \mathbb{E}[N]) \bar{H}\left(\frac{u}{X_N}\right) \times \frac{f(X_N)}{f_\theta(X_N)}$ .
- 

**4. Efficiency of the Modified Amussen–Kroese Estimator**

In this section, we investigate the asymptotic relative error of the estimators  $\hat{\ell}(u)$ . Since  $\ell(u) = \mathbb{E}[\hat{\ell}(u)]$ , the estimator  $\hat{\ell}(u)$  is consistent and we have the following second-order efficiency measures (see [15]),

1. The estimator is said to be logarithmically efficient if the following condition on the logarithmic rates holds:

$$\liminf_{u \rightarrow \infty} \frac{\log(\text{Var}(\hat{\ell}(u)))}{2 \log(\ell(u))} \geq 1.$$

2. The estimator is said to have bounded relative error if

$$\limsup_{u \rightarrow \infty} \frac{\text{Var}(\hat{\ell}(u))}{\ell(u)^2} \leq K < \infty.$$

where  $K$  is a constant that does not depend on  $u$ .

3. The estimator is said to have (asymptotically) vanishing relative error if

$$\limsup_{u \rightarrow \infty} \frac{\text{Var}(\hat{\ell}(u))}{\ell(u)^2} = 0.$$

We focus on subexponential distributions and related distributions. The definitions of these relevant distributions are listed as follows.

**Definition 2.** A distribution function  $F$  with support on  $(0, \infty)$  is subexponential, if  $\forall n \geq 2$ , it satisfies

$$\lim_{x \rightarrow \infty} \frac{\bar{F}^{*n}(x)}{\bar{F}(x)} = n,$$

where  $\bar{F}^{*n}(x)$  denotes the  $n$ -fold convolution of  $\bar{F}$ . Denote the class of subexponential distribution functions by  $\mathcal{S}$ .

**Definition 3.** A distribution function  $F$  with support on  $(0, \infty)$  is dominatedly varying, if  $\forall y \in (0, 1)$ , it satisfies

$$\limsup_{x \rightarrow \infty} \frac{\bar{F}(yx)}{\bar{F}(x)} < \infty.$$

Denote the class of dominatedly varying distribution functions by  $\mathcal{D}$ .

**Definition 4.** A distribution function  $F$  with support on  $(0, \infty)$  is long-tailed, if  $\forall y \geq 0$ , it satisfies

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x - y)}{\bar{F}(x)} = 1.$$

Denote the class of long-tailed distribution functions by  $\mathcal{L}$ .

We also exploit the definitions of Weibullian tail from [16].

**Definition 5 ([16]).** A distribution  $F$  on  $[0, \infty)$  is said to have a Weibullian tail

$$\bar{F}(z) \sim Cz^\gamma \exp[-\beta z^\alpha], \text{ where } \alpha, \beta, C > 0, \gamma \in (-\infty, \infty).$$

We write  $Z \in \mathcal{W}(\alpha, \beta, \gamma, C)$  if  $Z$  satisfies Definition 5.

The following examples are two examples of widely used distributions that have Weibullian tails.

**Example 1.** The CDF of a Weibull distribution with shape parameter  $b$  and scale parameter  $a$  is

$$\bar{H}(w) = e^{-(w/a)^b}, \quad x \geq 0,$$

denoted as Weibull( $b, a$ ), and  $W \in \mathcal{W}(b, a^{-b}, 0, 1)$ .

**Example 2.** The asymptotic form of the tail of a general PH distribution has the form

$$\bar{F}(x) \sim Cx^k e^{-\eta x},$$

where  $C, \eta > 0$  and  $k \in \{0, 1, 2, \dots\}$  (cf. [1]), so that we can write  $X \in \mathcal{W}(1, \eta, k, C)$ .

Lemma 1 in [16] provides the way to calculate the parameters of product of two random variables having Weibullian tails.

**Lemma 1 ([16]).** Let  $W \in \mathcal{W}(\alpha_1, \beta_1, \gamma_1, C_1)$ ,  $X \in \mathcal{W}(\alpha_2, \beta_2, \gamma_2, C_2)$  be independent non-negative random variables. Then,  $W \cdot X \in \mathcal{W}(\alpha, \beta, \gamma, C)$  with

$$\begin{aligned} \alpha &= \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}, \\ \beta &= \beta_1^{\alpha_2 / (\alpha_1 + \alpha_2)} \beta_2^{\alpha_1 / (\alpha_1 + \alpha_2)} \left[ \left( \frac{\alpha_1}{\alpha_2} \right)^{\alpha_2 / (\alpha_1 + \alpha_2)} \left( \frac{\alpha_2}{\alpha_1} \right)^{\alpha_1 / (\alpha_1 + \alpha_2)} \right], \\ \gamma &= \frac{\alpha_1 \alpha_2 + 2\alpha_1 \gamma_2 + 2\alpha_2 \gamma_1}{2(\alpha_1 + \alpha_2)}, \\ C &= \sqrt{2\pi} C_1 C_2 \frac{1}{\sqrt{\alpha_1 + \alpha_2}} (\alpha_1 \beta_1)^{(\alpha_2 - 2\gamma_1 + 2\gamma_2) / (2(\alpha_1 + \alpha_2))} (\alpha_2 \beta_2)^{(\alpha_1 - 2\gamma_2 + 2\gamma_1) / (2(\alpha_1 + \alpha_2))}. \end{aligned}$$

Suppose that we have the following conditions for the random variables involved.

**Condition 1.** Assume that  $N$  is an integer-valued random variable with  $\mathbb{E}[N^2] < \infty$  and  $\mathbf{X} = (X_1, X_2, X_3, \dots)$  is a sequence of i.i.d. random variables with continuous distribution function  $B \in \mathcal{S}$ , independent of  $N$ .

**Condition 2.** There exists a function  $a(u)$  that satisfies one of the following:

- (1)  $\lim_{u \rightarrow \infty} \frac{\widehat{\ell}(u)}{a(u)} = N$  almost surely;
- (2)  $\lim_{u \rightarrow \infty} \frac{\widehat{\ell}(u)}{a(u)} = \mathbb{E}[N]$ .

**Condition 3.** There exists a random variable  $U$  with  $\mathbb{E}[U^2] < \infty$ , such that for all  $u > 0$ ,

$$U \geq \frac{\widehat{\ell}(u)}{a(u)},$$

almost surely.

**Lemma 2.** If Condition 2 (1) and Condition 3 are fulfilled, we obtain

$$\lim_{u \rightarrow \infty} \frac{\text{Var}[\widehat{\ell}(u)]}{\mathbb{E}[\widehat{\ell}(u)]^2} = \frac{\text{Var}[N]}{\mathbb{E}[N]^2}. \tag{14}$$

If Condition 2 (2) and Condition 3 are fulfilled, we obtain

$$\lim_{u \rightarrow \infty} \frac{\text{Var}[\widehat{\ell}(u)]}{\mathbb{E}[\widehat{\ell}(u)]^2} = 0. \tag{15}$$

**Proof.**

$$\lim_{u \rightarrow \infty} \frac{\text{Var}[\widehat{\ell}(u)]}{\mathbb{E}[\widehat{\ell}(u)]^2} = \lim_{u \rightarrow \infty} \mathbb{E} \left[ \frac{\widehat{\ell}(u)}{a(u)} \right]^{-2} \left( \mathbb{E} \left[ \left( \frac{\widehat{\ell}(u)}{a(u)} \right)^2 \right] - \mathbb{E} \left[ \frac{\widehat{\ell}(u)}{a(u)} \right]^2 \right). \tag{16}$$

If Condition 3 is fulfilled, by dominated convergence, we then interchange limit and expectation in (16) to obtain the result.  $\square$

Note that if Condition 1 is fulfilled, then (14) is bounded.

We list the conditions and lemma above which are mirrored from [13] for the reader’s convenience.

**Proposition 1.** The estimators  $\widehat{\ell}_{\text{ConAK1}}(u)$  and  $\widehat{\ell}_{\text{ConAK2}}(u)$

- Have bounded relative error when  $N$  is light-tailed, and  $W \in \mathcal{D} \cap \mathcal{L}$ ;
- Are logarithmically efficient when  $N$  is deterministic, and  $W$  has a Weibullian tail with shape parameter  $\in (0, \log(3/2)/\log(4/3))$ .

**Proof.** Based on the law of total variance  $\text{Var}(Y) = \mathbb{E}[\text{Var}(Y | X)] + \text{Var}[\mathbb{E}[Y | X]]$ , we can conclude that the conditional Monte Carlo estimator has lower variance than the estimator  $Y$ . Thus, we can obtain  $\text{Var}[\widehat{\ell}_{\text{ConAK1}}(u)] \leq \text{Var}[\widehat{\ell}_{\text{AK}}(u)]$  and  $\text{Var}[\widehat{\ell}_{\text{ConAK2}}(u)] \leq \text{Var}[\widehat{\ell}_{\text{AK}}(u)]$  for conditional Asmussen–Kroese estimators  $\widehat{\ell}_{\text{ConAK1}}(u)$  and  $\widehat{\ell}_{\text{ConAK2}}(u)$ .

The Asmussen–Kroese estimator (2) was proven to be logarithmically efficient when  $Z$  has a Weibullian tail in [7], and to have bounded relative error in [13] for the following three cases:

- $N$  is light-tailed, and  $Z \in \mathcal{D} \cap \mathcal{L}$ .
- $N$  is light-tailed, and  $Z \sim \log\text{-normal}(\mu, \sigma^2)$ .
- $N$  is bounded, and  $Z$  is Weibull-distributed with shape parameter  $\in (0, \log(3/2)/\log(3))$ .

Here, we extended the result to scaling random variables in similar cases for the estimators  $\widehat{\ell}_{\text{ConAK1}}(u)$  and  $\widehat{\ell}_{\text{ConAK2}}(u)$ , which are as follows:

- $W \in \mathcal{D} \cap \mathcal{L}$

As a consequence of Corollary 3.3 in [17], if  $F$  is a non-degenerate phase-type distri-

bution, and the scaling distribution  $H \in \mathcal{D} \cap \mathcal{L}$ , then the phase-type scale mixture distribution  $B \in \mathcal{D} \cap \mathcal{L}$ . Therefore, the estimators  $\widehat{\ell}_{\text{ConAK1}}(u)$  and  $\widehat{\ell}_{\text{ConAK2}}(u)$  attain the same efficiency as the Asmussen–Kroese estimator, as they have lower variance.

- $W \in \mathcal{W}(\alpha, \beta, \gamma, C)$

By Lemma 1, we observe that a phase-type distribution has a Weibullian tail with shape parameter being 1, and if the scaling random variable  $W \in \mathcal{W}(\alpha, \beta, \gamma, C)$ , then  $Z \in \mathcal{W}(\alpha_Z, \beta_Z, \gamma_Z, C_Z)$ . Furthermore, the shape parameter of  $Z$  is  $\alpha_Z = \frac{\alpha}{\alpha+1} \in (0, 1)$ ; thus,  $Z$  is heavy-tailed.

In [7], the authors restricted to the Weibull-type distribution with shape parameter in  $(0, \log(3/2)/\log(2))$  and showed that the Asmussen–Kroese estimator is logarithmically efficient when  $N$  is deterministic. By Lemma 1, we obtain that if the shape parameter of the scaling random variable  $\alpha \in (0, \log(3/2)/\log(4/3))$ , then the shape parameter of the phase-type scale mixture distribution is  $\alpha_Z \in (0, \log(3/2)/\log(2))$ . Therefore, the estimators  $\widehat{\ell}_{\text{ConAK1}}(u)$  and  $\widehat{\ell}_{\text{ConAK2}}(u)$  attain the same efficiency as the Asmussen–Kroese estimator if the scaling random variable  $W \in \mathcal{W}(\alpha, \beta, \gamma, C)$  with  $\alpha \in (0, \log(3/2)/\log(4/3))$ .

□

Further combining with a control variate for the number of summands  $N$  will additionally improve the efficiency of the estimators  $\widehat{\ell}_{\text{ConAK1}}(u)$  and  $\widehat{\ell}_{\text{ConAK2}}(u)$ .

### 5. Numerical Experiments

In this section, we provide a variety of numerical experiments to illustrate the proposed simulation methods. In all illustrative cases, we consider each scaling random variable  $W$  to have support on all of  $[0, \infty)$  and each  $X$  to be Erlang-distributed with shape parameter being  $k$  and rate parameter being  $\beta$ . Erlang distributions are selected as an example PH distribution because they play an important role in the class of PH distributions ([18]). The class of generalised Erlang distributions (a series of  $k$  exponentials with their own rates  $\beta_i$  where  $i \in \{1, 2, \dots, k\}$ ) is dense in the set of all probability distributions on the non-negative half-line.

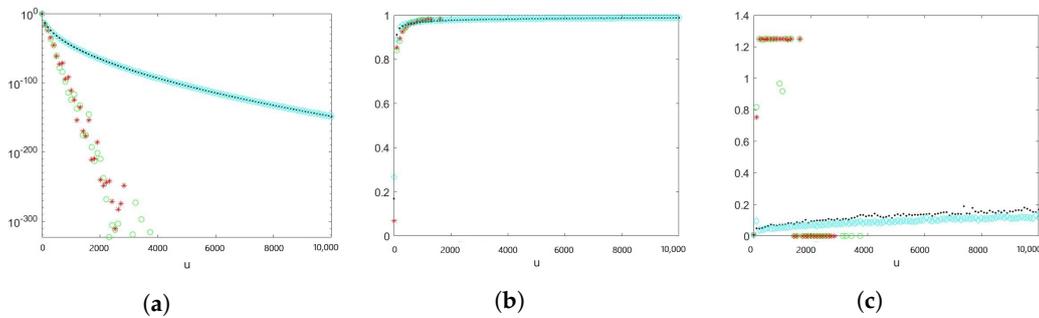
We remind the reader that in all cases, the resulting product  $Z = W \cdot X$  is heavy-tailed since the scaling random variable has unbounded support.

We explore the estimators  $\widehat{\ell}_{\text{ConAK2+CV}}$  which are the conditional Asmussen–Kroese estimator with control variates for the number of summands, and the estimators  $\widehat{\ell}_{\text{ConAK1+IS+CV}}$  and  $\widehat{\ell}_{\text{ConAK2+IS+CV}}$ , which are the conditional Asmussen–Kroese estimator with importance sampling on the last term of the summands and control variate for the number of summands.

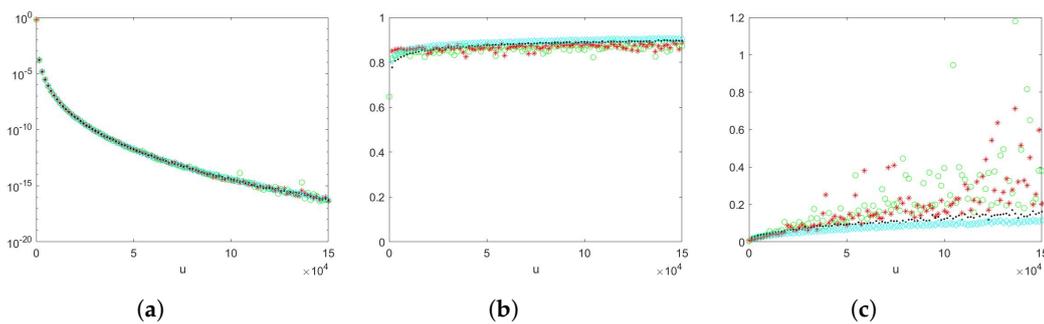
The figures we include in this paper are the estimates of  $\ell(u)$  for  $u$  in the range  $[0, 10^4)$ , empirical logarithmic rates obtained by estimating  $\log(\widehat{\text{Var}}(\widehat{\ell}(u)))/2\log(\ell(u))$ , and empirical relative errors obtained by estimating  $\widehat{\text{Var}}(\widehat{\ell}(u))/\ell(u)^2$ .

We explore the estimators combined with control variates for the number of summands to the exclusion of the estimators not combined with control variates in later examples, because combining with control variates will slightly improve the efficiency of the proposed estimators. We provide a preliminary comparison for the estimators  $\widehat{\ell}_{\text{ConAK1}}$ ,  $\widehat{\ell}_{\text{ConAK1+CV}}$ ,  $\widehat{\ell}_{\text{ConAK1+IS}}$  and  $\widehat{\ell}_{\text{ConAK1+IS+CV}}$  for the scaling random variable being exponential, and the estimators  $\widehat{\ell}_{\text{ConAK2}}$ ,  $\widehat{\ell}_{\text{ConAK2+CV}}$ ,  $\widehat{\ell}_{\text{ConAK2+IS}}$  and  $\widehat{\ell}_{\text{ConAK2+IS+CV}}$  for the scaling random variable being Weibull, aiming to rule out estimators that do not perform well in later examples. The details of density selection regarding importance sampling for the preliminary examples are discussed in Examples 3 and 4, and the results are summarised in Figures 1 and 2, respectively. We observe that in Figure 1, estimators  $\widehat{\ell}_{\text{ConAK1}}$  and  $\widehat{\ell}_{\text{ConAK1+CV}}$  perform poorly due to numerical instability. When  $u$  exceeds around 2600, it is beyond the capability of data calculation. Thus, we do not include the estimators  $\widehat{\ell}_{\text{ConAK1}}$  and  $\widehat{\ell}_{\text{ConAK1+CV}}$  in later examples. In Figure 2, all estimators appear to provide sharp estimates. We also observe that the empirical logarithmic rate tends to 1 when  $u$  gets large, which suggests that the

estimators are logarithmically efficient. Nevertheless, the empirical relative errors for the estimators without importance sampling is increasing and not as stable when  $u$  becomes large. Since we could observe a slight improvement when combining with control variates from both Figures 1 and 2, we include the estimators combined with control variates exclusively.



**Figure 1.** The random variable  $N$  is geometrically distributed with success probability  $p = 0.2$ ,  $W_i \stackrel{i.i.d}{\sim} \text{Exp}(1)$ ,  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(1,3)$ , where  $i = 1, \dots, N$ . Green circles: ConAK1. Red stars: ConAK1+CV. Black dots: ConAK1+IS (conditioning and IS on exponential). Cyan diamonds: ConAK1+IS+CV (conditioning and IS on exponential). (a) Estimates of  $\ell(u)$  as a function of  $u$  on a logarithmic scale. (b) Empirical logarithmic rates as a function of  $u$ . (c) Empirical relative errors as a function of  $u$ .



**Figure 2.** The random variable  $N$  is geometrically distributed with success probability  $p = 0.2$ ,  $W_i \stackrel{i.i.d}{\sim} \text{Weibull}(0.35,2)$ ,  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(2,3)$ , where  $i = 1, \dots, N$ . Green circles: ConAK2. Red stars: ConAK2+CV. Black dots: ConAK2+IS (conditioning and IS on Erlang using an exponential PDF). Cyan diamonds: ConAK2+IS+CV (conditioning and IS on Erlang using an exponential PDF). (a) Estimates of  $\ell(u)$  as a function of  $u$  on a logarithmic scale. (b) Empirical logarithmic rates as a function of  $u$ . (c) Empirical relative errors as a function of  $u$ .

**Example 3** (Light-tailed Scaling: Exponential). Let  $W_i \stackrel{i.i.d}{\sim} \text{Exp}(\mu)$  be the scaling random variable, with PDF

$$h(w_i) = \mu e^{-\mu w_i}, \quad w_i \geq 0,$$

and consider  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(k, \beta)$ , with PDF

$$f(x_i) = \frac{\beta^k x_i^{k-1} e^{-\beta x_i}}{(k-1)!}, \quad x_i \geq 0,$$

where  $i = 1, \dots, N$ .

The Erlang distribution is the same as a gamma distribution with the shape parameter being an integer.

(1) When conditioning on the scaling random variable, the resulting exponentially-twisted PDF is

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

where  $\mu_\theta < \mu$ . Hence, the likelihood ratio is given by

$$\frac{h(W_N)}{h_\theta(W_N)} = \frac{\mu}{\mu_\theta} e^{-(\mu-\mu_\theta)W_N}.$$

The choice of  $\mu_\theta$ , conditioning on  $N$ , is determined by solving  $\zeta'(\theta(N)) = \frac{Z_{N-1}^* \vee (u - S_{N-1})}{\mathbb{E}[X_N]}$  using  $\zeta(\theta(N)) = \ln \frac{\mu}{\mu - \theta(N)}$ , resulting in  $\theta(N) = \mu - \frac{\mathbb{E}[X_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ , and  $\mu_\theta = \frac{\mathbb{E}[X_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ . The likelihood ratio for (7), conditioning on  $N$ , is given by

$$\frac{h(W_N)}{h_\theta(W_N)} = \frac{(Z_{N-1}^* \vee (u - S_{N-1}))\beta\mu}{k} e^{-\left(\mu - \frac{k}{\beta(Z_{N-1}^* \vee (u - S_{N-1}))}\right)W_N}.$$

- (2) When conditioning on the PH random variable, the resulting PDF of the Erlang random variable is  $\text{Erlang}(k, \beta_\theta)$ . Then, the likelihood ratio is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \left(\frac{\beta}{\beta_\theta}\right)^k e^{-(\beta-\beta_\theta)X_N}.$$

We solve  $\zeta'(\theta(N)) = \frac{Z_{N-1}^* \vee (u - S_{N-1})}{\mathbb{E}[W_N]}$  using  $\zeta(\theta(N)) = k \ln \frac{\beta}{\beta - \theta(N)}$ , conditioning on  $N$ , resulting in  $\theta(N) = \beta - \frac{k\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ , and  $\beta_\theta = \frac{k\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ . The likelihood ratio for (10), conditioning on  $N$ , is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \left(\frac{(Z_{N-1}^* \vee (u - S_{N-1}))\beta\mu}{k}\right)^k e^{-\left(\beta - \frac{k}{\mu(Z_{N-1}^* \vee (u - S_{N-1}))}\right)X_N}.$$

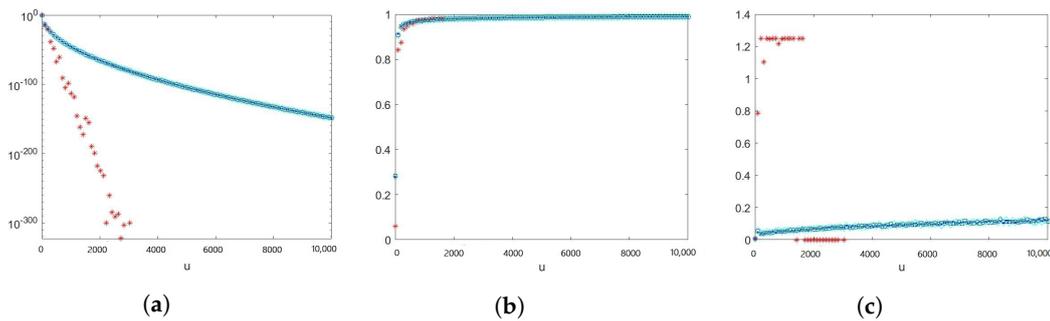
An alternative option is to perform importance sampling using the exponential PDF. Then, the likelihood ratio is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \frac{\beta^k X_N^{(k-1)} e^{-(\beta-\beta_\theta)X_N}}{\beta_\theta (k-1)!}.$$

We solve  $\zeta'(\theta(N)) = \frac{Z_{N-1}^* \vee (u - S_{N-1})}{\mathbb{E}[W_N]}$  using  $\zeta(\theta(N)) = \ln \frac{\beta}{\beta - \theta(N)}$ , conditioning on  $N$ , resulting in  $\theta(N) = \beta - \frac{\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ , and  $\beta_\theta = \frac{\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ . The likelihood ratio for (10), conditioning on  $N$ , is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \frac{(Z_{N-1}^* \vee (u - S_{N-1}))\beta^k \mu X_N^{(k-1)}}{(k-1)!} e^{-\left(\beta - \frac{1}{\mu(Z_{N-1}^* \vee (u - S_{N-1}))}\right)X_N}.$$

We take a sample size of  $10^5$ ,  $W_i \stackrel{i.i.d}{\sim} \text{Exp}(\mu)$ ,  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(k, \beta)$ , where  $i = 1, \dots, N$  and  $N \sim \text{Geo}(p)$  with  $\mu = 1$ ,  $k = 1$ ,  $\beta = 3$ , and  $p = 0.2$ . The results are shown in Figure 3. Proposition 1 does not provide a theoretical efficiency guarantee for this example as the number of summands is random, although the scaling random variable has a Weibullian tail. From numerical example results, we can observe that estimator  $\hat{\ell}_{\text{ConAK2+CV}}$  is not providing reliable estimates; however, estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  perform better.



**Figure 3.** The random variable  $N$  is geometrically distributed with success probability  $p = 0.2$ ,  $W_i \stackrel{i.i.d}{\sim} \text{Exp}(1)$ ,  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(1,3)$ , where  $i = 1, \dots, N$ . Black squares: ConAK1+IS+CV (conditioning and IS on exponential). Red stars: ConAK2+CV. Blue dots: ConAK2+IS+CV (conditioning and IS on Erlang using an exponential PDF). Cyan diamonds: ConAK2+IS+CV (conditioning and IS on Erlang using an Erlang PDF). (a) Estimates of  $\ell(u)$  as a function of  $u$  on a logarithmic scale. (b) Empirical logarithmic rates as a function of  $u$ . (c) Empirical relative errors as a function of  $u$ .

**Example 4 (Heavy-tailed Scaling: Weibull).** Let the scaling random variable  $W_i$  be Weibull-distributed, with PDF

$$h(w_i) = \frac{b}{a} \left(\frac{w_i}{a}\right)^{b-1} e^{-(w_i/a)^b}, \quad w_i \geq 0,$$

and consider  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(k, \beta)$ , where  $i = 1, \dots, N$ .

- (1) When conditioning on the scaling random variable, the resulting hazard-rate twisted PDF of the scaling random variable is

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

Thus, conditioning on  $N$ , and taking into account  $\mathbb{E}[X_N]$ , the likelihood ratio of a single element associated with this change of measure is

$$\frac{h(W_N)}{h_\theta(W_N)} = \frac{1}{1 - \theta(N)} e^{-\theta(N)\Lambda(W_N)}.$$

The likelihood ratio for (7), conditioning on  $N$ , is given by

$$\frac{h(W_N)}{h_\theta(W_N)} = -\ln \bar{H} \left( \frac{Z_{N-1}^* \vee (u - S_{N-1})}{\mathbb{E}[X_N]} \right) (\bar{H}(W_N))^{1 + \frac{1}{\ln \bar{H}((Z_{N-1}^* \vee (u - S_{N-1}))/\mathbb{E}[X_N])}}.$$

where  $\theta(N) = 1 - 1/\Lambda((Z_{N-1}^* \vee (u - S_{N-1}))/\mathbb{E}[X_N])$ . We note that from (11), we have  $\bar{H}_\theta(W_N) = [\bar{H}(W_N)]^{1-\theta}$ , and so it is straightforward to generate the twisted scaling random variables by using the inverse transform method.

- (2) When conditioning on the PH random variable, the resulting PDF of the Erlang random variable is  $\text{Erlang}(k, \beta_\theta)$ ; then, the likelihood ratio is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \left(\frac{\beta}{\beta_\theta}\right)^k e^{-(\beta-\beta_\theta)X_N}.$$

We solve  $\zeta'(\theta(N)) = \frac{Z_{N-1}^* \vee (u - S_{N-1})}{\mathbb{E}[W_N]}$  using  $\zeta(\theta(N)) = k \ln \frac{\beta}{\beta - \theta(N)}$ , conditioning on  $N$ , resulting in  $\theta(N) = \beta - \frac{k\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ , and  $\beta_\theta = \frac{k\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ . The likelihood ratio for (10), conditioning on  $N$ , is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \left( \frac{(Z_{N-1}^* \vee (u - S_{N-1}))\beta}{ka\Gamma(1 + 1/b)} \right)^k e^{-\left(\beta - \frac{ka\Gamma(1+1/b)}{(Z_{N-1}^* \vee (u - S_{N-1}))}\right)X_N}.$$

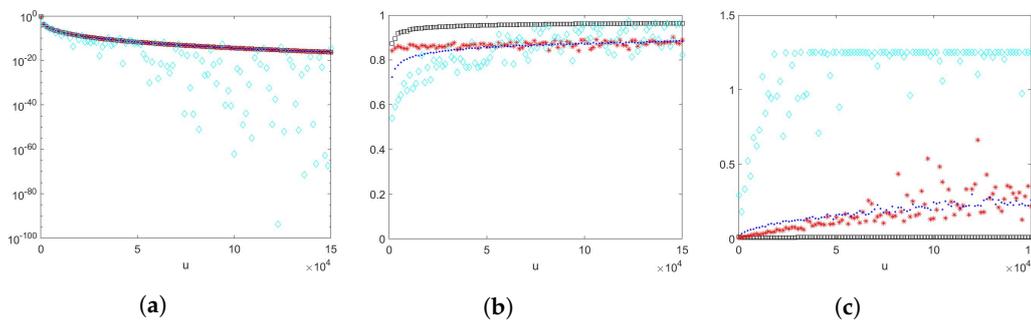
An alternative option is to perform importance sampling using the exponential PDF. Then, the likelihood ratio is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \frac{\beta^k X_N^{(k-1)} e^{-(\beta-\beta_\theta)X_N}}{\beta_\theta (k-1)!}.$$

We solve  $\zeta'(\theta(N)) = \frac{Z_{N-1}^* \vee (u - S_{N-1})}{\mathbb{E}[W_N]}$  using  $\zeta(\theta(N)) = \ln \frac{\beta}{\beta - \theta(N)}$ , conditioning on  $N$ , resulting in  $\theta(N) = \beta - \frac{\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ , and  $\beta_\theta = \frac{\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ . The likelihood ratio for (10), conditioning on  $N$ , is given by

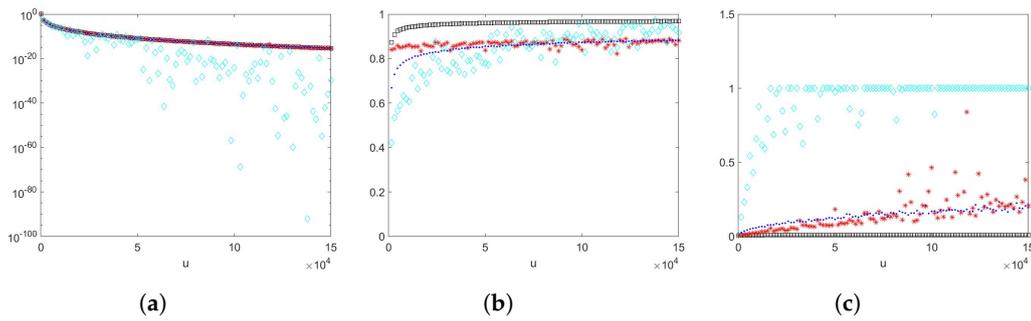
$$\frac{f(X_N)}{f_\theta(X_N)} = \frac{(Z_{N-1}^* \vee (u - S_{N-1})) \beta^k X_N^{(k-1)}}{a \Gamma(1 + 1/b) (k-1)!} e^{-\left(\beta - \frac{a \Gamma(1+1/b)}{(Z_{N-1}^* \vee (u - S_{N-1}))}\right) X_N}.$$

We take a sample size of  $10^5$ ,  $W_i \stackrel{i.i.d}{\sim} \text{Weibull}(b, a)$ ,  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(k, \beta)$ , where  $i = 1, \dots, N$ , and  $N \sim \text{Geo}(p)$  with scale parameter  $a = 2$ , shape parameter  $b = 0.35$ ,  $k = 2$ ,  $\beta = 3$ , and  $p = 0.2$ . The results are shown in Figure 4. Proposition 1 does not provide a theoretical efficiency guarantee for this example as the number of summands is random, although the scaling random variable has a Weibullian tail. From numerical example results, we can observe that estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling density using an exponential PDF) perform well. However, estimators  $\hat{\ell}_{\text{ConAK2+CV}}$  and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling density using an Erlang PDF) appear to fluctuate more when  $u$  becomes larger, and they have increasing empirical relative errors.



**Figure 4.** The random variable  $N$  is geometrically distributed with success probability  $p = 0.2$ ,  $W_i \stackrel{i.i.d}{\sim} \text{Weibull}(0.35, 2)$ ,  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(2, 3)$ , where  $i = 1, \dots, N$ . Black squares: ConAK1+IS+CV (conditioning and IS on Weibull). Red stars: ConAK2+CV (conditioning on Erlang). Blue dots: ConAK2+IS+CV (conditioning and IS on Erlang using an exponential PDF). Cyan diamonds: ConAK2+IS+CV (conditioning and IS on Erlang using an Erlang PDF). (a) Estimates of  $\ell(u)$  as a function of  $u$  on a logarithmic scale. (b) Empirical logarithmic rates as a function of  $u$ . (c) Empirical relative errors as a function of  $u$ .

Next, we change random variable  $N$  from geometrically distributed to deterministic, and keep the rest of the parameters the same as above. The results are shown in Figure 5. Proposition 1 suggests that the estimator  $\hat{\ell}_{\text{ConAK2+CV}}$  is logarithmically efficient as this example satisfies the second case of Proposition 1. From numerical example results, we can observe that estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling density using an exponential PDF) perform well. However, estimators  $\hat{\ell}_{\text{ConAK2+CV}}$  and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling density using an Erlang PDF) appear to fluctuate more when  $u$  becomes larger, and they have increasing empirical relative errors.



**Figure 5.** The number of summands  $N = 30$ ,  $W_i \stackrel{i.i.d}{\sim} \text{Weibull}(0.35, 2)$ ,  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(2, 3)$ , where  $i = 1, \dots, N$ . Black squares: ConAK1+IS+CV (conditioning and IS on Weibull). Red stars: ConAK2+CV (conditioning on Erlang). Blue dots: ConAK2+IS+CV (conditioning and IS on Erlang using an exponential PDF). Cyan diamonds: ConAK2+IS+CV (conditioning and IS on Erlang using an Erlang PDF). (a) Estimates of  $\ell(u)$  as a function of  $u$  on a logarithmic scale. (b) Empirical logarithmic rates as a function of  $u$ . (c) Empirical relative errors as a function of  $u$ .

**Example 5 (Heavy-tailed Scaling: Pareto).** Let the scaling random variable  $W_i$  be Pareto-distributed, with PDF

$$h(w_i) = \frac{\delta}{w_i^{\delta+1}}, \quad w_i \geq 1,$$

and consider  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(k, \beta)$ , where  $i = 1, \dots, N$ . Note that Pareto distributions  $\in \mathcal{D} \cap \mathcal{L}$ ; see [5].

- (1) When conditioning on the scaling random variable, the resulting hazard-rate twisted PDF of the scaling random variable is

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

Hence, the likelihood ratio is given by

$$\frac{h(W_N)}{h_\theta(W_N)} = \frac{1}{1-\theta} W_N^{-\theta\delta}.$$

Thus, conditioning on  $N$ , we set  $\theta(N) = 1 - 1/\Lambda((Z_{N-1}^* \vee (u - S_{N-1}))/\mathbb{E}[X_N])$ . Since  $\Lambda(W_N) = \delta \ln(W_N)$  in this case, the likelihood ratio of single element for (7), conditioning on  $N$ , is

$$\frac{h(W_N)}{h_\theta(W_N)} = \left( \delta \ln \left( \frac{(Z_{N-1}^* \vee (u - S_{N-1}))\beta}{k} \right) \right) W_N^{-(\delta-1/\ln((Z_{N-1}^* \vee (u - S_{N-1}))\beta/k))}.$$

- (2) When conditioning on the PH random variable, the resulting PDF of the Erlang random variable is  $\text{Erlang}(k, \beta_\theta)$ ; then, the likelihood ratio is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \left( \frac{\beta}{\beta_\theta} \right)^k e^{-(\beta-\beta_\theta)X_N}.$$

We solve  $\zeta'(\theta(N)) = \frac{Z_{N-1}^* \vee (u - S_{N-1})}{\mathbb{E}[W_N]}$  using  $\zeta(\theta(N)) = k \ln \frac{\beta}{\beta - \theta(N)}$ , conditioning on  $N$ , resulting in  $\theta(N) = \beta - \frac{k\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ , and  $\beta_\theta = \frac{k\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ . The likelihood ratio for (10), conditioning on  $N$ , is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \left( \frac{(Z_{N-1}^* \vee (u - S_{N-1}))\beta(\delta - 1)}{k\delta} \right)^k e^{-\left( \beta - \frac{k\delta}{(Z_{N-1}^* \vee (u - S_{N-1}))(\delta - 1)} \right) X_N}.$$

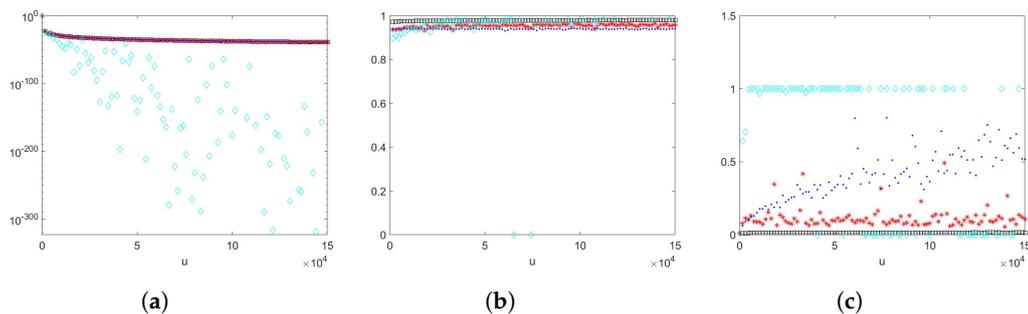
An alternative option is to perform importance sampling using the exponential PDF. Then, the likelihood ratio for is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \frac{\beta^k X_N^{(k-1)} e^{-(\beta-\beta_\theta)X_N}}{\beta_\theta (k-1)!}.$$

We solve  $\zeta'(\theta(N)) = \frac{Z_{N-1}^* \vee (u - S_{N-1})}{\mathbb{E}[W_N]}$  using  $\zeta(\theta(N)) = \ln \frac{\beta}{\beta - \theta(N)}$ , conditioning on  $N$ , resulting in  $\theta(N) = \beta - \frac{\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ , and  $\beta_\theta = \frac{\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ . The likelihood ratio for (10), conditioning on  $N$ , is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \frac{(Z_{N-1}^* \vee (u - S_{N-1})) \beta^k (\delta - 1) X_N^{(k-1)}}{\delta (k-1)!} e^{-\left(\beta - \frac{\delta}{(Z_{N-1}^* \vee (u - S_{N-1}))^{(\delta-1)}}\right) X_N}.$$

We take a sample size of  $10^5$ ,  $W_i \stackrel{i.i.d}{\sim} \text{Pareto}(\delta)$ ,  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(k, \beta)$ , where  $i = 1, \dots, N$ , and  $N \sim \text{Geo}(p)$  with  $\delta = 8$ ,  $k = 2$ ,  $\beta = 3$ , and  $p = 0.2$ . The results are shown in Figure 6. Proposition 1 suggests that the estimator  $\hat{\ell}_{\text{ConAK2+CV}}$  will have bounded relative error, since  $\text{Pareto} \in \mathcal{D} \cap \mathcal{L}$ . From numerical example results, we can observe that the estimator  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  performs well. The estimator  $\hat{\ell}_{\text{ConAK2+CV}}$  has a few larger empirical relative error estimates as  $u$  is varied. Estimators  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling density using an Erlang PDF and using an exponential PDF) appear to fluctuate more when  $u$  becomes larger, and they have increasing empirical relative errors.



**Figure 6.** The random variable  $N$  is geometrically distributed with success probability  $p = 0.2$ ,  $W_i \stackrel{i.i.d}{\sim} \text{Pareto}(8)$ ,  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(2, 3)$ , where  $i = 1, \dots, N$ . Black squares: ConAK1+IS+CV (conditioning and IS on Pareto). Red stars: ConAK2+CV (conditioning on Erlang). Blue dots: ConAK2+IS+CV (conditioning and IS on Erlang using an exponential PDF). Cyan diamonds: ConAK2+IS+CV (conditioning and IS on Erlang using an Erlang PDF). (a) Estimates of  $\ell(u)$  as a function of  $u$  on a logarithmic scale. (b) Empirical logarithmic rates as a function of  $u$ . (c) Empirical relative errors as a function of  $u$ .

**Example 6 (Heavy-tailed Scaling: Log-normal).** Consider now  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(k, \beta)$ , and the scaling random variable  $W_i \stackrel{i.i.d}{\sim} \text{LogN}(\mu, \sigma^2)$  with PDF

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

with corresponding CDF

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

where  $i = 1, \dots, N$ , and  $\Phi(\cdot)$  is the CDF of a standard normal random variate.

(1) When conditioning on the scaling random variable, the resulting hazard-rate twisted PDF of the scaling random variable is

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

The likelihood ratio is given by

$$\frac{h(W_N)}{h_\theta(W_N)} = \frac{1}{1 - \theta(N)} [\overline{H}(W_N)]^{\theta(N)},$$

where  $\theta(N) = 1 - 1/\Lambda((Z_{N-1}^* \vee (u - S_{N-1}))/\mathbb{E}[X_N])$ . The likelihood ratio for (7), conditioning on  $N$ , is given by

$$\frac{h(W_N)}{h_\theta(W_N)} = -\ln \overline{H}\left(\frac{Z_{N-1}^* \vee (u - S_{N-1})}{\mathbb{E}[X_N]}\right) (\overline{H}(W_N))^{1 + \frac{1}{\ln \overline{H}((Z_{N-1}^* \vee (u - S_{N-1}))/\mathbb{E}[X_N])}}.$$

where  $\theta(N) = 1 - 1/\Lambda((Z_{N-1}^* \vee (u - S_{N-1}))/\mathbb{E}[X_N])$ .

We note that from (11), we have  $\overline{H}_\theta(W_N) = [\overline{H}(W_N)]^{1-\theta}$ , and so it is straightforward to generate the twisted scaling random variables by using the inverse transform method.

- (2) When conditioning on the PH random variable, the resulting PDF of the Erlang random variable is  $\text{Erlang}(k, \beta_\theta)$ ; then, the likelihood ratio is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \frac{f(X_N)}{f_\theta(X_N)} = \left(\frac{\beta}{\beta_\theta}\right)^k e^{-(\beta - \beta_\theta)X_N}.$$

We solve  $\zeta'(\theta(N)) = \frac{Z_{N-1}^* \vee (u - S_{N-1})}{\mathbb{E}[W_N]}$  using  $\zeta(\theta(N)) = k \ln \frac{\beta}{\beta - \theta(N)}$ , conditioning on  $N$ , resulting in  $\theta(N) = \beta - \frac{k\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ , and  $\beta_\theta = \frac{k\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ . The likelihood ratio for (10), conditioning on  $N$ , is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \left(\frac{(Z_{N-1}^* \vee (u - S_{N-1}))\beta}{k\mathbb{E}[W_N]}\right)^k e^{-\left(\beta - \frac{k\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}\right)X_N}.$$

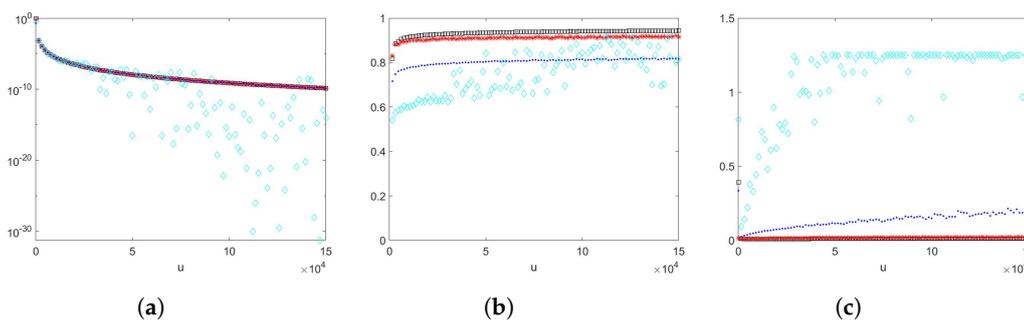
An alternative option is to perform importance sampling using the exponential PDF. Then, the likelihood ratio is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \frac{\beta^k X_N^{(k-1)} e^{-(\beta - \beta_\theta)X_N}}{\beta_\theta (k-1)!}.$$

We solve  $\zeta'(\theta(N)) = \frac{Z_{N-1}^* \vee (u - S_{N-1})}{\mathbb{E}[W_N]}$  using  $\zeta(\theta(N)) = \ln \frac{\beta}{\beta - \theta(N)}$ , conditioning on  $N$ , resulting in  $\theta(N) = \beta - \frac{\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ , and  $\beta_\theta = \frac{\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}$ . The likelihood ratio for (10), conditioning on  $N$ , is given by

$$\frac{f(X_N)}{f_\theta(X_N)} = \frac{(Z_{N-1}^* \vee (u - S_{N-1}))\beta^k X_N^{(k-1)}}{\mathbb{E}[W_N](k-1)!} e^{-\left(\beta - \frac{\mathbb{E}[W_N]}{Z_{N-1}^* \vee (u - S_{N-1})}\right)X_N}.$$

We take a sample size of  $10^5$ ,  $W_i \stackrel{i.i.d}{\sim} \text{LogN}(\mu, \sigma^2)$ ,  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(k, \beta)$ , where  $i = 1, \dots, N$ , and  $N \sim \text{Geo}(p)$  with  $\mu = 2$ ,  $\sigma = 1.5$ ,  $k = 2$ ,  $\beta = 3$ , and  $p = 0.2$ . The results are shown in Figure 7. Proposition 1 does not provide a theoretical efficiency guarantee for this example as the log-normal distribution is neither  $\mathcal{D} \cap \mathcal{L}$  nor has a Weibullian tail. From numerical example results, we can observe that estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$ ,  $\hat{\ell}_{\text{ConAK2+CV}}$ , and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling density using an exponential PDF) perform well.  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling density using an exponential PDF) turns out to have slightly lower empirical logarithmic rates and higher empirical relative errors than estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  and  $\hat{\ell}_{\text{ConAK2+CV}}$ . The estimator  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling density using an Erlang PDF) appears to fluctuate more when  $u$  becomes larger, and it has increasing empirical relative errors.



**Figure 7.** The random variable  $N$  is geometrically distributed with success probability  $p = 0.2$ ,  $W_i \stackrel{i.i.d}{\sim} \text{LogN}(2, 2.25)$ ,  $X_i \stackrel{i.i.d}{\sim} \text{Erlang}(2, 3)$ , where  $i = 1, \dots, N$ . Black squares: ConAK1+IS+CV (conditioning and IS on log-normal). Red stars: ConAK2+CV (conditioning on Erlang). Blue dots: ConAK2+IS+CV (conditioning and IS on Erlang using an exponential PDF). Cyan diamonds: ConAK2+IS+CV (conditioning and IS on Erlang using an Erlang PDF). (a) Estimates of  $\ell(u)$  as a function of  $u$  on a logarithmic scale. (b) Empirical logarithmic rates as a function of  $u$ . (c) Empirical relative errors as a function of  $u$ .

### 6. 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. We combined Asmussen–Kroese estimation with the conditional Monte Carlo method and further exploited importance sampling applied to the last term of the summands. 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.

When combining Asmussen–Kroese estimation with the conditional Monte Carlo method, we can either condition on the scaling random variable or PH random variable. The estimators are denoted as  $\hat{\ell}_{\text{ConAK1}}$  and  $\hat{\ell}_{\text{ConAK2}}$ , respectively. In the given examples, we compared the estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$ ,  $\hat{\ell}_{\text{ConAK2+CV}}$ ,  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling using an exponential PDF), and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling using an Erlang PDF). We took the distribution of  $N$  to be geometrically distributed in most cases, with one extra example for  $N$  as deterministic.

We observed that in Example 3, where the scaling random variable was exponentially distributed, the estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  with importance sampling density being either exponential or Erlang gave sharp estimates and empirical logarithmic rates which tended to 1 as  $u$  became large, and also had lower empirical relative errors than estimator  $\hat{\ell}_{\text{ConAK2+CV}}$ . The estimator  $\hat{\ell}_{\text{ConAK2+CV}}$  performed poorly due to the calculation capacity issue, as we observed that when  $u$  exceeded around 1500, the empirical relative errors were 0 in Figure 3c. The poor performance of estimator  $\hat{\ell}_{\text{ConAK2+CV}}$  is not surprising as Proposition 1 does not provide theoretical efficiency guarantees when the scaling random variable has a Weibullian tail and the number of summands is random.

In Example 4, where the scaling random variable was Weibull-distributed, we took the number of summands  $N$  being geometrically distributed or being deterministic. The results are summarised in Figures 4 and 5. We observed that all the estimators in both examples performed similar. The estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$ ,  $\hat{\ell}_{\text{ConAK2+CV}}$ , and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling using an exponential PDF) gave sharp estimates and empirical logarithmic rates which tended to 1 as  $u$  became large, and also had lower empirical relative errors than estimator  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  with importance sampling density being Erlang. Despite positive indications from the empirical logarithmic efficiency estimates, the empirical relative errors for the estimator  $\hat{\ell}_{\text{ConAK2+CV}}$  had an increasing trend and were fluctuating. Among the other two better-performing estimators, the estimator  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  had the lowest empirical relative errors, and empirical logarithmic rates which tended to 1 faster.

In Example 5, where the scaling random variable was Pareto-distributed, the estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$ ,  $\hat{\ell}_{\text{ConAK2+CV}}$ , and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling using an

exponential PDF) gave sharp estimates and empirical logarithmic rates which tended to 1 as  $u$  became large, and also had lower empirical relative errors than estimator  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  with importance sampling density being Erlang. Despite positive indications from the empirical logarithmic efficiency estimates, the empirical relative errors for the estimator  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling using an exponential PDF) had an increasing trend and were fluctuating. Among the other two better-performing estimators, the estimator  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  had the lowest empirical relative errors, and empirical logarithmic rates which tended to 1 faster.

In Example 6, where the scaling random variable was log-normal-distributed, we observed similar trends as in Example 5. In this example, we observed that the estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$ ,  $\hat{\ell}_{\text{ConAK2+CV}}$ , and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling using an exponential PDF) gave sharp estimates. Empirical logarithmic rates of all the estimators appeared to tend to 1 as  $u$  became large. The estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$ ,  $\hat{\ell}_{\text{ConAK2+CV}}$ , and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling using an exponential PDF) had lower empirical relative errors than estimator  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  with importance sampling density being Erlang. The estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  and  $\hat{\ell}_{\text{ConAK2+CV}}$  performed similarly, and showed better performance with empirical relative errors closer to 0, and higher empirical logarithmic rates in this example.

In all the above examples, the estimator  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling using an exponential PDF) appeared to perform better than the estimator  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling using an Erlang PDF), with sharp estimates, empirical logarithmic rates closer to 1, and lower empirical relative errors.

Finally, the examples given suggest that the estimator  $\hat{\ell}_{\text{ConAK2+CV}}$  performs well when the scaling random variable has a heavier tail. The efficiency study suggests that  $\hat{\ell}_{\text{ConAK1}}$  and  $\hat{\ell}_{\text{ConAK2}}$  will attain the same efficiency as the Asmussen–Kroese estimator for the classes of distributions in Proposition 1. However, when conditioning on the scaling random variable, we do not always obtain reliable estimates due to the numerical limitations. In this case, the numerical experiments suggest that further exploiting importance sampling on the scaling random variable of the last term of the summands will provide more reliable estimates. For the log-normal case which does not fall in the cases in Proposition 1, we do not apply Proposition 1 to conclude theoretical efficiency for the scaling random variable being log-normal-distributed. However, the numerical experiment suggests that the estimator  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  performs the best.

In conclusion, numerical studies suggest that, in general, the estimator  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  will provide more reliable estimates when estimating tail probabilities of random sums of PH scale mixture random variables. When the scaling random variables have much heavier tail (e.g., log-normal case), the estimator  $\hat{\ell}_{\text{ConAK2+CV}}$  performs better than in other cases. For all cases, the estimator  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  (importance sampling using an exponential PDF) appears to provide more reliable and stable estimates than using an Erlang PDF, with empirical logarithmic rates which tend to 1 as  $u$  becomes large, and relatively low empirical relative errors.

At present, the importance sampling densities and related parameters are chosen heuristically to ensure that the probabilities of interest are no longer rare under these changes of measure. One key question which we did not address in this work is how to choose these densities and parameters in a principled and ideally asymptotically optimal way. We studied the efficiency of conditional Asmussen–Kroese estimators. Proof of the theoretical efficiency properties of the proposed estimators  $\hat{\ell}_{\text{ConAK1+IS+CV}}$  and  $\hat{\ell}_{\text{ConAK2+IS+CV}}$  is a subject for future research. Furthermore, in the present work, we assumed for simplicity that all of the random variables in the sum are independent, and PH random variables were chosen to be Erlang-distributed. An interesting avenue for future work is to develop effective simulation methods for this problem when there is structured dependence between the random variables. The PH assumption can be also relaxed as long as the product of the two random variables is one of  $\mathcal{D} \cap \mathcal{L}$ , Weibull with appropriate shape parameter, or is log-normal.

**Author Contributions:** Conceptualisation, H.Y. and T.T.; methodology, H.Y. and T.T.; software, H.Y.; validation, H.Y. and T.T.; formal analysis, H.Y. and T.T.; investigation, H.Y.; resources, H.Y.; writing—original draft preparation, H.Y.; writing—review and editing, H.Y. and T.T.; visualisation, H.Y. and T.T.; supervision, T.T.; project administration, H.Y. and T.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All code for the numerical experiments found in this paper is available at <https://github.com/HY9412/Estimating-Tail-Probabilities-of-Random-Sums-of-Phase-type-Scale-Mixture-Random-Variables> (accessed on 5 August 2022).

**Acknowledgments:** The authors acknowledge the valuable input of Leonardo Rojas-Nandayapa on an earlier version of this manuscript. The authors thank the reviewers for their comments which improved the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Asmussen, S.; Albrecher, H. *Ruin Probabilities*, 2nd ed.; World Scientific Publishing: Singapore, 2010.
2. Asmussen, S. *Applied Probabilities and Queues*, 2nd ed.; Springer: New York, NY, USA, 2003.
3. Asmussen, S.; Nerman, O.; Olsson, M. Fitting phase-type distributions via the EM algorithm. *Scand. J. Stat.* **1996**, *23*, 419–441.
4. Bladt, M.; Rojas-Nandayapa, L. Fitting phase-type scale mixtures to heavy-tailed data and distributions. *Extremes* **2018**, *21*, 285–313. [\[CrossRef\]](#)
5. Foss, S.; Korshunov, D.; Zachary, S. *An Introduction to Heavy-Tailed and Subexponential Distributions*, 2nd ed.; Springer: New York, NY, USA, 2013.
6. Asmussen, S.; Binswanger, K.; Højgaard, B. Rare Events Simulation for Heavy-tailed Distributions. *Bernoulli* **2000**, *6*, 303–322. [\[CrossRef\]](#)
7. Asmussen, S.; Kroese, D. Improved algorithms for rare event simulation with heavy tails. *Adv. Appl. Probab.* **2006**, *38*, 545–558. [\[CrossRef\]](#)
8. Juneja, S.; Shahabuddin, P. Simulating Heavy Tailed Processes Using Delayed Hazard Rate Twisting. *ACM Trans. Model. Comput. Simul. (TOMACS)* **2002**, *12*, 94–118. [\[CrossRef\]](#)
9. Keilson, J.; Steutel, F. Mixtures of distributions, moment inequalities and measures of exponentiality and normality. *Ann. Probab.* **1974**, *2*, 112–130. [\[CrossRef\]](#)
10. Bladt, M.; Nielsen, B.F.; Samorodnitsky, G. Calculation of Ruin Probabilities for a Dense Class of Heavy Tailed Distributions. *Scand. Actuar. J.* **2015**, *2015*, 573–591. [\[CrossRef\]](#)
11. Rojas-Nandayapa, L.; Xie, W. Asymptotic tail behaviour of phase-type scale mixture distributions. *Ann. Actuar. Sci.* **2018**, *12*, 412–432. [\[CrossRef\]](#)
12. Neuts, M. *Probability Distributions of Phase Type*; Department of Mathematics, University of Louvain: Ottignies-Louvain-la-Neuve, Belgium, 1975; pp. 173–206.
13. Hartinger, J.; Kortschak, D. On the Efficiency of the Asmussen–Kroese-estimator and its Application to Stop-loss Transforms. *Blätter der DGVM* **2009**, *30*, 363–377. [\[CrossRef\]](#)
14. Siegmund, D. Importance Sampling in the Monte Carlo Study of Sequential Tests. *Ann. Stat.* **1976**, *4*, 673–684. [\[CrossRef\]](#)
15. Kroese, D.; Taimre, T.; Botev, Z. *Handbook of Monte Carlo Methods*, 1st ed.; John Wiley and Sons: Hoboken, NJ, USA, 2011.
16. Arendarczyk, M.; Dębicki, K. Asymptotics of supremum distribution of a Gaussian process over a Weibullian time. *Bernoulli* **2011**, *17*, 194–210. [\[CrossRef\]](#)
17. Su, C.; Chen, Y. On the behavior of the product of independent random variables. *Sci. China* **2006**, *A49*, 342–359. [\[CrossRef\]](#)
18. O’Cinneide, C.A. Phase-type Distributions and Majorization. *Ann. Appl. Probab.* **1991**, *1*, 219–227. [\[CrossRef\]](#)