Importance Sampling for Sums of Random Variables with Regularly Varying Tails

Paul Dupuis,* Kevin Leder and Hui Wang†
Lefschetz Center for Dynamical Systems
Brown University
Providence, R.I. 02912
USA

Abstract

Importance sampling is a variance reduction technique for efficient estimation of rare-event probabilities by Monte Carlo. For random variables with heavy tails there is little consensus on how to choose the change of measure used in importance sampling. In this paper we study dynamic importance sampling schemes for sums of independent and identically distributed random variables with regularly varying tails. The number of summands can be random but must be independent of the summands. For estimating the probability that the sum exceeds a given threshold, we explicitly identify a class of dynamic importance sampling algorithms with bounded relative errors. In fact, these schemes are nearly asymptotically optimal in the sense that the second moment of the corresponding importance sampling estimator can be made as close as desired to the minimal possible value.

1 Introduction

Suppose one wishes to estimate the quantity $p_b = P(S_n > b)$, where $S_n = X_1 + \cdots + X_n$ and the X_i's are real-valued, independent and identically distributed (iid). A simple and often effective means is to use Monte Carlo simulation. One generates K iid replicas $\{S_n^k\}$ of the random variable S_n,

*Research of this author supported in part by the National Science Foundation (NSF-DMS-0306070 and NSF-DMS-0404806) and the Army Research Office (DAAD19-02-1-0425 and W911NF-05-1-0289).

†Research of this author supported in part by the National Science Foundation (NSF-DMS-0103669 and NSF-DMS-0404806).

and forms the estimate $Z_K \doteq (\sum_{k=1}^K I_{\{S_n^k > b\}})/K$. The rate of convergence of Z_K is determined by its variance:

$$\mathrm{var}(Z_K) = (p_b - p_b^2)/K.$$

Note that $p_b \to 0$ as $b \to \infty$ implies $\mathrm{var}(Z_K) \to 0$ as $b \to \infty$. However, when estimating small probabilities a more important statistic is the relative error of the estimate:

$$RE(Z_K) \doteq \frac{\text{standard deviation of } Z_K}{\text{mean of } Z_k} = \frac{1}{\sqrt{K}} \cdot \sqrt{\frac{1-p_b}{p_b}}.$$

Hence for bounded relative error it is necessary that K grows as fast as $1/p_b$, and because of this standard Monte Carlo simulation is rarely used to estimate rare event probabilities.

An alternative approach to the problem of estimating small probabilities is importance sampling, where instead of sampling from the original distribution samples are drawn from a new distribution under which the rare events are no longer rare. More specifically, iid samples of the random variable $I_{\{\tilde{S}_n > b\}}$ are drawn, where $\tilde{S}_n = \tilde{X}_1 + \cdots + \tilde{X}_n$ and the vector $(\tilde{X}_1, \ldots, \tilde{X}_n)$ has an alternative distribution, say ν_n^b. The corresponding importance sampling estimator is just the sample average of iid copies of

$$\hat{p}_b \doteq I_{\{\tilde{S}_n > b\}} \frac{d\mu}{d\nu_n^b}(\tilde{X}_1, \ldots, \tilde{X}_n),$$

where μ denotes the distribution of $(X_1, \ldots, X_n)$. Clearly this estimator is unbiased. The goal of importance sampling is to choose ν_n^b so as to minimize the variance, or equivalently, the second moment of $\hat{p}_b$:

$$E\left[\hat{p}_b^2\right] = E\left[I_{\{S_n > b\}} \frac{d\mu}{d\nu_n^b}(X_1, \ldots, X_n)\right].$$

It turns out that solving for the unconstrained minimization problem over all possible distributions requires knowing p_b. Instead, one typically searches within a parametric family of changes of measure and looks for a distribution that satisfies an optimality criterion. Jensen's inequality implies

$$E\left[\hat{p}_b^2\right] \geq (E\left[\hat{p}_b\right])^2 = p_b^2,$$

thus giving a lower bound on the second moment. A change of measure ν_n^b is said to be *asymptotically optimal*, or have *asymptotically optimal relative error*, if

$$\lim_{b\to\infty} \frac{E\left[\hat{p}_b^2\right]}{p_b^2} = \frac{E\left[I_{\{S_n > b\}} d\mu/d\nu_n^b(X_1, \ldots, X_n)\right]}{p_b^2} = 1. \tag{1.1}$$

One would like to construct schemes whose asymptotic relative error is close to or equal to this minimal value 1.

In [6, 7] it was shown that ideas from stochastic control and game theory can be used effectively in the design of importance sampling schemes for random variables with finite moment generating functions. This paper is concerned with sums of non-negative random variables with heavy tailed distributions (by which we mean $E[\exp(tX_i)] = \infty$ for all $t > 0$). For this setup, there was no general theory for choosing sampling distributions ν_n^b that satisfy this asymptotic optimality criterion, or even distributions that have uniformly (in b) bounded relative error. A goal of the current paper is to demonstrate that the techniques of control theory can again serve as basic tools in the design and analysis of asymptotically optimal importance sampling schemes for heavy tailed distributions.

The paper is organized as follows. Section 2 introduces a parametric family of alternative sampling distributions (i.e., controls) ν_n^b. In Section 3, we use weak convergence arguments to show that, when the number of summands is fixed, such changes of measure induce estimators with bounded relative errors. Moreover, one can always identify nearly asymptotically optimal schemes in the sense that the corresponding importance sampling estimators come within an (arbitrarily) prescribed error of the absolute lower bound 1 in (1.1). In Section 4 we adapt this construction to estimate

$$\rho_b = P(X_1 + \cdots + X_N > b)$$

when N is a random variable that is independent of $\{X_i, i \in \mathbb{N}\}$ and satisfies $E[z^N] < \infty$ for some $z > 1$. For this case we are also able to identify importance sampling schemes that are nearly asymptotically optimal. Section 5 presents a collection of numerical results. We compare our scheme with two existing simulation methods, one of which is based on conditional Monte Carlo rather than importance sampling [1], and the other is based on delayed hazard rate twisting [9]. It is worth mentioning that the conditional Monte Carlo algorithm produces estimates that have bounded relative errors, although it is not known whether they satisfy the asymptotic optimality criterion.

2 Problem setup

Consider a sequence of iid non-negative random variables $\{X_i, i \in \mathbb{N}\}$ with tail probability $\bar{F}(x) \doteq P(X_i > x)$. Let $S_n \doteq X_1 + \cdots + X_n$. Assume that, for some $\alpha > 0$, the function $\bar{F}$ satisfies

$$\lim_{b\to\infty} \frac{\bar{F}(ab)}{\bar{F}(b)} = a^{-\alpha} \text{ for all } a > 0. \tag{2.1}$$

A random variable with this property is said to have *regularly varying tails*. It is well known that such random variables are *subexponential* [1, page 253, Proposition 1.4] in the sense that

$$\lim_{b\to\infty} \frac{P(S_n > b)}{P(X_1 > b)} = n \tag{2.2}$$

for every $n \in \mathbb{N}$. An in-depth account of heavy-tailed distributions can be found in [8].

We wish to estimate $P(S_N > b)$ when b is a large positive number and N is an $\mathbb{N}$-valued random variable independent of $\{X_i\}$. In preparation, we first study the special case where $N \equiv n$ is a fixed number. As discussed in the Introduction, the samples are drawn from an alternative distribution ν_n^b. Our goal is to find, for each $\varepsilon > 0$, measures ν_n^b (we omit the ε-dependence in the notation) such that

$$\lim_{b\to\infty} \frac{E[I_{\{S_n>b\}} d\mu/d\nu_n^b(X_1, \ldots, X_n)]}{P(S_n > b)^2} \leq 1 + \varepsilon. \tag{2.3}$$

When ε is small, the importance sampling scheme based on ν_n^b achieves a nearly asymptotic optimal relative error [compare with (1.1)]. Using the subexponential property (2.2), (2.3) reduces to

$$\lim_{b\to\infty} \frac{E[I_{\{S_n>b\}} d\mu/d\nu_n^b(X_1, \ldots, X_n)]}{\bar{F}(b)^2} \leq (1 + \varepsilon)n^2. \tag{2.4}$$

The algorithm for this special case can then be adapted to case where N is random. This extension will be discussed in Section 4.

Remark 2.1 We will assume throughout that the random variable X_i has a density f. This condition is not essential and is imposed simply for convenience of exposition.

2.1 A parameterized family of sampling distributions

In the setting of light-tailed random variables (i.e., those with finite moment generating functions in a neighborhood of the origin), it is customary to consider sampling distributions that belong to the class of "exponential tilts" and/or their mixtures, and indeed one can obtain very good results by doing so.

However, the situation is less clear for random variables with regularly varying tails. A contribution of the present paper is the identification of a class of sampling distributions that can yield asymptotically optimal performance and are simple to implement. The main requirement is that one should be able to sample from the tail distribution with density $f(y)I_{\{y>c\}}/\bar{F}(c)$ for all $c \geq 0$.

Fix $n \in \mathbb{N}$. Each distribution in our class will be determined by parameters $(a, p_{i,n}, q_{i,n})$, where $a \in (0,1)$ and $\{p_{i,n}, q_{i,n}, 1 \leq i \leq n-1\}$ is a sequence of non-negative numbers such that $p_{i,n} + q_{i,n} = 1$ and $q_{i,n} > 0$ for every i. It is easiest to describe the distribution of interest as that induced by random variables $(Y_{1,n}, Y_{2,n}, \ldots, Y_{n,n})$. Here $Y_{1,n}$ has the density

$$f_{1,n}^b(y) = p_{1,n}f(y) + q_{1,n}\frac{f(y)}{\bar{F}(ab)}I_{\{y>ab\}}.$$

For $1 < i < n$ the conditional density of $Y_{i,n}$, given $S_{i-1,n} \doteq Y_{1,n} + \cdots + Y_{i-1,n} = s_{i-1} \leq b$, is

$$f_{i,n}^b(y) = p_{i,n}f(y) + q_{i,n}\frac{f(y)}{\bar{F}(a(b-s_{i-1}))}I_{\{y>a(b-s_{i-1})\}},$$

and $f_{i,n}^b = f$ if $s_{i-1} > b$. Lastly, if $s_{n-1} \leq b$ then $Y_{n,n}$ has conditional density

$$f_{n,n}^b(y) = \frac{f(y)}{\bar{F}(b-s_{n-1})}I_{\{y>b-s_{n-1}\}},$$

and otherwise the conditional density of $Y_{n,n}$ is f.

Note that it is not difficult to simulate from this distribution. When drawing the sample $Y_{i,n}$, if $S_{i-1,n} = s_{i-1} \leq b$ then one first flips a coin that is heads with probability $p_{i,n}$. If heads comes up then we sample from the original distribution. Otherwise we sample from the original distribution conditioned on the event that the outcome is greater than $a(b - s_{i-1})$. If $s_{i-1} > b$ then of course we sample from the original distribution.

Remark 2.2 If $s_{i-1} < b$, then $(b - s_{i-1})$ is the residual distance to go before the sample sum exceeds the threshold b. The role of the parameter $a \in (0,1)$ is to determine how close we will come to jumping all the required distance when the coin turns up tails (except for $i = n$). Since $a < 1$ we do not attempt to jump over the threshold w.p.1, but rather with positive probability we come close to but not over the threshold. It will turn out that the asymptotic performance (as $b \uparrow \infty$) depends on a, and that as $a \uparrow 1$ this asymptotic performance approaches optimality. Hence it is tempting to use $a = 1$ in the prelimit also. However, it turns out that the limits $a \uparrow 1$ and $b \uparrow \infty$ do not permute. As a consequence, the corresponding importance sampling scheme does not even achieve good asymptotic performance if one sets $a = 1$ in the prelimit.

3 Near asymptotic optimality for fixed n

In this section we analyze, via weak convergence methods, the asymptotic performance of the parametric family of changes of measure defined in Section 2. For each fixed choice of the parameters $(\alpha, p_{i,n}, q_{i,n})$ [i.e., controls], we obtain a cost. Thus finding a good change of measure amounts to solving a deterministic, discrete time control problem. Nearly optimal controls are identified, which in turn yield asymptotically nearly optimal changes of measure for the importance sampling problem.

3.1 A weak convergence analysis

Proposition 3.1 *Fix $n \in \mathbb{N}$. Let ν_n^b be the importance sampling distributions defined in Section 2 with parameters $(\alpha, p_{i,n}, q_{i,n})$. Then*

$$\lim_{b\to\infty} \frac{E[I_{\{S_n>b\}} d\mu/d\nu_n^b(X_1,\ldots,X_n)]}{\bar{F}(b)^2} = \prod_{j=1}^{n-1} \frac{1}{p_{j,n}} + a^{-\alpha} \sum_{i=1}^{n-1} \frac{1}{q_{i,n}} \prod_{j=1}^{i-1} \frac{1}{p_{j,n}}.$$

The limit will be shown using weak convergence methods. After setting up the notation, we present a few preliminary results before returning to the proof of the proposition. To begin, we rewrite the expected value as

$$\frac{1}{\bar{F}(b)^2} \int_{\mathbb{R}_+^n} I_{\{y_1+\cdots+y_n>1\}} \frac{d\mu}{d\nu_n^b}(b\mathbf{y})\mu(bd\mathbf{y}),$$

where $\mathbf{y} = (y_1, \ldots, y_n)$. Define

$$\mathsf{K} \doteq \{\mathbf{y} \in \mathbb{R}_+^n : y_1 + \cdots + y_n > 1\}$$

and a family of measures on $\mathbb{R}_+^n$ by

$$\theta^b(A) \doteq \frac{\mu(b(A \cap \mathsf{K}))}{\bar{F}(b)}$$

(recall that μ is the product probability measure induced by the iid random variables $X_1, \ldots, X_n$). Then the integral can be rewritten in the form

$$\int_{\mathbb{R}_+^n} \frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}) \theta^b(d\mathbf{y}).$$

For the rest of the proof we use the definitions $M_n \doteq \max(X_1, \ldots, X_n)$ and $S_n \doteq X_1 + \cdots + X_n$.

Lemma 3.2 *If* $(X_1, \ldots, X_n)$ *are iid non-negative random variables with the subexponential property, then*

$$\lim_{b \to \infty} P(M_n \leq b \mid S_n > b) = 0.$$

Proof. Observe the following result due to the inclusion-exclusion principle:

$$P(M_n > b) = nP(X_1 > b) + C_2(P(X_1 > b))^2 + \cdots + C_{n-1}(P(X_1 > b))^{n-1},$$

where $C_2, \ldots, C_n$ are some constants. It follows that

$$\lim_{b \to \infty} \frac{P(M_n > b)}{P(X_1 > b)} = n.$$

Thanks to the subexponential property (2.2),

$$P(M_n \leq b \mid S_n > b) = 1 - P(M_n > b \mid S_n > b) = 1 - \frac{P(M_n > b)}{P(S_n > b)} \to 0.$$

This completes the proof. ■

We can now analyze the weak convergence of θ^b as $b \to \infty$. Although the θ^b's are not necessarily probability measures, there is an obvious extension of the notion of weak convergence to non-negative measures with uniformly bounded mass [5, page 373]. In the following θ_j is defined as the probability measure on $\mathbb{R}_+^n$ generated by the random vector $(Y_1^j, \ldots, Y_n^j)$, where $Y_i^j = 0$ for $i \neq j$ and Y_j^j has density $\alpha y^{-\alpha-1} 1_{\{y \geq 1\}}$.

Lemma 3.3 $\theta^b \Rightarrow \theta \doteq \sum_{j=1}^n \theta_j$.

Proof. For any vector $a \in \mathbb{R}_+^n \setminus \{0\}$ define the rectangle $R_a \doteq \{y \in \mathbb{R}_+^n : y_1 \leq a_1, \ldots, y_n \leq a_n\}$. Since these rectangles are convergence determining

[3, Example 2.3] and $\lim_{b\to\infty}\theta^b(\mathbb{R}^n_+) = n = \theta(\mathbb{R}^n_+)$ thanks to (2.2), it suffices to show that

$$\lim_{b\to\infty}\theta^b(R_a) = \theta(R_a) \tag{3.1}$$

for all those $a \in \mathbb{R}^n_+$ such that $\theta(\partial R_a) = 0$.

To this end we first consider the case $\max\{a_1,\ldots,a_n\} \le 1$. It is obvious that $\theta(R_a) = 0$, so we only have to prove $\theta^b(R_a) \to 0$. This follows immediately from Lemma 3.2 and the subexponential property (2.2), since

$$\theta^b(R_a) \le \frac{P(M_n \le b, S_n > b)}{\bar{F}(b)} = P(M_n \le b \mid S_n > b)\frac{P(S_n > b)}{\bar{F}(b)} \to 0.$$

Next consider the case $\max\{a_1,\ldots,a_n\} > 1$, and without loss of generality assume that $a_j > 1$ for $1 \le j \le k$ only. We can also assume that $a_i > 0$ for every i since $\theta(\partial R_a) > 0$ otherwise. Define

$$U_0 \doteq \{y_1 \le 1,\ldots,y_k \le 1, y_{k+1} \le a_{k+1},\ldots,y_n \le a_n\},$$

and for $1 \le j \le k$

$$U_j \doteq \{y_1 \le 1,\ldots,y_{j-1} \le 1, 1 < y_j \le a_j, y_{j+1} \le a_{j+1},\ldots,y_n \le a_n\}.$$

Clearly the U_j's are disjoint and $R_a = U_0 \cup U_1 \cup \cdots \cup U_k$. All we need to show is that $\theta^b(U_j) \to \theta(U_j)$ for every $0 \le j \le k$.

The convergence of $\theta^b(U_0) \to \theta(U_0) = 0$ is already established since $U_0 = R_{\bar{a}}$ where $\bar{a} = (1,\ldots,1,a_{k+1},\ldots,a_n)$ and $\max\{\bar{a}_1,\ldots,\bar{a}_n\} \le 1$. It remains to show for the case where $j \ge 1$. Using the definition of θ and the fact that θ_j is supported on points where $y_i = 0$ if $i \ne j$, we see that

$$\theta(U_j) = \theta_j(U_j) = \alpha\int_1^{a_j} y^{-\alpha-1}dy = 1 - a_j^{-\alpha}.$$

Since $U_j \subset \mathsf{K}$, it follows from the definition of θ^b that

$$\begin{aligned}
\theta^b(U_j) &= \frac{1}{\bar{F}(b)}P\left\{(X_1,\ldots,X_n) \in bU_j\right\} \\
&= \frac{1}{\bar{F}(b)}P(b < X_j \le a_j b)\cdot\prod_{i<j}P(X_i \le b)\cdot\prod_{i>j}P(X_i \le a_i b) \\
&= \frac{\bar{F}(b) - \bar{F}(a_j b)}{\bar{F}(b)}\cdot\prod_{i<j}P(X_i \le b)\cdot\prod_{i>j}P(X_i \le a_i b).
\end{aligned}$$

Since $a_i > 0$ for every i, the regularly varying tail property implies

$$\lim_{b\to\infty}\theta^b(U_j) = 1 - a_j^{-\alpha} = \theta(U_j).$$

This completes the proof. ∎

Lemma 3.4 *There exists $M < \infty$ such that for any $b \in [0, \infty)$ and any $y \in \mathsf{K}$,*

$$\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(by) \leq M.$$

Proof. For $y \in \mathsf{K}$ set $s_0 = 0, s_j = y_1 + \cdots + y_j$, and define $\tau(y) \doteq \min\{j \geq 1 : s_j > 1\}$. We consider the cases $\tau(y) = n$ and $\tau(y) < n$ separately.

Case 1: Assume for now that $y \in \mathsf{K}$ and $\tau(y) = n$. Then by definition of ν_n^b,

$$\begin{aligned}\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}) \;=\; & \frac{\bar{F}(b(1-s_{n-1}))}{\bar{F}(b)} \prod_{j=1}^{n-1} \Bigg(\frac{1}{p_{j,n}} I_{\{y_j \leq a(1-s_{j-1})\}} \\ & + \frac{\bar{F}(ab(1-s_{j-1}))}{p_{j,n}\bar{F}(ab(1-s_{j-1})) + q_{j,n}} I_{\{y_j > a(1-s_{j-1})\}} \Bigg).\end{aligned}$$

Consider the decomposition $\mathsf{K} = \mathsf{K}_1 \cup \mathsf{K}_2$ where $\mathsf{K}_1 = \{y \in \mathbb{R}_+^n : y_j \leq a(1-s_{j-1})$ for $1 \leq j \leq n-1$ and $s_n > 1\}$ and $\mathsf{K}_2 = \mathsf{K} \setminus \mathsf{K}_1$.

For $y \in \mathsf{K}_1$, it is not difficult to argue by induction that $s_j \leq 1 - (1-a)^j$ for $1 \leq j \leq n-1$. Therefore,

$$\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}) = \frac{\bar{F}(b(1-s_{n-1}))}{\bar{F}(b)} \prod_{j=1}^{n-1} \frac{1}{p_{j,n}} \leq \frac{\bar{F}(b(1-a)^{n-1})}{\bar{F}(b)} \prod_{j=1}^{n-1} \frac{1}{p_{j,n}}. \quad (3.2)$$

For any $y \in \mathsf{K}_2$, let $J \doteq \{j : y_j > a(1-s_{j-1}), j = 1, \ldots, n-1\}$, which is non-empty. Define j^* to be the smallest element in J, and let

$$q \doteq \min\{q_{j,n} : 1 \leq j \leq n-1\}. \quad (3.3)$$

Note that for all j

$$\frac{\bar{F}(ab(1-s_{j-1}))}{p_{j,n}\bar{F}(ab(1-s_{j-1})) + q_{j,n}} < \frac{1}{p_{j,n}}.$$

Then the following bound is obtained:

$$\begin{aligned}\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}) \;&=\; \frac{\bar{F}(b(1-s_{n-1}))}{\bar{F}(b)} \cdot \prod_{j \in J} \frac{\bar{F}(ab(1-s_{j-1}))}{p_{j,n}\bar{F}(ab(1-s_{j-1})) + q_{j,n}} \cdot \prod_{j \in J^c} \frac{1}{p_{j,n}} \\ &\leq\; \frac{1}{\bar{F}(b)} \frac{\bar{F}(ab(1-s_{j^*-1}))}{p_{j^*,n}\bar{F}(ab(1-s_{j^*-1})) + q_{j^*,n}} \cdot \prod_{j \neq j^*} \frac{1}{p_{j,n}} \\ &\leq\; \frac{1}{\bar{F}(b)} \frac{\bar{F}(ab(1-s_{j^*-1}))}{q} \cdot \prod_{j=1}^{n-1} \frac{1}{p_{j,n}}.\end{aligned}$$

Since for every $k \in \{1, \dots, j^*-1\}$ we have $y_k \le a(1-s_{k-1})$, induction yields that $s_k \le 1-(1-a)^k$ for all such k's. In particular, $s_{j^*-1} \le 1-(1-a)^{j^*-1} \le 1-(1-a)^{n-2}$. Thus for every $y \in \mathsf{K}_2$

$$\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}) \le \frac{\bar{F}(ab(1-a)^{n-2})}{q\bar{F}(b)} \prod_{j=1}^{n-1} \frac{1}{p_{j,n}}. \tag{3.4}$$

Thanks to (3.2), (3.4), observing $q \le 1$ and $\min\{(1-a)^{n-1}, a(1-a)^{n-2}\} \ge a(1-a)^{n-1}$, we obtain the bound

$$\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}) \le \frac{1}{q} \prod_{j=1}^{n-1} \frac{1}{p_{j,n}} \cdot \frac{\bar{F}(ba(1-a)^{n-1})}{\bar{F}(b)} \tag{3.5}$$

for every $y \in \mathsf{K}$ and $\tau(y) = n$.

Case 2: Assume that $y \in \mathsf{K}$ and $\tau \doteq \tau(y) < n$. In this case we have

$$\begin{aligned}
\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}) &= \frac{1}{\bar{F}(b)} \frac{\bar{F}(ab(1-s_{\tau-1}))}{p_{\tau,n}\bar{F}(ab(1-s_{\tau-1})) + q_{\tau,n}} \prod_{j=1}^{\tau-1} \Bigg(\frac{1}{p_{j,n}} I_{\{y_j \le a(1-s_{j-1})\}} \\
&\qquad + \frac{\bar{F}(ab(1-s_{j-1}))}{p_{j,n}\bar{F}(ab(1-s_{j-1})) + q_{j,n}} I_{\{y_j > a(1-s_{j-1})\}} \Bigg) \\
&\le \frac{1}{q_{\tau,n}} \frac{\bar{F}(ab(1-s_{\tau-1}))}{\bar{F}(b)} \prod_{j=1}^{\tau-1} \Bigg(\frac{1}{p_{j,n}} I_{\{y_j \le a(1-s_{j-1})\}} \\
&\qquad + \frac{\bar{F}(ab(1-s_{j-1}))}{p_{j,n}\bar{F}(ab(1-s_{j-1})) + q_{j,n}} I_{\{y_j > a(1-s_{j-1})\}} \Bigg).
\end{aligned}$$

Using the same argument in Case 1 (replace n by τ), we obtain the bound

$$\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}) \le \frac{1}{q_{\tau,n}} \cdot \frac{1}{q} \prod_{j=1}^{\tau-1} \frac{1}{p_{j,n}} \cdot \frac{\bar{F}(ba(1-a)^{\tau-1})}{\bar{F}(b)} \tag{3.6}$$

for $y \in \mathsf{K}$ and $\tau = \tau(y) < n$.

To summarize, since $q \le q_{\tau,n}$ and $p_{j,n} \le 1$, (3.5) and (3.6) imply that

$$\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}) \le \frac{1}{q^2} \prod_{j=1}^{n-1} \frac{1}{p_{j,n}} \cdot \frac{\bar{F}(ba(1-a)^{n-1})}{\bar{F}(b)} \tag{3.7}$$

for every $y \in \mathsf{K}$. Because $\bar{F}$ is assumed to have regularly varying tails, the right-hand-side of (3.7) is bounded from above by a constant independent of b. ■

Proof of Proposition 3.1. For each $1 \le i \le n$ define the set $A_i \doteq \{\mathbf{y} \in \mathbb{R}_+^n : y_k \le a(1-s_{k-1}), 1 \le k \le i-1, s_i > 1\}$. Note that for $1 \le i \le n-1$ and for $\mathbf{y} \in A_i$, as b tends to infinity,

$$\begin{aligned} \frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}) &= \frac{1}{\bar{F}(b)} \cdot \frac{\bar{F}(ab(1-s_{i-1}))}{p_{i,n}\bar{F}(ab(1-s_{i-1})) + q_{i,n}} \prod_{j=1}^{i-1} \frac{1}{p_{j,n}} \\ &\to (a(1-s_{i-1}))^{-\alpha} \frac{1}{q_{i,n}} \prod_{j=1}^{i-1} \frac{1}{p_{j,n}}. \end{aligned} \tag{3.8}$$

Since $1 - s_{i-1} \ge (1-a)^{i-1}$ in A_i the convergence is uniform for all y in A_i. This follows from a well known theorem, which states that if $\bar{F}$ is of regular variation then

$$\lim_{x\to\infty} \frac{\bar{F}(ax)}{\bar{F}(x)}$$

is uniform for a in any compact subset of $(0,1]$ [4, page 22, Theorem 1.5.2]. Similarly note that on A_n we have the following uniform convergence:

$$\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}) = \frac{\bar{F}(b(1-s_{n-1}))}{\bar{F}(b)} \prod_{j=1}^{n-1} \frac{1}{p_{j,n}} \to (1-s_{n-1})^{-\alpha} \prod_{j=1}^{n-1} \frac{1}{p_{j,n}}. \tag{3.9}$$

Let $A \doteq \bigcup_{i=1}^n A_i$, and g be a bounded continuous function on $\mathbb{R}_+^n$ that satisfies

$$g(\mathbf{y}) \doteq \lim_b \frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y}), \quad \text{for every } \mathbf{y} \in A.$$

Such g always exists since the closures of A_i are disjoint.

Thanks to the uniform convergence and that θ^b has bounded mass, we have

$$\lim_{b\to\infty} \int_A \frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y})\theta^b(d\mathbf{y}) = \lim_{b\to\infty} \int_A g(\mathbf{y})\theta^b(d\mathbf{y}).$$

Observing that $\theta(\partial A) = 0$, the weak convergence $\theta^b \Rightarrow \theta$ implies that

$$\lim_{b\to\infty} \int_A g(\mathbf{y})\theta^b(d\mathbf{y}) = \int_A g(\mathbf{y})\theta(d\mathbf{y}),$$

as well as

$$\theta^b(A^c) \to \theta(A^c) = 0.$$

The last display, Lemma 3.4 and that supp $(\theta^b) \subset \mathsf{K}$, in turn yield

$$\lim_{b\to\infty} \int_{A^c} \frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y})\theta^b(d\mathbf{y}) = 0 = \int_{A^c} g(\mathbf{y})\theta(d\mathbf{y}).$$

It follows that

$$\lim_{b\to\infty} \int_{\mathbb{R}_+^n} \frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(b\mathbf{y})\theta^b(d\mathbf{y}) = \int_{\mathbb{R}_+^n} g(\mathbf{y})\theta(d\mathbf{y}).$$

Since the support of θ is those $\mathbf{y} = (y_1, \ldots, y_n)$ where $y_j \geq 1$ for a single j and $y_i = 0$ for $i \neq j$, it is not difficult to check that, thanks to (3.8) and (3.9),

$$\int_{\mathbb{R}_+^n} g(\mathbf{y})\theta(d\mathbf{y}) = \prod_{j=1}^{n-1} \frac{1}{p_{j,n}} + a^{-\alpha} \sum_{i=1}^{n-1} \frac{1}{q_{i,n}} \prod_{j=1}^{i-1} \frac{1}{p_{j,n}}.$$

This completes the proof. ■

3.2 Solution to the limit problem

In this section we argue that one can choose $(a, p_{i,n}, q_{i,n})$ appropriately so that the corresponding change of measure ν_n^b attains nearly asymptotically optimal relative error; see (2.4). We need the following result.

Lemma 3.5 *Given parameters $(a, p_{i,n}, q_{i,n})$, define*

$$J(a; p_{i,n}, q_{i,n}) \doteq \prod_{j=1}^{n-1} \frac{1}{p_{j,n}} + a^{-\alpha} \sum_{i=1}^{n-1} \frac{1}{q_{i,n}} \prod_{j=1}^{i-1} \frac{1}{p_{j,n}}.$$

Then for any fixed $a \in (0, 1)$, the function $J(a; \cdot, \cdot)$ is minimized at

$$p_{k,n}^* \doteq \frac{(n-k-1)a^{-\alpha/2} + 1}{(n-k)a^{-\alpha/2} + 1}, \quad q_{k,n}^* = 1 - p_{k,n}^*, \qquad 1 \leq k \leq n-1,$$

with minimum

$$J^*(a) = \Big((n-1)a^{-\alpha/2} + 1\Big)^2.$$

Proof. We use an argument of dynamic programming type. For $1 \leq k \leq n$, define

$$J_k(a; p_{i,n}, q_{i,n}) \doteq \prod_{j=k}^{n-1} \frac{1}{p_{j,n}} + a^{-\alpha} \sum_{i=k}^{n-1} \frac{1}{q_{i,n}} \prod_{j=k}^{i-1} \frac{1}{p_{j,n}},$$

and

$$V_k(a) \doteq \inf_{\{p_{i,n},q_{i,n}\}} J_k(a;p_{i,n},q_{i,n}).$$

Note that J_k is independent of those $(p_{i,n}, q_{i,n})$ where $i \leq k-1$ and that the original problem corresponds to $k = 1$ (i.e, $J = J_1$). It is not difficult to check by definition

$$J_k(a;p_{i,n},q_{i,n}) = a^{-\alpha}\frac{1}{q_{k,n}} + \frac{1}{p_{k,n}}J_{k+1}(a;p_{i,n},q_{i,n}),$$

which in turn yield the dynamic programming equation (DPE)

$$V_k(a) = \inf\left\{a^{-\alpha}\frac{1}{q_{k,n}} + \frac{1}{p_{k,n}}V_{k+1}(a) : \ p_{k,n} \geq 0, q_{k,n} > 0, p_{k,n} + q_{k,n} = 1\right\}.$$

Since $V_n(a) \equiv 1$ by definition, one can easily use backward induction (we omit the details) to show that

$$V_k(a) = \Big((n-k)a^{-\alpha/2} + 1\Big)^2,$$

and that the right-hand-side of the DPE is minimized at $(p^*_{k,n}, q^*_{k,n})$. This completes the proof. ■

The following corollary, which states the existence of nearly optimal importance sampling schemes, is immediate.

Corollary 3.6 *Let $\varepsilon > 0$ be given. Then there exists $a \in (0,1)$ such that $J^*(a) \leq (1+\varepsilon)n^2$. Let $(p^*_{i,n}, q^*_{i,n})$ be the optimal weights defined as in Lemma 3.5. Then the change of measure ν^b_n with parameters $(a, p^*_{i,n}, q^*_{i,n})$ is nearly asymptotically optimal in that*

$$\lim_{b\to\infty} \frac{E[I_{\{S_n>b\}}d\mu/d\nu^b_n(X_1,\ldots,X_n)]^2}{\bar{F}(b)^2} \leq (1+\varepsilon)n^2.$$

4 Importance sampling for random N

In this section we address the problem of estimating

$$\rho_b \doteq P(X_1 + X_2 + \cdots X_N > b)$$

where N is a $\mathbb{N}$-valued random variable that is independent of $\{X_i\}$. Throughout we assume $E[z^N] < \infty$ for some $z > 1$. Let $s_n \doteq P(N = n)$ and $c \doteq E[N]$. Observe that $\{ns_n/c\}$ defines a probability measure on $\mathbb{N}$.

Importance sampling algorithm: The scheme is parameterized by (a_0, a_1, K) where $a_0 \in (0,1)$, $a_1 \in (0, 1 - z^{-1/\alpha})$, and $K \in \mathbb{N}$. Each independent sample is constructed in the following fashion.

- Generate a random variable $\tilde{N}$ according to $P(\tilde{N} = n) = ns_n/c$.
- If $\tilde{N} = n \leq K$, then draw the random vector $(\tilde{X}_1, \ldots, \tilde{X}_n)$ from the distribution ν_n^b with parameter $(a_0, p^*_{i,n}, q^*_{i,n})$ where $(p^*_{i,n}, q^*_{i,n})$ are the optimal weights defined in Lemma 3.5 with $a = a_0$.
- If $\tilde{N} = n > K$, then draw the random vector $(\tilde{X}_1, \ldots, \tilde{X}_n)$ from the distribution ν_n^b with parameter $(a_1, p^*_{i,n}, q^*_{i,n})$ where $(p^*_{i,n}, q^*_{i,n})$ are the optimal weights defined in Lemma 3.5 with $a = a_1$.
- Define

$$\hat{\rho}_b \doteq I_{\left\{\tilde{X}_1 + \cdots + \tilde{X}_{\tilde{N}} > b\right\}} \frac{c}{\tilde{N}} \frac{d\mu}{d\nu^b_{\tilde{N}}} (\tilde{X}_1, \ldots, \tilde{X}_{\tilde{N}}).$$

The importance sampling estimator is just the sample average of independent copies of $\hat{\rho}_b$.

The following result characterizes the asymptotic performance of this importance sampling scheme.

Theorem 4.1 *Consider the importance sampling scheme with parameter* (a_0, a_1, K). *Then*

$$\lim_{b\to\infty} \frac{1}{\bar{F}(b)^2} E\left[\hat{\rho}_b^2\right] = \sum_{n=1}^{K} \frac{cs_n}{n} \left((n-1)a_0^{-\alpha/2} + 1\right)^2 + \sum_{n=K+1}^{\infty} \frac{cs_n}{n} \left((n-1)a_1^{-\alpha/2} + 1\right)^2. \tag{4.1}$$

In particular, for any $\varepsilon > 0$, *there exist* (a_0, a_1, K) *such that*

$$\lim_{b\to\infty} \frac{1}{\bar{F}(b)^2} E\left[\hat{\rho}_b^2\right] \leq (1+\varepsilon)c^2.$$

Before proceeding with the proof, let us check that this indeed describes a nearly asymptotically optimal scheme. By Jensen's inequality

$$E\left[\hat{\rho}_b^2\right] \geq (E[\hat{\rho}_b])^2 = P(S_N > b)^2.$$

Also, since the random variables X_i are subexponential and $E[z^N] < \infty$ for some $z > 1$, [1, page 259, Lemma 2.2] asserts that

$$\lim_{b\to\infty} \frac{P(S_N > b)}{\bar{F}(b)} = E[N] = c.$$

It follows that

$$\liminf_{b\to\infty} \frac{1}{\bar{F}(b)^2} E\left[\hat{\rho}_b^2\right] \geq c^2.$$

Hence such a scheme is indeed nearly asymptotically optimal.

Remark 4.1 As we will see, the introduction of the cutoff K and the use of a different parameter a_1 for $\tilde{N} > K$ are for technical reasons in order to facilitate an interchange needed in the proof. It is not known at this time if this setup is necessary, or if one can work with a single parameter $a_0 \in (0, 1)$ and $K = \infty$.

Proof of Theorem 4.1. When the samples are generated according to this scheme,

$$\begin{aligned} \frac{1}{\bar{F}(b)^2} E\left[\hat{\rho}_b^2\right] &= \frac{1}{\bar{F}(b)^2} E\left[I_{\{S_N > b\}} \frac{c}{N} \frac{d\mu}{d\nu_N^b}(X_1, \ldots, X_N) \right] \\ &= \frac{1}{\bar{F}(b)^2} \sum_{n=1}^{K} E\left[I_{\{S_n > b\}} \frac{d\mu}{d\nu_n^b}(X_1, \ldots, X_n) \right] \frac{cs_n}{n} \\ &\quad + \frac{1}{\bar{F}(b)^2} \sum_{n=K+1}^{\infty} E\left[I_{\{S_n > b\}} \frac{d\mu}{d\nu_n^b}(X_1, \ldots, X_n) \right] \frac{cs_n}{n} \end{aligned}$$

We next take b to ∞ in the previous display. Assume for now that the interchange of limit and the infinite sum is valid – the justification will be given momentarily. Then (4.1) follows immediately from Proposition 3.1 and Lemma 3.5. Since $a_0, a_1 < 1$ and $\sum_n ns_n = c$, it is not difficult to see that the right-hand-side of (4.1) is bounded from above by

$$\sum_{n=1}^{\infty} \frac{cs_n}{n} n^2 a_0^{-\alpha} + \sum_{n=K+1}^{\infty} \frac{cs_n}{n} n^2 a_1^{-\alpha} = a_0^{-\alpha} c^2 + a_1^{-\alpha} c \sum_{n=K+1}^{\infty} ns_n. \qquad (4.2)$$

For any $\varepsilon > 0$, the conclusion of the theorem follows by taking K large enough and a_0 sufficiently close to 1.

It remains to justify the interchange of limit with the infinite sum. This will be done by finding a dominating function for

$$\frac{1}{\bar{F}(b)^2} E\left[I_{\{S_n > b\}} \frac{d\mu}{d\nu_n^b}(X_1, \ldots, X_n) \right] \frac{c s_n}{n}$$

when $n > K$. Recall that in this case ν_n^b is defined with parameter $(a_1, p_{i,n}^*, q_{i,n}^*)$ where $(p_{i,n}^*, q_{i,n}^*)$ are the optimal weights given by Lemma 3.5 with $a = a_1$.

By inequality (3.7) we have,

$$\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(\mathbf{x}) \leq \frac{1}{q^2} \prod_{j=1}^{n-1} \frac{1}{p_{j,n}^*} \cdot \frac{\bar{F}(b a_1 (1 - a_1)^{n-1})}{\bar{F}(b)}$$

on the set $\{\mathbf{x} \in \mathbb{R}_+^n : x_1 + \cdots + x_n > b\}$, where q [defined in (3.3)] is

$$q \doteq \min\{q_{j,n}^*\} = \min\{1 - p_{j,n}^*\} = \frac{a_1^{-\alpha/2}}{(n-1) a_1^{-\alpha/2} + 1} < 1.$$

Using this and the particular form of the weights $p_{j,n}^*$ from Lemma 3.5,

$$\frac{1}{\bar{F}(b)} \frac{d\mu}{d\nu_n^b}(\mathbf{x}) \leq \frac{\bar{F}(b a_1 (1 - a_1)^{n-1})}{\bar{F}(b)} \left((n-1) a_1^{-\alpha/2} + 1 \right)^3 a_1^{\alpha},$$

which in turn implies

$$\begin{aligned} &\frac{1}{\bar{F}(b)^2} E\left[I_{\{S_n > b\}} \frac{d\mu}{d\nu_n^b}(X_1, \ldots, X_n) \right] \\ &\quad \leq \frac{\bar{F}(b a_1 (1 - a_1)^{n-1})}{\bar{F}(b)} \frac{P(S_n \geq b)}{\bar{F}(b)} \left((n-1) a_1^{-\alpha/2} + 1 \right)^3 a_1^{\alpha}. \end{aligned} \tag{4.3}$$

A well known result from the theory of subexponential distributions (see, e.g., [1, page 255, Lemma 1.8]) states that for all $\gamma > 0$ there is $K(\gamma)$ such that the following bound holds for all $b \geq 0$:

$$\frac{P(S_n \geq b)}{\bar{F}(b)} \leq K(\gamma)(1 + \gamma)^n. \tag{4.4}$$

Another result [4, page 25, Theorem 1.5.6] states the following: for any $\delta > 0$ there exists $A(\delta) > 1$ such that for all $0 \leq y \leq x$,

$$\frac{\bar{F}(y)}{\bar{F}(x)} \leq A(\delta) \left(\frac{x}{y} \right)^{\delta + \alpha}. \tag{4.5}$$

Now choose $\gamma, \delta > 0$ so that

$$\frac{1+\gamma}{(1-a_1)^{\alpha+\delta}} < z.$$

Such γ and δ always exists, thanks to the assumption that $0 < a_1 < 1 - z^{-1/\alpha}$. We now apply the bounds in equations (4.4) and (4.5) to inequality (4.3). Observing that $s_n \leq Cz^{-n}$ for some constant C since $E[z^N] < \infty$, it is not difficult to show that there is a finite constant $\bar{C}$ such that

$$\frac{1}{\bar{F}(b)^2} E\left[I_{\{S_n > b\}} \frac{d\mu}{d\nu_n^b}(X_1, \ldots, X_n) \right] \frac{cs_n}{n} \leq \bar{C} n^2 \beta^n,$$

where $\beta \doteq (1+\gamma)(1-a_1)^{-(\alpha+\delta)}/z < 1$. The right-hand-side then serves as a summable dominating function. ■

5 Numerical Results

In this section we present some numerical results for the estimation of

$$\rho_b \doteq P(X_1 + \cdots + X_N > b),$$

where the X_i's are iid random variables with regularly varying tails. The simulation results from the algorithms outlined in this paper are denoted by DIS (for dynamic importance sampling). For comparison, we also include results from the weighted delayed hazard twisting algorithm of [9], denoted by WDHT, and the conditional Monte Carlo algorithm from the report [2], marked as CMC.

In all the tables, N is a random variable independent of $\{X_i\}$ with distribution $P(N = n) = \rho(1-\rho)^{n-1}$ for $n \geq 1$. In Tables 1 and 2, we assume X_i has tail distribution $P(X_i > b) = (1+b)^{-\alpha}$ for various values of α, while Table 5 uses tail distribution of form $P(X_i > b) = (1+b/2)^{-2.15} \log(2+b)/\log 2$.

For the WDHT algorithm we use the parameters used in the paper [9]. For our algorithm (described in Section 4), we must choose (a_0, a_1, K). For a given $\varepsilon > 0$, we set

$$a_0 = \left(1 + \frac{\varepsilon}{2}\right)^{-1/\alpha}, \quad a_1 = \frac{1 - (1-\rho)^{1/\alpha}}{2}$$

and

$$K = \lfloor \max\{-\delta \log A, 2\delta^2\} + 1 \rfloor, \quad \text{where } \delta = \frac{-1}{\log \sqrt{1-\rho}} \text{ and } A = \frac{\varepsilon a_1^\alpha}{2(1+\rho)}.$$

Note that under this choice of (a_0, a_1, K), the conditions of Theorem 4.1 are satisfied and the right-hand-side of (4.1) is bounded from above by $(1+\varepsilon)E[N]^2$ (see the Appendix), so that the scheme is nearly asymptotically optimal. This is not the only choice that has such properties. The performance, however, does not vary much when using other choices.

It has been shown that the WDHT algorithm is logarithmically asymptotically optimal (which means that the log of the second moment divided by the log of the probability of interest converges to 2 as $b \to \infty$), and that the CMC algorithm has bounded relative error (though not necessarily nearly asymptotically optimal relative error). The numerical results show that our algorithm has the best performance for all the parameter values considered, with the standard error better than that in the CMC algorithm by at least a factor of 10.

Remark 5.1 It is not standard in the literature on this topic to report simulation results for deterministic N. However, we did test such problems, and in some cases found that the performance of our algorithm and CMC was similar. We conjecture that for these cases the CMC algorithm is actually nearly asymptotically optimal, even though there is no proof to support this conjecture. In all cases that we tested where N was random the two algorithms were not comparable, with differences similar to those in the presented examples, and thus for random N it seems that the CMC algorithm is not nearly asymptotically optimal. In all cases, both algorithms out-performed the WDHT algorithm.

Appendix

In this appendix we show that for the choice of (a_0, a_1, K) defined in Section 5, the right-hand-side of (4.1) is bounded from above by $(1+\varepsilon)E[N]^2$. Note that (4.2) gives an upper bound for the right-hand-side of (4.1). It suffices to show that for this choice of (a_0, a_1, K),

$$a_0^{-\alpha} c^2 + a_1^{-\alpha} c \sum_{n=K+1}^{\infty} n s_n \leq (1+\varepsilon) c^2,$$

which is itself implied by

$$\sum_{n=K+1}^{\infty} n s_n / c \leq \frac{\varepsilon a_1^{\alpha}}{2}.$$

Since $s_n = P(N = n) = \rho(1-\rho)^{n-1}$, $c = E[N] = 1/\rho$, we need to show

$$\sum_{n=K+1}^{\infty} n\rho^2(1-\rho)^{n-1} \leq \frac{\varepsilon a_1^\alpha}{2}.$$

But simple algebra yields

$$\sum_{n=K+1}^{\infty} n\rho^2(1-\rho)^{n-1} = (K\rho+1)(1-\rho)^K \leq K(\rho+1)(1-\rho)^K,$$

whence it remains to show

$$K(1-\rho)^K \leq A.$$

To this end, observing that $K \geq 2\delta^2$ and using inequality $e^x \geq x^2/2$, we have

$$e^{K/\delta} \geq (K/\delta)^2/2 \geq K.$$

Therefore

$$K(1-\rho)^K \leq e^{K/\delta}(1-\rho)^K = \left[e^{1/\delta}(1-\rho)\right]^K = (e^{-1/\delta})^K.$$

This completes the proof since $K \geq -\delta \log A$. ■

ρ	b	True Value	DIS	WDHT	CMC	
0.25	$1e+06$	0.004	0.003997 $2.037e-06$ [0.003993, 0.004001]	0.004018 0.0001703 [0.003678, 0.004359]	0.003954 $2.427e-05$ [0.003906, 0.004003]	Estimate Std. Error Confid. Interval
	$1e+12$	$4e-06$	$3.997e-06$ $1.879e-09$ [$3.994e-06$, $4.001e-06$]	$4.253e-06$ $2.902e-07$ [$3.673e-06$, $4.833e-06$]	$4.016e-06$ $2.44e-08$ [$3.967e-06$, $4.064e-06$]	
	$1e+18$	$4e-09$	$3.999e-09$ $1.804e-12$ [$3.995e-09$, $4.002e-09$]	$3.748e-09$ $3.549e-10$ [$3.038e-09$, $4.458e-09$]	$4.041e-09$ $2.469e-11$ [$3.992e-09$, $4.091e-09$]	
0.5	$1e+06$	0.002	0.002001 $6.507e-07$ [0.001999, 0.002002]	0.001948 $7.946e-05$ [0.001789, 0.002107]	0.00199 $1.002e-05$ [0.00197, 0.00201]	
	$1e+12$	$2e-06$	$2e-06$ $6.897e-10$ [$1.999e-06$, $2.002e-06$]	$1.941e-06$ $1.256e-07$ [$1.69e-06$, $2.192e-06$]	$1.993e-06$ $9.876e-09$ [$1.973e-06$, $2.012e-06$]	
	$1e+18$	$2e-09$	$2e-09$ $7.041e-13$ [$1.999e-09$, $2.001e-09$]	$1.887e-09$ $1.644e-10$ [$1.558e-09$, $2.216e-09$]	$1.999e-09$ $9.888e-12$ [$1.98e-09$, $2.019e-09$]	
0.75	$1e+06$	0.001333	0.001333 $3.643e-07$ [0.001332, 0.001334]	0.001372 $4.821e-05$ [0.001275, 0.001468]	0.001335 $4.698e-06$ [0.001326, 0.001345]	
	$1e+12$	$1.33e-06$	$1.333e-06$ $3.272e-10$ [$1.333e-06$, $1.334e-06$]	$1.332e-06$ $7.492e-08$ [$1.182e-06$, $1.481e-06$]	$1.343e-06$ $4.796e-09$ [$1.334e-06$, $1.353e-06$]	
	$1e+18$	$1.33e-09$	$1.334e-09$ $3.061e-13$ [$1.333e-09$, $1.334e-09$]	$1.517e-09$ $1.085e-10$ [$1.3e-09$, $1.734e-09$]	$1.334e-09$ $4.681e-12$ [$1.325e-09$, $1.344e-09$]	

Table 1. Estimates for $P(X_1 + \cdots + X_N > b)$ where $P(N = n) = \rho(1-\rho)^{n-1}$ and $P(X_i > b) = (1+b)^{-1/2}$. All the results use 20,000 iterations. The true value is obtained by running our algorithm and the CMC algorithm for 500,000 iterations. The tolerance for our algorithm is set to $\varepsilon = 0.01$.

ρ	b	True Value	DIS	WDHT	CMC	
0.25	1000	0.0001286	0.0001279 $9.87e-08$ $[0.0001277, 0.0001281]$	0.0001297 $7.482e-06$ $[0.0001148, 0.0001447]$	0.0001289 $8.049e-07$ $[0.0001273, 0.0001305]$	Estimate Std. Error Confid. Interval
	$1e+05$	$1.266e-07$	$1.266e-07$ $5.337e-11$ $[1.264e-07, 1.267e-07]$	$1.407e-07$ $1.092e-08$ $[1.189e-07, 1.626e-07]$	$1.255e-07$ $7.671e-10$ $[1.239e-07, 1.27e-07]$	
	$1e+08$	$4e-12$	$4.003e-12$ $1.527e-15$ $[4e-12, 4.006e-12]$	$3.938e-12$ $4.288e-13$ $[3.081e-12, 4.796e-12]$	$3.995e-12$ $2.443e-14$ $[3.946e-12, 4.043e-12]$	
0.5	1000	$6.352e-05$	$6.341e-05$ $1.84e-08$ $[6.337e-05, 6.345e-05]$	$6.286e-05$ $3.429e-06$ $[5.601e-05, 6.972e-05]$	$6.35e-05$ $3.206e-07$ $[6.285e-05, 6.414e-05]$	
	$1e+05$	$6.329e-08$	$6.324e-08$ $2.26e-11$ $[6.32e-08, 6.329e-08]$	$5.864e-08$ $4.629e-09$ $[4.938e-08, 6.79e-08]$	$6.352e-08$ $3.199e-10$ $[6.288e-08, 6.416e-08]$	
	$1e+08$	$2e-12$	$2.001e-12$ $6.098e-16$ $[2e-12, 2.002e-12]$	$1.745e-12$ $1.864e-13$ $[1.372e-12, 2.118e-12]$	$2.005e-12$ $1.002e-14$ $[1.985e-12, 2.025e-12]$	
0.75	1000	$4.218e-05$	$4.217e-05$ $7.219e-09$ $[4.215e-05, 4.218e-05]$	$4.02e-05$ $1.981e-06$ $[3.624e-05, 4.416e-05]$	$4.231e-05$ $1.506e-07$ $[4.201e-05, 4.261e-05]$	
	$1e+05$	$4.218e-08$	$4.217e-08$ $9.719e-12$ $[4.215e-08, 4.219e-08]$	$4.154e-08$ $2.674e-09$ $[3.619e-08, 4.689e-08]$	$4.207e-08$ $1.479e-10$ $[4.177e-08, 4.236e-08]$	
	$1e+08$	$1.33e-12$	$1.334e-12$ $3.061e-16$ $[1.333e-12, 1.334e-12]$	$1.376e-12$ $1.134e-13$ $[1.149e-12, 1.603e-12]$	$1.333e-12$ $4.69e-15$ $[1.324e-12, 1.343e-12]$	

Table 2. Estimates for $P(X_1 + \cdots + X_N > b)$ where $P(N = n) = \rho(1-\rho)^{n-1}$ and $P(X_i > b) = (1+b)^{-3/2}$. All the results use 20,000 iterations. The true value is obtained by running our algorithm and the CMC algorithm for 500,000 iterations. The tolerance for our algorithm is set to $\varepsilon = 0.01$.

ρ	b	True Value	DIS	CMC	
0.25	$1e+04$	$5.951e-07$	$5.939e-07$ $2.756e-11$ $[5.939e-07, 5.94e-07]$	$5.935e-07$ $3.665e-09$ $[5.862e-07, 6.009e-07]$	Estimate Std. Error Confid. Interval
	$1e+07$	$3.678e-13$	$3.68e-13$ $4.075e-17$ $[3.679e-13, 3.681e-13]$	$3.684e-13$ $2.261e-15$ $[3.639e-13, 3.729e-13]$	
	$1e+09$	$2.371e-17$	$2.371e-17$ $3.749e-21$ $[2.37e-17, 2.371e-17]$	$2.359e-17$ $1.449e-19$ $[2.33e-17, 2.388e-17]$	
0.5	$1e+04$	$2.965e-07$	$2.964e-07$ $1.739e-11$ $[2.964e-07, 2.965e-07]$	$2.99e-07$ $1.505e-09$ $[2.96e-07, 3.02e-07]$	
	$1e+07$	$1.84e-13$	$1.84e-13$ $1.84e-17$ $[1.839e-13, 1.84e-13]$	$1.83e-13$ $9.116e-16$ $[1.812e-13, 1.849e-13]$	
	$1e+09$	$1.186e-17$	$1.185e-17$ $1.186e-21$ $[1.185e-17, 1.186e-17]$	$1.19e-17$ $5.899e-20$ $[1.178e-17, 1.201e-17]$	
0.75	$1e+04$	$1.976e-07$	$1.975e-07$ $5.908e-12$ $[1.975e-07, 1.975e-07]$	$1.974e-07$ $7.05e-10$ $[1.96e-07, 1.988e-07]$	
	$1e+07$	$1.227e-13$	$1.227e-13$ $6.134e-18$ $[1.226e-13, 1.227e-13]$	$1.226e-13$ $4.315e-16$ $[1.217e-13, 1.235e-13]$	
	$1e+09$	$7.9e-18$	$7.903e-18$ $3.953e-22$ $[7.903e-18, 7.904e-18]$	$7.938e-18$ $2.81e-20$ $[7.881e-18, 7.994e-18]$	

Table 3. Estimates for $P(X_1+\cdots+X_N > b)$ where $P(N=n) = \rho(1-\rho)^{n-1}$ and $P(X_i > b) = (1+b/2)^{-2.15}\log(2+b)/\log 2$. All the results use 20,000 iterations. The true value is obtained by running our algorithm and the CMC algorithm for 500,000 iterations. The tolerance for our algorithm is set to $\varepsilon = 0.01$.

References

[1] S. Asmussen. *Ruin Probabilities.* World Scientific, Singapore, 2000.

[2] S. Asmussen and D. Kroese. Improved algorithms for rare event simulation with heavy tails. Technical report, The Danish National Research Foundation: Network in Mathematical Physics and Stochastics, 2004.

[3] P. Billingsley. *Convergence of Probability Measures.* John Wiley, New York, 1999.

[4] N. Bingham, C. Goldie, and J. Teugels. *Regular Variation.* Cambridge University Press, Cambridge, 1987.

[5] P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations.* John Wiley & Sons, New York, 1997.

[6] P. Dupuis and H. Wang. Importance sampling, large deviations, and differential games. *Stochastics and Stochastics Reports*, 76:481–508, 2004.

[7] P. Dupuis and H. Wang. Dynamic importance sampling for uniformly recurrent markov chains. *Annals of Applied Probability*, 15:1–38, 2005.

[8] P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extremal events for Insurance and Finance.* Springer-Verlag, Germany, 2003.

[9] S. Juneja and P. Shahabuddin. Simulating heavy tailed processes using delayed hazard rate twisting. *ACM Transactions on Modeling and Computer Simulations*, pages 94–118, 2002.