

Efficient Rare Event Simulation for Heavy-tailed Compound Sums

Jose Blanchet and Chenxin Li
Harvard University

January 24, 2008

Abstract

We develop an efficient importance sampling algorithm for estimating the tail distribution of heavy-tailed compound sums, i.e. random variables of the form $S_M = Z_1 + \dots + Z_M$ where the Z_i 's are i.i.d. random variables in \mathbb{R} and M is a non-negative, integer-valued random variable independent of the Z_i 's. We construct the first estimator that can be rigorously shown to be strongly efficient only under the assumption that the Z_i 's are subexponential and M is light-tailed. Our estimator is based on state-dependent importance sampling and we use Lyapunov-type inequalities to control its second moment. The performance of our estimator is empirically illustrated in various instances involving popular heavy-tailed models.

1 Introduction

Let $(Z_k : k \geq 1)$ be a sequence of independent and identically distributed (i.i.d.) random variables (r.v.'s) in \mathbb{R} and let M be a r.v. independent of the Z_k 's taking values on non-negative integers. The r.v. $S_M = \sum_{k=1}^M Z_k$ is called a compound sum and it figures prominently in a variety of stochastic modeling contexts. For instance, it is well known that the steady-state waiting time of an M/G/1 queue can be represented as a geometric compound sum, i.e., M is geometrically distributed. The text by Kalashnikov (1997) is dedicated entirely to the analysis of geometric compound sums and their applications. Other types of sums that are often used in insurance applications are the so-called Polya sums, in which M follows a negative binomial distribution (see, for instance, Willmot and Lin (2001) and Embrechts, Klüppelberg and Mikosch (1997)). Polya sums arise when modeling

a sequence of rewards / costs associated to a Poisson stream of arrivals during a fixed time interval with random (gamma distributed) arrival rate. The algorithms developed here, in particular, can be used to efficiently estimate the tail of heavy-tailed Polya sums.

Our goal is to develop efficient simulation methodologies for estimating

$$u(b) = \mathbb{P}(S_M > b)$$

as $b \nearrow \infty$. Estimating $u(b)$ to a high relative accuracy via simulation is challenging because the event $\{S_M > b\}$ becomes rare as $b \nearrow \infty$, i.e., $u(b) \searrow 0$ as $b \nearrow \infty$. A performance measure that is widely used in rare-event simulation is that of *strong efficiency* (see Section 3), which in our context requires the coefficient of variation of an unbiased estimator for $u(b)$ to be uniformly bounded over b . Assuming basically that M has exponentially decaying tails and that the Z_k 's are subexponential we construct such a strongly efficient estimator.

Rare-event simulation methodology for heavy-tailed compound sums has been studied substantially in the literature. Intuitively, since M possesses light tails, large excedences of S_M are a lot more likely to be caused by the Z_k 's than by M . This statement can be given precise mathematical substance, see Embrechts et al (1997), p.450 for a result that covers geometric sums of the M/G/1 type. Consequently, the literature has primarily focused on developing efficient methodology for estimating the tail probability $\mathbb{P}(S_n > b)$ for fixed n , as $b \nearrow \infty$. The first provable efficient estimator for $\mathbb{P}(S_n > b)$ (in the regularly varying case, which basically means power-law type tails) was given by Asmussen and Binswanger (1997) and it was based on Conditional Monte Carlo (CMC). See also the paper by Asmussen and Kroese (2006) for refined CMC algorithms. Other algorithms, for instance, Juneja and Shahabuddin (2002), Boots and Shahabuddin (2001) and Dupuis, Leder and Wang (2007), have been proposed which are based on various importance sampling ideas. All the algorithms that have been developed so far can be proved to be efficient only for certain types of heavy-tailed distributions such as regularly varying or some cases of Weibull tails. We present a state-dependent importance sampling algorithm for compound sums that can be proved to be strongly efficient assuming subexponential increments. Subexponential r.v.'s in particular include Weibull tails, regularly varying distributions, lognormals and many more other distributions as special cases. This is the first algorithm that can be shown to achieve such performance under almost minimal conditions. We shall present numerical results that support the excellent performance of our estimator and compare against some of the estimators mentioned above. The mathematical

techniques behind the performance analysis of our state-dependent algorithms are based on Lyapunov-type inequalities, similar to those introduced in Blanchet and Glynn (2007) for the G/G/1 queue.

As we mentioned before, our proposed simulation algorithms is based on importance sampling. It is well known (Bucklew (2004), p.61) that the zero-variance change-of-measure (which is the optimal choice of importance sampler) corresponds to sampling S_M from its conditional distribution given that $S_M > b$. In our compound sums context, this optimal change-of-measure is also linked to the so-called Doob's h -transform (see Section 2). That being said, this scheme requires knowledge of $u(b)$ for its specification and therefore is not available for importance sampling. Nevertheless, it still suggests using an importance sampler similar to the optimal choice, which admits a Markovian characterization based on an integral equation that is satisfied by $u(\cdot)$ and is related to the so-called harmonic property. This characterization suggests a natural way to mimic the behavior of the zero-variance importance sampler if one can provide an approximation to $u(\cdot)$ – an idea that has been empirically studied in the literature in different settings, see for instance Juneja and Shahabuddin (2006). Such approximation, fortunately, is available from the theory of subexponential approximations (see, Embrechts et al (1997), p. 45). To be more specific, we have that (under our current assumptions)

$$u(b) \sim \mathbb{P}(Z_1 > b) \cdot \mathbb{E}M$$

as $b \nearrow \infty$. Our algorithm, therefore, enhances the previous approximation because it allows to increase its accuracy in the prelimit, i.e. for finite b , by combining it with a computational procedure that is efficient regardless of the size of b .

We mentioned before that our importance sampler is constructed to mimic the behavior of the optimal change-of-measure through a Markovian description. In order to prove strong efficiency we need to bound the second moment of our likelihood ratio, which in turn is the product of the local likelihood ratios applied to each of the increments. Our strategy is to develop a system of linear inequalities that can be used to bound the second moment of our estimator (as a function of b). In particular, a solution to such system of linear inequalities is called a Lyapunov function and it provides the required upper bound. We are able to find a Lyapunov function with the appropriate asymptotic behavior as $b \nearrow \infty$, which enables us to show strong efficiency.

The rest of the paper is organized as follows. In Section 2 we provide a harmonic representation of the zero-variance importance sampling distribu-

tion and review the notion of efficiency in rare-event simulation. In Section 3 we describe a natural strategy to construct efficient state-dependent importance sampling algorithms and means to bound its second order moment. In Section 4, a specific Lyapunov function is constructed for the problem of compound sums with heavy-tailed increments. This is the major contribution of our paper. Readers who focus more on the empirical applications can skip right ahead to Section 5, where an implementable algorithm and numerical examples are discussed.

2 Zero-variance change-of-measure

A compound sum can be viewed as a special type of Markov chains and $(u(b) : b \in \mathbb{R})$ can be embedded into a sequence of functions that satisfy certain harmonic property. More precisely, consider a Markov chain $S = (S_n : n \geq 0)$ with transition kernel $P_n(s_n, ds_{n+1}) = \mathbb{P}(Z_1 \in ds_{n+1} - s_n)$ and $S_0 = 0$, then S is just the random walk generated by the Z_k 's, i.e., $S_n = Z_1 + Z_2 + \dots + Z_n$. Now, let $u_k(x)$ be defined as

$$u_k(x; b) \triangleq u_k(x) = \mathbb{P}(S_{M-k} > b - x; M \geq k). \quad (1)$$

Then, $u(b)$ is embedded in (1) so that $u(b) = u_0(0; b) = u_0(0)$ (we will omit the explicit dependence on b when referring to the u_k 's in the rest of paper). It follows that $(u_k(x) : x \in \mathbb{R}, k \geq 0)$ satisfies

$$\begin{aligned} u_k(x) &= \mathbb{P}(S_0 > b - x; M = k) + \mathbb{P}(S_{M-k} > b - x; M \geq k + 1) \\ &= \int_{-\infty}^{\infty} P_k(x, dy) \cdot \mathbb{P}(S_{M-k-1} + y - x > b - x; M \geq k + 1) \\ &= \int_{-\infty}^{\infty} P_k(x, dy) \cdot u_{k+1}(y), \quad \forall k \in \mathbb{N}, x < b \end{aligned} \quad (2)$$

with boundary conditions $u_k(x) = \mathbb{P}(M \geq k) \triangleq \bar{F}_M(k), \forall x \geq b$. Equation (2) displays a fundamental relation satisfied by the u_k 's. To be more specific, we have the following general definition:

Definition 1 *Let $X = (X_n : n \geq 0)$ be a time inhomogeneous Markov chain on $(\Omega, \mathcal{F}, \mathbb{P})$, we call the sequence $(h_k(\cdot) : k \geq 0)$ space-time superharmonic in a set $D \subset \Omega$ if it satisfies*

$$h_k(x) \geq \mathbb{E}[h_{k+1}(X_{k+1}) | X_k = x] \quad \text{for } \forall x \in D. \quad (3)$$

$(h_k(\cdot) : k \geq 0)$ is called space-time harmonic if the equality holds in (3).

Therefore, $\{u_k(x)\}$ can then be viewed as a space-time harmonic sequence with set $D = \{x : x < b\}$. We will use the notation $\mathbb{E}_{x,k}(\cdot) \triangleq \mathbb{E}[\cdot | X_k = x]$ and $\mathbb{P}_{x,k}(\cdot) \triangleq \mathbb{P}(\cdot | X_k = x)$ throughout the rest of the paper.

It is well known, thanks to the work of Doob (1983), that the connection between (super)harmonic functions and (super)martingales often provides useful probabilistic representations of (super)harmonic functions. For instance, for the harmonic sequence of u_k 's, we have the following lemma:

Lemma 1 *The elements of the harmonic sequence $(u_k(\cdot) : k \geq 0)$ can be represented as*

$$u_k(x) = \mathbb{E}_{x,k}[\bar{F}_M(T_k(x))I(T_k(x) < \infty)] \quad (4)$$

where $T_k(x) = \inf\{t \geq k : S_t \geq b | S_k = x\}$ is the stopping time of the Markov chain $\{S_n\}$ to set D^c .

Proof. We fix $X_k = x$ and define $\widetilde{M}_{n \wedge T_k(x)} = u_{n \wedge T_k(x)}(S_{n \wedge T_k(x)})$ for $n \geq k$, where $a \wedge b \triangleq \min(a, b)$. Then, from the harmonic property $u_k(x) = \mathbb{E}_{x,k}[u_{k+1}(S_{k+1})]$, it easily follows that $(\widetilde{M}_n : n \geq k)$ is a martingale with respect to the filtration generated by the Markov chain $(S_n : n \geq k)$. In addition, notice that $\widetilde{M}_n \leq 1$, so we have that $\{\widetilde{M}_n\}$ is uniformly bounded and therefore uniformly integrable. By the Martingale Convergence Theorem, \widetilde{M}_n converges almost surely in L^1 , therefore

$$\begin{aligned} u_k(x) &= \mathbb{E}_{x,k}[\widetilde{M}_n] = \lim_{n \rightarrow \infty} \mathbb{E}_{x,k}[\widetilde{M}_n] \\ &= \mathbb{E}_{x,k}[\lim_{n \rightarrow \infty} \widetilde{M}_n] \\ &= \mathbb{E}_{x,k}[u_{T_k(x)}I(T_k(x) < \infty)] + I(T_k = \infty) \lim_{n \rightarrow \infty} u_n(S_n). \end{aligned}$$

Recall the definition of u_k 's (1), we know that $u_n(x) \leq P(M > n)$, hence $\lim_{n \rightarrow \infty} u_n(S_n) = 0$. We conclude

$$u_k(x) = \mathbb{E}_{x,k}[\bar{F}_M(T_k(x))I(T_k(x) < \infty)].$$

■

The connection between positive harmonic functions and their probabilistic representation (when available) is fundamental to the construction of our simulation schemes. In general, given a time inhomogeneous Markov chain $X = (X_n : n \geq 0)$ and a stopping set $D^c \subset \Omega$, we can define stopping times $T_k(x) = \inf\{n \geq k : X_n \in D^c | X_k = x \in D\}$. Consequently, it is easy to see that for a set of bounded non-negative numbers $\{c_k \in \mathbb{R}^+\}$,

$$h_k(x) = \mathbb{E}_{x,k}(c_{T_k(x)}I(T_k(x) < \infty)) \quad (5)$$

defines a non-negative space-time harmonic sequence. Note that for the case of the u_k 's, we have $c_k = F_M(k)$, $D = \{x : x < b\}$ and $X_n = S_n$. In the rest of this subsection we shall study the connection between harmonic sequences and zero-variance estimator for expectations such as the h_k 's in (5). Obviously, a change-of-measure that in principle provides a zero-variance estimator for $\{h_k(\cdot)\}$ is obtained if we apply importance sampling to $T_k \triangleq T_k(x)$ according to the probability mass function

$$\tilde{f}_{T_k}(t) = \frac{\mathbb{P}_{x,k}(T_k = t) \cdot c_t}{h_k(x)}.$$

Such change-of-measure has two problematic aspects from a simulation standpoint. First, and most obvious, is the fact that $h_k(x)$ is unknown. Second is that, typically, the distribution of T_k is also unknown because one can only simulate T_k via the underlying Markov chain. As a consequence, it would be more useful to understand the change-of-measure in terms of the state process $(X_n : n \geq 0)$. The next lemma shows that there is a particular construction of a probability measure \mathbb{Q}^* that preserves the Markov property of $(X_n : n \geq 0)$.

Lemma 2 *Suppose $(h_k(\cdot) : k \geq 0)$ is a non-negative space-time harmonic sequence defined in the form of (5) for the Markov chain $(X_n : n \geq 0)$. Then, under the change-of-measure \mathbb{Q}^* induced by the Markov transition kernel*

$$Q_n^*(x, dy) = \frac{P_n(x, dy)h_{k+1}(y)}{h_k(x)}, \quad (6)$$

$T_k(x)$ has the distribution given by $\tilde{f}_{T_k}(t)$.

Proof. In order to determine the distribution of the stopping time under \mathbb{Q}^* , we only need to notice that

$$I(T_k(x) = t) = I(X_{k+t} \notin D) \prod_{j=k+1}^{t-1} I(X_j \in D).$$

Therefore

$$\begin{aligned}
\mathbb{Q}_{x,k}^*(T_k(x) = t) &= t = \mathbb{E}_{x,k}^{\mathbb{Q}^*}[I(T_k(x) = t)] \\
&= \mathbb{E}_{x,k}^{\mathbb{Q}^*} \left[I(X_{k+t} \notin D) \prod_{j=k+1}^{k+t-1} I(X_j \in D) \right] \\
&= \mathbb{E}_{x,k}^{\mathbb{P}} \left[I(X_{k+t} \notin D) \prod_{j=k+1}^{k+t-1} I(X_j \in D) \prod_{j=k}^{k+t-1} \frac{h_{k+1}(X_{j+1})}{h_k(X_j)} \right] \\
&= \frac{c_t}{h_k(x)} \mathbb{P}_{x,k}(T_k(x) = t), \quad \text{for } t > k. \\
\mathbb{Q}_{x,k}^*(T_k(x) = k) &= k = \mathbb{E}_{x,k}^{\mathbb{Q}^*}[I(X_k \notin D)] = \mathbb{E}_{x,k}^{\mathbb{P}}[I(X_k \notin D)] \\
&= \frac{c_k}{c_k} \mathbb{P}_{x,k}(T_k(x) = k)
\end{aligned}$$

Because $\sum_{t=k}^{\infty} c_t \mathbb{P}(T_k(x) = t) / h_k(x) = 1$, we know that $\mathbb{Q}^*(T_k(x) = \infty) = 0$.

In fact, for $x \in D$, the estimator $L_k(x)$ under \mathbb{Q}^* is

$$\begin{aligned}
L_k(x) &= c_{T_k} \prod_{j=k}^{T_k-1} \frac{h_j(X_j)}{h_{j+1}(X_{j+1})} \\
&= c_{T_k} \frac{h_k(x)}{h_{T_k}(X_{T_k})} \equiv h_k(x).
\end{aligned} \tag{7}$$

■

The transition kernel (6) proposed in Lemma 2 is in fact the well-known Doob h -transform. The Doob h -transform addresses the second problem we noted when we introduce $\tilde{f}(\cdot)$. We still need to deal with the fact that $h_k(\cdot)$ is unknown. However, as we shall see in Sections 3 and 4, this issue can be alleviated by replacing $h_k(\cdot)$ with some suitable approximations. And, unlike the issue raised by introducing \tilde{f}_{T_k} directly, the use of such approximation guarantees that the Markov property of $(X_n : n \geq 0)$ is preserved and therefore simulation under the change-of-measure can be, in principle, easily applied.

3 State-dependent Sampler and Lyapunov Bounds

Before we move to the detailed discussion of suitable approximations, we shall note that the importance sampling estimators obtained by approximating $h_k(\cdot)$, naturally, will not be zero-variance. Therefore, we have to

provide the means to assess the performance of such estimators. There are extensive discussion on this subject, see, for example, Asmussen and Glynn (2007), Juneja and Shahabuddin (2006) or Bucklew (2004). A popular notion in this context is *strong efficiency*:

Definition 2 Suppose $\alpha(b) = \mathbb{E}[L(b)] > 0$ and $\alpha(b) \searrow 0$ as $b \nearrow \infty$. We say that $L(b)$ is a strongly efficient estimator of $\alpha(b)$ if there exists an $\zeta \in (0, \infty)$ such that

$$\sup_{b>0} \text{cv}(b)^2 \leq \zeta.$$

where $\text{cv}(b)^2 = \text{Var}(L(b))/\alpha(b)^2$.

The motivation of this concept is that, in the presence of strong efficiency, the number of simulation replications needed to obtain a given relative accuracy of $u(b)$ is bounded in b . If $L(b)$ is a strongly efficient estimator of $\alpha(b)$, then in order to obtain an ε -relative accurate estimator of $\alpha(b)$ with $1 - \delta$ confidence, one only needs to sample $n_0 = \lceil \zeta \varepsilon^{-2} \delta^{-1} \rceil$ number of replications of $L(b)$. Note that n_0 depends on neither b nor $\alpha(b)$, this makes $L(b)$ a convenient estimator of $\alpha(b)$.

Let's continue discussing the optimal change-of-measure introduced in Section 2. As we indicated in the previous section, the optimal choice of importance samplers is obtained using Doob's h -transform (6). So it makes sense to look for importance samplers that share this form. For example, instead of using the unknown harmonic sequence $\{h_k(\cdot)\}$ in the definition of the kernel given in (6) to construct the change-of-measure, we can try to replace it with some known approximation. Heuristically, if we can find a series of functions $\{v_k(\cdot)\}$ such that (in some sense)

$$\begin{aligned} v_k(x) &\approx h_k(x) = \mathbb{E}_{x,k}^{\mathbb{P}}[h_{k+1}(X_{k+1})] \\ w_k(x) &\triangleq \mathbb{E}_{x,k}^{\mathbb{P}}[v_{k+1}(X_{k+1})] \approx v_k(x), \end{aligned} \tag{8}$$

then, given that $v_k(\cdot)$ is (in some sense) close to be harmonic, it is reasonable to expect that the induced importance sampling transition kernel by $\{v_k(\cdot)\}$

$$Q_k(x, dy) = P_k(x, dy) \frac{v_{k+1}(y)}{w_k(x)}, \tag{9}$$

has the chance of producing an estimator for $h_k(x)$ with low variance properties.

If (9) is indeed our proposal for the importance sampling change-of-measure, then similar to (7) in Section 2, we have that

$$L_k(x) = c_{T_k(x)} I(T_k(x) < \infty) \prod_{j=k}^{T_k(x)-1} \frac{w_j(X_j)}{v_{j+1}(X_{j+1})} \Bigg|_{X_k=x} \quad (10)$$

is an unbiased estimator for $h_k(x)$ under the change-of-measure \mathbb{Q} induced by (9).

In order to determine the efficiency of a given estimator, we need to provide means to study its second moment. In particular, we are interested in studying

$$s_k(x) = \mathbb{E}_{x,k}^{\mathbb{Q}}[L_k(x)^2]$$

for $x \in D$. It is not hard to see that $s_k(x)$ satisfies similar equations as those defining a space-time harmonic sequence. The bound of $s_k(x)$ can usually be obtained by constructing a sequence of superharmonic envelope functions. This is the so called Lyapunov method. More generally, we have the following result.

Theorem 1 *Let $\{g_k(\cdot) : k \geq 0\}$ be a family of finite-valued non-negative functions for which*

$$\mathbb{E}_{x,k}^{\mathbb{Q}}[g_{k+1}(X_{n+1})] \cdot y_k(x)^2 \leq g_k(x)v_k(x)^2, \quad (11)$$

where y_k 's are non-negative functions and g_k 's are subject to the boundary conditions $g_k(x)v_k(x)^2 \geq c_k^2$, for $x \in D^c$. Then we have

$$g_k(x)v_k(x)^2 \geq \mathbb{E}_{x,k}^{\mathbb{Q}} \left[c_{T_k(x)}^2 I(T_k(x) < \infty) \prod_{j=k}^{T_k(x)-1} \frac{y_j(X_j)^2}{v_{j+1}(X_j)^2} \right].$$

Proof. Fix k and $S_k = x$, define, for $n \geq k$,

$$M_n = g_{T_k(x) \wedge n}(X_{T_k(x) \wedge n}) \prod_{j=k}^{T_k(x) \wedge n - 1} \frac{y_j(X_j)^2}{v_j(X_j)^2}.$$

Then, condition (11) implies that

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[M_{n+1}|\mathcal{F}_n] &= \mathbb{E}^{\mathbb{Q}}[M_{n+1}I(T_k(x) > n)|\mathcal{F}_n] + \mathbb{E}^{\mathbb{Q}}[M_{n+1}I(T_k(x) \leq n)|\mathcal{F}_n] \\
&= I(T_k(x) > n)\mathbb{E}^{\mathbb{Q}}\left[g_{n+1}(X_{n+1})\prod_{j=k}^n\left(\frac{y_j(X_j)}{v_j(X_j)}\right)^2\middle|\mathcal{F}_n\right] + M_nI(T_k(x) \leq n) \\
&\leq I(T_k(x) > n)g_n(X_n)\prod_{j=k}^{n-1}\left(\frac{y_j(X_j)}{v_j(X_j)}\right)^2 + M_nI(T_k(x) \leq n) \\
&= M_n.
\end{aligned}$$

Therefore, $\{M_k\}$ is a non-negative supermartingale under measure \mathbb{Q} . So it converges a.s. as $n \rightarrow \infty$. Moreover, by Fatou's Lemma, we have

$$\begin{aligned}
M_k &= \lim_{n \rightarrow \infty} \mathbb{E}_{x,k}^{\mathbb{Q}}[M_n] \\
&\geq \mathbb{E}_{x,k}^{\mathbb{Q}}[\underline{\lim}_{x \rightarrow \infty} M_n] \\
&= \mathbb{E}_{x,k}^{\mathbb{Q}}\left[g_{T_k(x)}(X_{T_k(x)})I(T_k(x) < \infty)\prod_{j=k}^{T_k(x)-1}\left(\frac{y_j(X_j)}{v_j(X_j)}\right)^2\right].
\end{aligned}$$

Hence,

$$\begin{aligned}
g_k(x)v_k^2(x) &= v_k^2(x)M_k \\
&\geq v_k^2(x)\mathbb{E}_{x,k}^{\mathbb{Q}}\left[g_{T_k(x)}(X_{T_k(x)})I(T_k(x) < \infty)\prod_{j=k}^{T_k(x)-1}\left(\frac{y_j(X_j)}{v_j(X_j)}\right)^2\right] \\
&\geq \mathbb{E}_{x,k}^{\mathbb{Q}}\left[v_{T_k(x)}(X_{T_k(x)})^2g_{T_k(x)}(X_{T_k(x)})I(T_k(x) < \infty)\prod_{j=k}^{T_k(x)-1}\left(\frac{y_j(X_j)}{v_{j+1}(X_j)}\right)^2\right] \\
&\geq \mathbb{E}_{x,k}^{\mathbb{Q}}\left[c_{T_k(x)}^2I(T_k(x) < \infty)\prod_{j=k}^{T_k(x)-1}\left(\frac{y_j(X_j)}{v_{j+1}(X_j)}\right)^2\right].
\end{aligned}$$

■

It does not take long to convince ourselves that the Lyapunov method is a very powerful tool in performance or efficiency analysis. To quickly illustrate the usage of the Lyapunov method, we apply it to the following example.

Example 1 (Light-tailed Geometric sum) If $M \sim \text{Geom}(\rho)$ and that there exists $\theta^* > 0$ such that $\mathbb{E} \exp(\theta^* Z_k) = (1-\rho)^{-1}$ and $\mathbb{E} Z_k \exp(\theta^* Z_k) < \infty$, find a strongly efficient estimator for $u(b)$ as $b \nearrow \infty$.

In this case, thanks to the celebrated Cramer's approximation (See Asmussen (2003), p.365), we have

$$\begin{aligned} u_k(x) &= u(b-x)\overline{F}_M(k) = (1-\rho)^k \mathbb{P}(S_M > b-x) \\ &\sim (1-\rho)^k C \exp(-\theta^*(b-x)) \end{aligned} \quad (12)$$

as $b \nearrow \infty$, therefore we shall use $v_k(x) = (1-\rho)^k \exp(-\theta^*(b-x))$ for $x < b$ to construct the measure \mathbb{Q} . Consequently, we have

$$\begin{aligned} w_k(x) &= (1-\rho)^{k+1} \int_{-\infty}^{\infty} P_k(x, dy) \exp(-\theta^*(b-y)) \\ &= (1-\rho)^{k+1} \int_{-\infty}^{\infty} \mathbb{P}_Z(z) \exp(-\theta^*(b-x-z)) dz \\ &= (1-\rho)^k \exp(-\theta^*(b-x)) = v_k(x); \\ Q_k(x, dy) &= P(x, dy) \frac{v_{k+1}(y)}{v_k(x)} \\ &= P(x, dy)(1-\rho) \exp(\theta^*(y-x)). \end{aligned}$$

\mathbb{Q} is in fact a state-independant change-of-measure and therefore, under \mathbb{Q} , the Z_k 's are i.i.d. Moreover, it follows easily by the convexity of the cumulant generating function of the Z_k 's that $\mathbb{E}_x^{\mathbb{Q}}[Z_{k+1}] > 0$, so $T_k(x) < \infty$. As we discussed before in (10), an unbiased estimator under \mathbb{Q} is

$$\begin{aligned} L_k(x) &= (1-\rho)^{T_k(x)} \frac{v_k(x)}{v_{T_k(x)}(S_{T_k(x)})} \Big|_{S_k=x} \\ &= (1-\rho)^k \exp(\theta^*(x - S_{T_k(x)})). \end{aligned}$$

If we can find $g(x)$ such that

$$\begin{aligned} \mathbb{E}_x^{\mathbb{Q}}[g(x + Z_k)] \cdot w_k(x)^2 &\leq g(x)v_k(x)^2, & \text{for } x < b, \\ g(x)v_k(x)^2 &\geq (1-\rho)^{2k}, & \text{for } x \geq b, \end{aligned}$$

then according to Theorem 5, coupled with the fact (12), we have

$$\text{cv}(L_k(x)) \leq \frac{\sqrt{g_k(x)v_k(x)^2}}{u_k(x)} \sim \frac{\sqrt{g_k(x)v_k(x)}}{v_k(x)} = \sqrt{g_k(x)}$$

as $b \nearrow \infty$. Therefore, the proof of strong efficiency is equivalent to finding a bounded envelope function $g_k(x)$. Indeed, this is easily achieved if we set $g_k(x) \equiv 1$. Hence,

$$L(b) = L_0(0) = \exp(-\theta^* S_{T(b)}),$$

where $T(b) = \inf\{n > 0 : S_n > b\}$, is a strongly efficient estimator for $u(b)$ under \mathbb{Q} as $b \nearrow \infty$.

4 Lyapunov bound for sub-exponential increments

In this section, we develop an estimator for heavy-tailed compound sums and prove its efficiency by means of a suitable Lyapunov function. We assume that the Z_k 's are subexponential, recall the following definition of subexponential r.v.'s:

Definition 3 *We call a non-negative random variable Y subexponential, if*

$$\mathbb{P}(Y_1 + Y_2 > t) \sim 2\mathbb{P}(Y > t)$$

as $t \nearrow \infty$ where Y_1 and Y_2 are i.i.d. copies of Y . A random variable Z is called subexponential if $Z_+ = \max(0, Z)$ is subexponential.

Subexponential distributions describe a very broad class of heavy-tailed distributions. For example, most of the distributions that are used in practice, such as Weibull, lognormal, Pareto, regularly varying distributions and stable laws, are all special cases of subexponential distributions. On the other hand, it has been argued that the subexponential property gives just enough analytical tractability to obtain exact asymptotic results for heavy-tailed compound sums.

For the random variable M , we require the following assumption:

Assumption 1 *Define $\lambda_k \triangleq \overline{F}_M(k+1)/\overline{F}_M(k)$, we assume that*

$$\overline{\lim}_{k \rightarrow \infty} \lambda_k < 1.$$

In addition, let us define

$$\begin{aligned} \mu_k &= \mathbb{E}(M - k | M \geq k) = \frac{\int_k^\infty \overline{F}_M(s) ds}{\overline{F}_M(k)} \\ &= \sum_{j=k}^\infty \frac{\overline{F}_M(j)}{\overline{F}_M(k)} = 1 + \lambda_k + \lambda_k \lambda_{k+1} + \lambda_k \lambda_{k+1} \lambda_{k+2} + \dots \end{aligned}$$

Immediately from Assumption 1, it is evident that there exists a constant $\mu^* \in (0, \infty)$ such that $\sup_{k < \infty} \mu_k \leq \mu^*$.

Going back to our development in Section 3, we are set to build a convenient change-of-measure \mathbb{Q} via some suitably constructed approximating functions ($v_k(\cdot) : k \geq 0$) that capture the asymptotic properties of the u_k 's.

Given each such v_k , the corresponding state-dependent change-of-measure \mathbb{Q} takes the form

$$\mathbb{Q}_{x,k}(Z_{k+1} \in dz) = \frac{v_{k+1}(x+z)}{w_k(x)} \mathbb{P}(Z_1 \in dz)$$

where $w_k(x) = \mathbb{E}^{\mathbb{P}}[v_{k+1}(x+Z)]$.

Fortunately, since the Z_k 's are subexponential, we have the following asymptotic result (Embrechts et al (1997), p.45):

$$\begin{aligned} u_k(x) &= \mathbb{P}(S_{M-k} > b-x; M \geq k) \\ &\sim \mu_k \mathbb{P}(Z_1 > b-x) \cdot \bar{F}_M(k) \\ &\sim \mu_k \mathbb{P}(Z_1 > b-x+a^*) \cdot \bar{F}_M(k) \end{aligned} \quad (13)$$

as $b \nearrow \infty$, where $a^* > 0$ may be any finite constant. The parameter a^* (independent of b) is introduced in (13) in order to give us more control over likelihood ratio on the region where $b-x$ is close to zero when constructing our algorithm. We shall discuss this issue in more detail in Lemma 3. Therefore, based on the asymptotic result (13), combined with the fact that $u_k(x) \leq \bar{F}_M(k)$, a natural candidate for $\{v_k(x)\}$ could be $\min(\mu_k \mathbb{P}(Z_1 > b-x+a^*), 1) \cdot \bar{F}_M(k)$. However, since the description of \mathbb{Q} is invariant under a multiplicative scaling of the v_k 's, we can further simplify our proposal for v_k 's to

$$v_k(x) = \min(\mu_k \mathbb{P}(Z_1 > b-x+a^*), 1). \quad (14)$$

Proposal (14) is very convenient to work with from a simulation standpoint because it can then be represented as the tail of a suitable r.v. V_k , i.e. $\mathbb{P}(V_k > b-x+a^*) = v_k(x)$. The twisted proposing distribution under \mathbb{Q} is then

$$\begin{aligned} \mathbb{Q}_{x,k}(Z_{k+1} \in dz) &= \frac{\mathbb{P}(V_{k+1} > b-x-z+a^*)}{\mathbb{P}(V_{k+1} + Z_1 > b-x+a^*)} \mathbb{P}(Z_1 \in dz) \\ &= \mathbb{P}(Z \in dz | V_{k+1} + Z_1 > b-x+a^*). \end{aligned}$$

Since $v_k(\cdot)$ is a good asymptotic description of $u_k(\cdot)$, it is conceivable that the approximated change-of-measure \mathbb{Q} will behave similarly as the zero-variance importance sampler when b is sufficiently large. Therefore, it is our hope to control the behavior of the likelihood ratio when $b-x+a^*$ is reasonable large. More precisely, we shall impose the following condition on a^* .

Condition 1 Given

$$\delta \in \left(0, \frac{1}{6(1 + \mu^*)}\right] \quad (15)$$

arbitrarily selected, choose $a^* > 0$ so that for all $b, k \geq 0$ and $s \in [0, b)$, we have

$$\frac{w_k(s)}{v_k(s)} \leq \left(\frac{\mu_{k+1}}{\mu_k} + \frac{1}{\mu_k}\right)(1 + \delta). \quad (16)$$

The next result shows that such selection of a^* is always possible.

Lemma 3 *If M satisfies Assumption 1, then for any $\delta > 0$, there exists an a^* such that (16) in Condition 1 holds.*

Proof. First of all, because M satisfies Assumption 1, we have that $\sup \mu_k = \sup \sum_{j=k}^{\infty} \bar{F}_M(j) / \bar{F}_M(k) < \mu^*$ is bounded. Consequently, we know that there exists a y_0 such that if $y = b - s + a^* > y_0$, then

$$\begin{aligned} \sup_{k \geq 0} \mu_k \mathbb{P}(Z_1 > y) &< 1, \\ \sup_{k \geq 0, y > y_0} \left| \mu_k \frac{\mathbb{P}(Z_1 > y)}{\mathbb{P}(V_k > y)} - 1 \right| &\leq \delta. \end{aligned}$$

This is convenient because for such y

$$\begin{aligned} \frac{w_k(s)}{v_k(s)} &= \frac{\mathbb{P}(V_k + Z_1 > y)}{\mathbb{P}(V_k > y)} \\ &= \frac{\mathbb{P}(Z_1 > y)}{\mu_k \mathbb{P}(Z_1 > y)} + \frac{\mathbb{P}(Z_1 \leq y; V_{k+1} > y - Z_1)}{\mu_k \mathbb{P}(Z_1 > y)}. \end{aligned} \quad (17)$$

For the second term,

$$\begin{aligned} &\frac{\mathbb{P}(Z_1 \leq y; V_{k+1} > y - Z_1)}{\mu_k \mathbb{P}(Z_1 > y)} \\ &= \int_{-\infty}^y \frac{\mathbb{P}(V_{k+1} > y - s)}{\mu_k \mathbb{P}(Z_1 > y)} \mathbb{P}(Z_1 \in ds) \\ &\leq \frac{\mu_{k+1}}{\mu_k \mathbb{P}(Z_1 > y)} \int_{-\infty}^y \mathbb{P}(Z_0 > y - s) \mathbb{P}(Z_1 \in ds). \end{aligned}$$

Because $\mathbb{P}(Z_0 > y - s)$ is monotonically increasing in s , we know that

$$\int_{-\infty}^y \mathbb{P}(Z_0 > y - s) \mathbb{P}(Z_1 \in ds) \leq \int_{-\infty}^y \mathbb{P}(Z_0 > y - s) \mathbb{P}(Z_1^+ \in ds).$$

Notice that $\mathbb{P}(Z_0 > y - s) = \mathbb{P}(Z_0^+ > y - s)$ and $\mathbb{P}(Z_1 > y) = \mathbb{P}(Z_1^+ > y)$, we have

$$\frac{\mathbb{P}(Z_1 \leq y; V_{k+1} > y - Z_1)}{\mu_k \mathbb{P}(Z_1 > y)} \leq \frac{\mu_{k+1} \mathbb{P}(Z_0^+ + Z_1^+ > y; Z_1^+ < y)}{\mu_k \mathbb{P}(Z_1^+ > y)}.$$

Remember that Z_0^+, Z_1^+ are non-negative subexponential r.v.'s. So Definition 3 implies

$$\begin{aligned} \mathbb{P}(Z_0^+ + Z_1^+ > y; Z_1^+ < y) &= \mathbb{P}(Z_0^+ + Z_1^+ > y) - \mathbb{P}(Z_1^+ > y) \\ &\sim \mathbb{P}(Z_1^+ > y) \quad \text{as } y \nearrow \infty. \end{aligned}$$

Therefore, we know that for any $\delta > 0$, there exists $y_1 > 0$ such that for $y > y_1$,

$$\frac{\mathbb{P}(Z_1 \leq y; V_{k+1} > y - Z_1)}{\mu_k \mathbb{P}(Z_1 > y)} \leq \frac{\mu_{k+1}}{\mu_k} (1 + \delta).$$

Overall, if we set $a^* = \max(y_0, y_1)$, then for $\forall s \in [0, b)$, $y = b - s + a^*$, we have

$$\frac{w_k(s)}{v_k(s)} \leq \left(\frac{\mu_{j+1}}{\mu_j} + \frac{1}{\mu_j} \right) (1 + \delta).$$

■

Lemma 3 is critical in the construction of our efficient estimator as the proof of our main result shows.

Theorem 2 *Assume that the Z_k 's are subexponential and that M satisfies Assumption 1. Let $T_k(x) = \inf\{m \geq k : S_{m-k} > b - x\}$, then*

i) *Under the proposed change-of-measure \mathbb{Q} ,*

$$\mathbb{E}_{x,k}^{\mathbb{Q}}[e^{\delta^* T_k(x)}] \leq 2e^{\delta^* k} < \infty$$

where $\delta^* = 2 \log[1 + (1 + \mu^*)^{-1}/6]$;

ii) *Moreover, the estimator*

$$L_k(x; b) \triangleq \bar{F}_M(T_k(x)) \cdot \prod_{j=k}^{T_k(x)-1} \frac{w_j(S_j)}{v_{j+1}(S_{j+1})} \Bigg|_{S_k=x} \quad (18)$$

is a strongly efficient estimator for $u_k(x; b)$ as $b \nearrow \infty$. In particular, $L_0(0)$ is a strongly efficient estimator for $u(b) = \mathbb{P}(S_M > b)$ as $b \nearrow \infty$.

Proof. i) If we propose $g_k(x) = 1 + I(x \leq b + a^*)$, then

$$\begin{aligned}
& (1 + \delta)^2 \mathbb{E}_{x,k}^{\mathbb{Q}}[g_{k+1}(x + Z_{k+1})] \\
&= (1 + \delta)^2 [1 + \mathbb{Q}_{x,k}(Z_{k+1} \leq b - x + a^*)] \\
&\leq (1 + \delta)^2 [1 + \mathbb{P}_{x,k}(Z \leq b - x + a^* | Z + V_{k+1} \geq b - x + a^*)] \\
&= (1 + \delta)^2 \left[2 - \frac{\mathbb{P}_{x,k}(Z > b - x + a^*)}{\mathbb{P}_{x,k}(Z + V_k > b - x + a^*)} \right] \\
&= (1 + \delta)^2 \left[2 - \frac{v_k(x)}{\mu_k w_k(x)} \right] \tag{19}
\end{aligned}$$

For any $\delta \in \left(0, \frac{1}{6(1+\mu^*)}\right]$, according to Lemma 3, we know that (19) is less or equal to

$$(1 + \delta)^2 \left[2 - \frac{1}{(1 + \delta)(1 + \mu_{k+1})} \right] \leq 2.$$

In particular, for $\forall x \leq b + a^*$, we have

$$(1 + \delta)^2 \mathbb{E}_{x,k}^{\mathbb{Q}}[g_{k+1}(x + Z_{k+1})] \leq 2 = g_k(x). \tag{20}$$

Equation (20) yields the bound

$$\begin{aligned}
\mathbb{E}_{x,k}^{\mathbb{Q}}[g_{k+1}(x + Z_{k+1})] &\leq g(x, k) - 2\delta \mathbb{E}_{x,k}^{\mathbb{Q}}[g_{k+1}(x + Z_{k+1})] \\
&\leq g(x, k) - 2\delta.
\end{aligned}$$

This bound implies, using a standard Lyapunov argument as in Proposition 11.3.3 of Meyn and Tweedie (1993), that

$$\mathbb{E}_{x,k}^{\mathbb{Q}}[T_k(x)] < \infty. \tag{21}$$

Therefore, under the proposed change-of-measure \mathbb{Q} , the stopping time $T_k(x)$ of our state-dependant importance sampling is almost surely finite. Moreover, from (20) it follows easily that $\widehat{M}_n = g_{n \wedge T_k}(S_{n \wedge T_k})(1 + \delta)^{2(n \wedge T_k)}$ is a supermartingale for $n \geq k$ (with respect to the filtration generated by the S_m 's). Therefore, using Fatou's Lemma, we obtain

$$\begin{aligned}
\mathbb{E}_{x,k}^{\mathbb{Q}}((1 + \delta)^{2T_k(x)}) &\leq \underline{\lim}_{n \rightarrow \infty} \mathbb{E}_{x,k}^{\mathbb{Q}}[\widehat{M}_n] \\
&\leq \widehat{M}_k = 2(1 + \delta)^{2k} < \infty.
\end{aligned}$$

Recall that δ can be any constant in the region $(0, (1 + \mu^*)^{-1}/6]$. Then, if we write $\delta^* = 2 \log[1 + (1 + \mu^*)^{-1}/6]$, we have

$$\mathbb{E}_{x,k}^{\mathbb{Q}}[e^{\delta^* T_k(x)}] \leq 2e^{\delta^* k} < \infty.$$

ii) By design, we know that $L_k(x; b)$ in (18) is the importance sampling estimator obtained by using the change-of-measure \mathbb{Q} generated by the transition kernel

$$Q_k(x, dy) = f_Z(y - x) \frac{v_{k+1}(y)}{w_k(x)} dy,$$

where

$$\begin{aligned} v_{k+1}(y) &= \mathbb{P}(V_{k+1} > b - y + a^*), \\ w_k(x) &= \mathbb{P}(V_{k+1} + Z_1 > b - x + a^*). \end{aligned}$$

It then follows that

$$\begin{aligned} u_k(x) &= \mathbb{E}_{x,k}^{\mathbb{Q}}[L_k(x; b)] \\ &= \bar{F}_M(k) \cdot \mathbb{E}_{x,k}^{\mathbb{Q}} \left[\prod_{j=k}^{T_k(x)-1} \lambda_j \frac{w_j(S_j)}{v_{j+1}(S_{j+1})} \right]. \end{aligned}$$

In order to analyze the efficiency of the estimator $L_k(x)$ we shall find an appropriate bound for

$$s_k(x) = \mathbb{E}_{x,k}^{\mathbb{Q}} \left[\prod_{j=k}^{T_k(x)-1} \frac{\lambda_j^2 w_j(S_j)^2}{v_{j+1}(S_{j+1})^2} \right].$$

According to Theorem 1, if we can find \tilde{g}_k 's such that

$$\mathbb{E}_{x,k}^{\mathbb{Q}} [\tilde{g}_{k+1}(x + Z_{k+1})] \frac{\lambda_k^2 w_k^2(x)}{v_k^2(x)} \leq \tilde{g}_k(x) \quad \text{for } x < b, \quad (22)$$

$$\tilde{g}_k(x) v_k(x)^2 \geq 1 \quad \text{for } x \geq b, \quad (23)$$

then

$$s_k(x) \leq \tilde{g}_k(x) v_k(x)^2.$$

Now, recall the selection of a^* in Lemma 3. We shall have that

$$\frac{w_k(s)}{v_k(s)} \leq \left(\frac{\mu_{k+1}}{\mu_k} + \frac{1}{\mu_k} \right) (1 + \delta). \quad (24)$$

Note that

$$\mu_k = 1 + \lambda_k \mu_{k+1},$$

which implies

$$1 - \frac{1}{\mu_k} \leq \lambda_k \left(\frac{\mu_{k+1}}{\mu_k} + \frac{1}{\mu_k} \right) = 1 - \frac{1 - \lambda_k}{\mu_k} \leq 1.$$

Hence,

$$\lambda_k \frac{w_k(s)}{v_k(s)} \leq (1 + \delta).$$

On the other hand, we know

$$\begin{aligned} v_k(x) &\geq \mu^* \mathbb{P}(Z > a^*) \\ 1/v_k(x)^2 &\leq \mu^{*-2} \mathbb{P}(Z > a^*)^{-2} \end{aligned}$$

Therefore, if we can find a bounded Lyapunov function $\{g_k(x)\}$ such that satisfies

$$(1 + \delta)^2 \mathbb{E}_{x,k}^{\mathbb{Q}} [\tilde{g}_{k+1}(x + Z_{k+1})] \leq \tilde{g}_k(x) \quad \text{for } x < b, \quad (25)$$

$$\tilde{g}_k(x) v_k(x)^2 \geq 1 \quad \text{for } x \geq b, \quad (26)$$

then (22) and (23) will be automatically satisfied. In fact, (25) is exactly the inequality (20) we obtained in part i). Noticing also that $1 \leq g_k(x) = 1 + I(x \leq b + a^*) \leq 2$, we know that $\tilde{g}_k(x) = g_k(x)/\mu^{*2} \mathbb{P}(Z > a^*)^2$ yields the suitable Lyapunov function.

Combining this with the subexponential property (13) of the Z_k 's, we now get

$$\begin{aligned} \frac{\mathbb{E}_{x,k}^{\mathbb{Q}}[L_k(x; b)^2]}{u(x; b)^2} &\sim \frac{\mathbb{E}_{x,k}^{\mathbb{Q}}[L_k(x; b)^2]}{v_k(x)^2 \overline{F}_M(k)^2} \\ &= \frac{s_k(x)}{v_k(x)^2} \\ &\leq g_k(x) \\ &\leq \frac{2}{\mu^{*2} \mathbb{P}(Z > a^*)^2} \end{aligned} \quad (27)$$

as $b \nearrow \infty$. Part ii) of the theorem follows. ■

As a remark, we note that Theorem 2 guarantees a termination time that has finite exponential moments uniformly in b , a feature that is very pleasing from a complexity standpoint.

5 Algorithm and Numerical Example

A precise description of our algorithm for estimating $\mathbb{P}(S_M > b)$ is stated in Algorithm 1 below.

An issue that arises when implementing the proposed algorithm involves the generation of Z_k under \mathbb{Q} in STEP 1. Note that this step involves a

Algorithm 1 Main Algorithm

STEP 0: Set $x \leftarrow 0$, $L \leftarrow 1$, $\tau \leftarrow 1$, $k \leftarrow 0$, select a $a^* > 0$ satisfying Condition 1;

STEP 1: Compute $\mu_{k+1} = \mathbb{E}(M - k | M > k)$ and sample Z_k according to

$$\mathbb{Q}_{x,k}(Z_{k+1} \in dz) = \frac{\mathbb{P}(V_{k+1} > b - x - z + a^*)}{\mathbb{P}(V_{k+1} + Z_1 > b - x + a^*)} \mathbb{P}(Z_1 \in dz), \quad (28)$$

where V_{k+1} is a random variable satisfying

$$\mathbb{P}(V_{k+1} > x) = \min(\mu_{k+1} \mathbb{P}(Z_1 > x), 1).$$

STEP 2: Set $x \leftarrow x + Z_{k+1}$, $k \leftarrow k + 1$,

$$L \leftarrow L \frac{\mathbb{P}(V_{k+1} + Z_k > b - x + a^*)}{\mathbb{P}(V_{k+1} > b - x - z + a^*)}.$$

STEP 3: Repeat STEP 1 and 2 until $x > b$, then move on to STEP 4.

STEP 4: Let $L(b) \leftarrow \bar{F}_M(k)L$ and output L .

rare-event-type simulation problem for large b . This problem, however, is significantly simpler than the original problem as it just involves two random variables; namely, V_{k+1} and Z_1 , whose distribution is computable. One simple approach that can be applied in a case-by-case basis consists in constructing an acceptance / rejection procedure with acceptance probability uniformly bounded away from zero as $b \nearrow \infty$. As it is discussed in detail in Blanchet and Glynn (2007), these types of procedures can be constructed with a low computational cost (in particular, bounded in b for regularly varying functions and with at most $O(b)$ operations for Weibull tails) and do not require the explicit evaluation of the normalizing constant that appears dividing in (28). Nevertheless, such normalizing constant must be computed at the subsequent x 's obtained during the course of the algorithm in STEP 2 in order to evaluate the overall estimator. When numerically computed, such normalizing constants must be evaluated using carefully implemented one dimensional numerical integration routines; note that the number of one dimensional integrals that are required is bounded by part i) of Theorem 2.

As an illustration of our proposed algorithm, we apply our approach to the following examples. The true value reported in the tables was obtained by running 10 million replications of SDIS.

Example 2 (Regularly varying increments) Assume that Z follows

a Pareto distribution with density

$$\mathbb{P}(Z \in dz) = \frac{\alpha}{(1+z)^{\alpha+1}} dz, \quad z \geq 0, \alpha > 1.$$

Thus $\mathbb{P}(Z \geq z) = (1+z)^{-\alpha}$. Compute the probability $\mathbb{P}(S_M > b)$ where $M \sim \text{Geom}(\rho)$.

In this particular example, $(\mu_{j+1}/\mu_j + 1/\mu_j) = 3/2$. We can set $a^* = 4$ for the case $\alpha = 1.5$ and $a^* = 0$ for case $\alpha = 50$ to satisfy condition (16). The numerical results are reported in Tables 1 and 2. We also tried our method with a^* taking values as low as 0 and the results do not seem to change significantly. One reason for this may be that inequality (16) controls the behavior of the likelihood ratio throughout the whole region $(-\infty, b]$. However, the process will spend most of the time in a region that is far away from b . Therefore, a worst-case scenario type bound such as (16) seems conservative. In the following tables, the result of our method (denote as SDIS – state-dependant importance sampling) is compared to two recently proposed methods: the importance sampling algorithm proposed by Dupuis, Leder and Wang (2007), denoted as DLW; and conditional Monte Carlo of Asmussen and Kroese (2006), denoted as CMC.

We now discuss a family of examples for which our method is the only approach that can be rigorously proved to provide a strongly efficient estimator. In particular, we apply our method to Weibullian tails and, as an empirical benchmark, we used once again both the conditional Monte Carlo method, CMC, and DLW’s estimator.

Example 3 (Weibull increments) Assume that Z follows a Weibull distribution with density

$$\mathbb{P}(Z \in dz) = \beta z^{\beta-1} e^{-z^\beta} dz, \quad z \geq 0, 0 < \alpha < 1.$$

Thus $\mathbb{P}(Z \geq z) = e^{-z^\beta}$. Compute the probability $\mathbb{P}(S_M > b)$ where $M \sim \text{Geom}(\rho)$. We selected $\beta = 3/4$, which is a value larger than the threshold at which CMC is shown to be strongly efficient for deterministic number terms; CMC has not been shown to be efficient for random number of increments, as we consider here. The results are shown in Table 3, where we set $a^* = 80$ when applying SDIS. The empirical analysis suggests a very good performance for CMC beyond the theoretical analysis reported in Asmussen and Kroese (2006). DLW’s procedure, on the other hand, consistently reports a confidence interval that fails to cover the tail probability of interest; despite suggesting a small standard deviation estimated from the output data. This type of situation, which is known to be difficult to

detect in practice, occurs in the design of importance sampling algorithm that have a highly skewed likelihood ratio. In fact, one can easily construct examples to illustrate that the coefficient of variation of DLW's estimator grows significantly for Weibullian tails as the tail parameter $b \nearrow \infty$.

References

- [1] Asmussen, S. (2003) *Applied Probability and Queues*. Springer-Verlag. New York.
- [2] Asmussen, S., Binswanger, K., and Hojgaard, B. (1997) Simulation of ruin probabilities for subexponential claims. *Astin Bulletin* **27**, 297-318.
- [3] Asmussen, S., Binswanger, K., and Hojgaard, B. (2000) Rare event simulation for heavy-tailed distributions. *Bernoulli* **6**, 303-322.
- [4] Asmussen, S. and Kroese, D. (2006) Improved algorithms for rare event simulation with heavy tails. *Adv. in Appl. Probab.* **38**, 545-558.
- [5] Asmussen, S. and Glynn, P. (2007) *Stochastic Simulation*. Springer. New York.
- [6] Blanchet, J. and Glynn, P. (2007) Efficient rare event simulation for the maximum of heavy-tailed random walks. *Ann. of Appl. Prob.* To appear.
- [7] Boots, N.K. and Shahabuddin, P. (2001) Simulating ruin probabilities in insurance risk processes with subexponential claims. *Proceedings of the 2001 Winter Simulation Conference*, pp. 468–476.
- [8] Bucklew, J. A. (2004) *Introduction to Rare Event Simulation*. Springer-Verlag. New York.
- [9] Doob, J. L. (1984) *Classical Potential Theory and Its Probabilistic Counterpart*. Springer-Verlag. New York.
- [10] Dupuis, P., Leder, K., and Wang, H. (2007) Importance Sampling for Sums of Random Variables with Regularly Varing Tails. *ACM TOMACS* **17**, No 3, Article 14.
- [11] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997) *Modelling extreme events with applications to insurance and finance*. Springer-Verlag. New York.

- [12] Juneja, S. and Shahabuddin, P. (2002) Simulating heavy-tailed processes using delayed hazard rate twisting. *ACM TOMACS* **12**, 94-118.
- [13] Juneja, S. and Shahabuddin, P. (2006) Rare Event Simulation Techniques: An Introduction and Recent Advances. *Handbook on Simulation*. Chapter 11. Elsevier. Editors: Shane Henderson and Barry Nelson 291-350.
- [14] Kalashnikov, V. (1997) *Geometric Sums: Bounds for Rare Events with Applications*. Kluwer. Dordrecht, The Netherlands.
- [15] Meyn, S.P. and Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability*. Springer-Verlag. London.
- [16] Siegmund, D. (1976) Importance Sampling in the Monte Carlo Study of Sequential Tests, *Ann. Stat.* **4**, 673-864.
- [17] Willmot, G.E., and Lin, S. (2000) *Lundberg approximations for compound distributions with insurance applications*. Lecture Notes in Statistics **156**, Springer-Verlag. New York.

[Estimation] [Std. Error] [Conf. Interval]	$b = 10^3$	$b = 10^5$	$b = 10^8$
True Value	6.352e-05	6.325e-08	2e-12
SDIS	6.341e-05 1.99e-07 [6.302e-05, 6.380e-05]	6.327e-08 2.01e-10 [6.323e-08, 6.331e-08]	2.004e-12 6.286e-15 [1.992e-12, 2.017e-12]
DLW	6.340e-05 2.042e-07 [6.336e-05, 6.344e-05]	6.325e-08 2.19e-11 [6.321e-08, 6.329e-08]	2.000e-12 7.694e-16 [1.999e-12, 2.002e-12]
CMC	6.335e-05 3.234e-07 [6.270e-05, 6.400e-05]	6.276e-08 3.116e-10 [6.214e-08, 6.338e-08]	2.007e-12 9.886e-15 [1.809e-12, 2.205e-12]

Table 1: Simulation Result of Example 2 with $\alpha = 3/2$, $\rho = 1/2$, and the number of replications is 20,000. Three values reported in each cell are estimation, standard error and confidence interval respectively.

[Estimation] [Std. Error] [Conf. Interval]	$b = 10^6$	$b = 10^{12}$	$b = 10^{18}$
True Value	0.002	2e-06	2e-09
SDIS	0.002 6.304e-06 [1.999e-03, 2.001e-03]	1.997e-06 6.336e-09 [1.984e-06, 2.009e-06]	2.008e-09 6.329e-12 [2.007e-09, 2.009e-09]
DLW	0.002001 8.019e-07 [1.999e-03, 2.002e-03]	2.000e-06 7.224e-10 [1.998e-06, 2.001e-06]	2.000e-09 7.225e-13 [1.999e-09, 2.001e-09]
CMC	0.00201 1.002e-05 [1.998e-03, 2.031e-03]	2.013e-06 1.024e-08 [1.993e-06, 2.033e-06]	2.012e-09 1.013e-11 [1.992e-09, 2.032e-09]

Table 2: Simulation result of Example 2 with $\alpha = 1/2$, $\rho = 1/2$ and 20,000 replications. Three values reported in each cell are estimation, standard error and confidence interval respectively.

[Estimation] [Std. Error] [Conf. Interval]	$b = 30$	$b = 40$	$b = 50$
True Value	6.0e-05	3.1e-06	1.8e-07
SDIS	5.73e-05 6.29e-06 [4.50e-05, 6.96e-05]	3.27e-06 5.49e-07 [2.15e-06, 4.31e-06]	1.57e-07 2.91e-08 [1e-07, 2.14e-07]
DLW	4.30e-05 1.40e-05 [1.55e-05, 7.04e-05]	1.33e-06 2.16e-07 [0.91e-06, 1.76e-06]	0.89e-07 2.55e-08 [0.39e-07, 1.39e-07]
CMC	6.59e-05 2.01e-05 [2.64e-05, 10.5e-05]	1.81e-06 2.22e-07 [1.38e-06, 2.25e-06]	1.09e-07 3.87e-08 [0.33e-07, 1.85e-07]

Table 3: Simulation result of Example 3 with $\beta = 3/4$, $\rho = 1/2$ and 10,000 replications. Three values reported in each cell are estimation, standard error and confidence interval respectively.