

RARE-EVENT SIMULATION OF HEAVY-TAILED RANDOM WALKS BY SEQUENTIAL IMPORTANCE SAMPLING AND RESAMPLING

HOCK PENG CHAN,* *National University of Singapore*

SHAOJIE DENG, *Microsoft*

TZE-LEUNG LAI,** *Stanford University*

Abstract

We introduce a new approach to simulating rare events for Markov random walks with heavy-tailed increments. This approach involves sequential importance sampling and resampling, and uses a martingale representation of the corresponding estimate of the rare-event probability to show that it is unbiased and to bound its variance. By choosing the importance measures and resampling weights suitably, it is shown how this approach can yield asymptotically efficient Monte Carlo estimates.

Keywords: Efficient simulation; heavy-tailed distribution; sequential Monte Carlo; regularly varying tail

2010 Mathematics Subject Classification: Primary 65C05
Secondary 60G50

1. Introduction

The past decade has witnessed many important advances in Monte Carlo methods for computing tail distributions and boundary crossing probabilities of multivariate random walks with independent and identically distributed (i.i.d.) or Markov-dependent increments; see the survey paper by Blanchet and Lam [5]. In particular, the case of heavy-tailed random walks has attracted much recent attention because of its applications to queueing and communication networks. A random variable is called *light tailed* if its moment generating function is finite in some neighborhood of the origin. It is said to be *heavy tailed* otherwise.

Another area of much recent interest is the development and the associated probability theory of the efficient Monte Carlo method to compute rare-event probabilities $\alpha_n = P(A_n)$ such that $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$. A Monte Carlo estimator $\hat{\alpha}_n$ of α_n using m simulation runs is said to be *logarithmically efficient* if

$$m\text{var}(\hat{\alpha}_n) \leq \alpha_n^{2+o(1)} \quad \text{as } n \rightarrow \infty; \quad (1)$$

it is said to be *strongly efficient* if

$$m\text{var}(\hat{\alpha}_n) = O(\alpha_n^2).$$

Strong efficiency means that, for every $\varepsilon > 0$,

$$\text{var}(\hat{\alpha}_n) \leq \varepsilon \alpha_n^2 \quad (2)$$

Received 8 September 2011; revision received 24 April 2012.

* Postal address: Department of Statistics and Applied Probability, National University of Singapore, Science Drive 1, 119260, Singapore.

** Postal address: Statistics Department, Stanford University, Stanford, CA 94305, USA.

can be achieved by using m simulation runs, with m depending on ε but not on n . In the case of logarithmic efficiency, (2) can be achieved by using m_n simulation runs, with $m_n = (\alpha_n^{-1})^{o(1)}$ to cancel the $\alpha_n^{o(1)}$ term in (1). Since the focus of this paper is on rare events associated with a random walk S_n , any Monte Carlo estimate of a rare-event probability has to generate the i.i.d. or Markov-dependent increments X_1, \dots, X_n of the random walk for each simulation run, and this computational task is linear in n . We call the Monte Carlo estimate *linearly efficient* if $m_n = O(n)$ simulation runs can be used to achieve (2). More generally, for any nondecreasing sequence of positive constants $C_n \rightarrow \infty$ such that $C_n = o(\alpha_n^{-1})$, we call the Monte Carlo estimate C_n -efficient if $m_n = O(C_n)$ simulation runs can achieve (2). Note in this connection that the variance of the direct Monte Carlo estimate of α_n using m_n independent simulation runs is $\alpha_n(1 - \alpha_n)/m_n$, and, therefore, (2) can be achieved only by choosing $m_n \geq (\varepsilon\alpha_n)^{-1}(1 - \alpha_n)$.

To achieve strong efficiency, Blanchet and Glynn [4] and Blanchet and Liu [6] made use of approximations of Doob's h -transform to develop an importance sampling method for computing $P(A)$ when the event A is related to a Markov chain Y_k that has transition probability densities $p_k(\cdot | Y_{k-1})$ with respect to some measure ν . Letting $h_k(Y_k) = P(A | Y_k)$, note that

$$E[h_k(Y_k) | Y_{k-1}] = E[P(A | Y_k) | \mathcal{F}_{k-1}] = P(A | \mathcal{F}_{k-1}) = h_{k-1}(Y_{k-1}),$$

i.e. $\int p_k(x, y)h_k(y) \, d\nu(y) = h_{k-1}(x)$. This yields the transition density

$$p_k^h(x, y) := p_k(x, y) \frac{h_k(y)}{h_{k-1}(x)} \tag{3}$$

of an importance measure $Q = P(\cdot | A)$, and p_k^h is called the h -transform of p_k . Although the likelihood ratio dP/dQ is equal to $P(A)$ and therefore has zero variance, this importance measure cannot be used in practice because $P(A)$ is the unknown probability to be estimated. On the other hand, one may be able to find a tractable approximation v_k of h_k for $k = 1, 2, \dots$ so that $p_k^h(x, y)$ can be approximated by

$$q_k(x, y) = p_k(x, y) \frac{v_k(y)}{\int p_k(x, y)v_k(y') \, d\nu(y')}, \tag{4}$$

which is the transition density function of an importance measure that can be used to perform importance sampling.

In this paper we propose a new approach to simulating rare-event probabilities for heavy-tailed random walks. This approach uses not only sequential (dynamic) importance sampling but also resampling. Chan and Lai [9] introduced the sequential importance sampling with resampling (SISR) methodology and applied it to simulate $P(g(S_n/n) \geq b)$ and $P(\max_{n_0 \leq n \leq n_1} ng(S_n/n) \geq c)$ for light-tailed random walks, where g is a general function and S_n is a random walk. Note that, unlike [3], we consider here the situation in which n approaches ∞ , rather than with fixed n . In [9], the importance measure is simply $Q = P$ and the resampling weights for the light-tailed case heavily depend on the finiteness of the moment generating function. Moreover, a distinguishing feature of a heavy-tailed random walk S_n is the possibility of a single large increment resulting in the exceedance of $g(S_n/n)$ or $\max_{n_0 \leq n \leq n_1} ng(S_n/n)$ over a threshold. An important idea underlying the SISR method to simulate rare-event probabilities for heavy-tailed random walks in Section 3 is to make use of the single large jump property to decompose the event of interest into two disjoint events, one of which involves the maximum increment being large. We use different Monte Carlo schemes to simulate these two events.

In Section 2 we describe another way of using SISR to simulate rare-event probabilities of heavy-tailed random walks. Here we start with a target importance measure, such as the one that

uses the transition density (4) to approximate the h -transform (3). The normalizing constant, which is the integral in (4), may be difficult to compute for general state spaces. Moreover, it may be difficult to sample from such density. The SISR procedure in Section 2 provides an alternative to this elaborate direct importance sampling procedure but still achieves its effect. The analysis of the two different SISR schemes for estimating rare-event probabilities, given in Sections 2 and 3, respectively, enables us to bound the variance of an SISR estimate. In Corollaries 1–5 in Sections 2 and 4 we use these bounds to show that the SISR estimates are efficient under certain regularity conditions. In Section 5 we provide numerical results to supplement the asymptotic theory and give further discussions on related literature.

2. Implementing a target importance measure by SISR

Let $\mathbf{Y}_n = (Y_1, \dots, Y_n)$, and let $p_k(\cdot \mid \mathbf{y}_{k-1})$ be the conditional density, with respect to some measure ν , of Y_k given $\mathbf{Y}_{k-1} = \mathbf{y}_{k-1}$. Let $p_n(\mathbf{y}_n) = \prod_{k=1}^n p_k(y_k \mid \mathbf{y}_{k-1})$. To evaluate a rare-event probability $\alpha = P(\mathbf{Y}_n \in \Gamma)$, direct Monte Carlo involves the generation of m independent samples from the density function $p_n(\mathbf{y}_n)$ and then estimating α by

$$\hat{\alpha}_D = m^{-1} \sum_{j=1}^m \mathbf{1}_{\{\mathbf{Y}_n^{(j)} \in \Gamma\}}.$$

Importance sampling involves the generation of m independent samples from an alternative density $\tilde{q}_k(\cdot \mid \mathbf{y}_{k-1})$ and then estimating α by

$$\hat{\alpha}_I = m^{-1} \sum_{j=1}^m \frac{p_n(\mathbf{Y}_n^{(j)}) \mathbf{1}_{\{\mathbf{Y}_n^{(j)} \in \Gamma\}}}{\tilde{q}_n(\mathbf{Y}_n^{(j)})},$$

where $\tilde{q}_n(\mathbf{y}_n) = \prod_{k=1}^n \tilde{q}_k(y_k \mid \mathbf{y}_{k-1})$ and satisfies $\tilde{q}_n(\mathbf{y}_n) > 0$ whenever $p_n(\mathbf{y}_n) \mathbf{1}_{\{\mathbf{y}_n \in \Gamma\}} > 0$. If one is able to choose \tilde{q}_n such that $p_n(\mathbf{y}_n) \mathbf{1}_{\{\mathbf{y}_n \in \Gamma\}} / \tilde{q}_n(\mathbf{y}_n) \leq c\alpha$ for some positive constant c , then one can ensure that

$$m E_{\tilde{Q}}(\hat{\alpha}_I^2) \leq c^2 \alpha^2, \tag{5}$$

yielding a strongly efficient $\hat{\alpha}_I$.

For the case in which Y_n is a random walk S_n and the rare event is $A = \{S_n \geq b\}$, a candidate for the choice of $\tilde{q}_k(\cdot \mid S_{k-1})$ is (4) in which v_k is an approximation to the h -transform. Large deviation or some other asymptotic method leads to an asymptotic approximation of the form

$$P(S_n \geq b \mid S_k) \sim g(b - S_k, n - k), \tag{6}$$

which can be used to derive v_k . As noted in Section 1, the normalizing constant (i.e. the denominator) in (4) is often difficult to evaluate and the target importance measure with transition density (4) may be difficult to sample from. We next show that we can bypass the normalizing constant by using SISR, which also enables us to weaken and generalize (6) to

$$c_n g_n(\mathbf{Y}_k, n - k) \leq P(A_n \mid \mathbf{Y}_k) \leq c'_n g_n(\mathbf{Y}_k, n - k) \tag{7}$$

for all n and k and almost all \mathbf{Y}_k , where c_n and c'_n are positive constants. In (7), \mathbf{Y}_k is a general stochastic sequence and we denote the event of interest by A_n to indicate that it is rare in the sense that $\alpha_n = P(A_n) \rightarrow 0$ as $n \rightarrow \infty$. The weakening of (6) to (7) is of particular importance for implementation since it allows one to choose g_n to be piecewise constant so that not only

can the normalizing constants in (8) below be easily computed but (8) is also convenient to sample from. Let

$$q_k(y_k | \mathbf{y}_{k-1}) = \frac{p_k(y_k | \mathbf{y}_{k-1})g_n(y_k, n - k)}{w_{k-1}(\mathbf{y}_{k-1})g_n(\mathbf{y}_{k-1}, n - k + 1)}, \tag{8}$$

in which $w_0 \equiv 1$ and $w_{k-1}(\mathbf{y}_{k-1})$ is a normalizing constant to make $q_k(\cdot | \mathbf{y}_{k-1})$ a density function for $k \geq 2$. From (7), it follows that

$$\kappa_n^{-1} \leq w_{k-1}(\mathbf{y}_{k-1}) \leq \kappa_n, \quad \text{where} \quad \kappa_n = \frac{c'_n}{c_n}. \tag{9}$$

To be more specific, we describe the SISR procedure in stages, initializing with $Y_0^{(\ell)} = y_0$, a specified initial state, or with $Y_0^{(1)}, \dots, Y_0^{(m)}$ generated from the initial distribution.

Step 1: importance sampling at stage k. Generate $\tilde{Y}_k^{(j)}$ from $q_k(\cdot | \mathbf{Y}_{k-1}^{(j)})$, and let $\tilde{\mathbf{Y}}_k^{(j)} = (\mathbf{Y}_{k-1}^{(j)}, \tilde{Y}_k^{(j)})$ for all $1 \leq j \leq m$.

Step 2: resampling at stage k. Let $\bar{w}_k = m^{-1} \sum_{\ell=1}^m w_k(\tilde{\mathbf{Y}}_k^{(\ell)})$ and the resampling weights

$$w_k^{(j)} = \frac{w_k(\tilde{\mathbf{Y}}_k^{(j)})}{m\bar{w}_k}. \tag{10}$$

Generate i.i.d. multinomial random variables b_1, \dots, b_m such that $P(b_1 = j) = w_k^{(j)}$ for $1 \leq j \leq m$. Let $\mathbf{Y}_k^{(\ell)} = \tilde{\mathbf{Y}}_k^{(b_\ell)}$ for all $1 \leq \ell \leq m$. If $k < n$, increment k by 1 and go to step 1; otherwise, end the procedure. There is no resampling at stage n .

After stage n , estimate α by

$$\hat{\alpha}_B = \frac{\bar{w}_1 \cdots \bar{w}_{n-1}}{m} \sum_{j=1}^m \frac{\mathbf{1}_{\{\tilde{\mathbf{Y}}_n^{(j)} \in \Gamma\}} g_n(Y_0^{(a_j)}, n)}{g_n(\tilde{\mathbf{Y}}_n^{(j)}, 0)}, \tag{11}$$

where $Y_0^{(a_j)}$ is the initial (ancestral) state of $\tilde{\mathbf{Y}}_n^{(j)}$. For notational simplicity, we assume a specified initial state $Y_0^{(\ell)} = y_0$ for all ℓ , and abbreviate $g_n(y_0, n)$ and α_n to g_0 and α , respectively.

Resampling is used in the above procedure to handle the normalizing constants in a target importance measure that approximates the h -transform. In [7], a computationally expensive discretization scheme, with partition width $1/n$, is used to implement the state-dependent importance sampling scheme based on the asymptotic approximation (6) in the case of regularly varying random walks. Using resampling as described in the preceding paragraph enables us to bypass the costly computation of the normalizing constants, and the SISR estimate $\hat{\alpha}_B$ is still linearly efficient in this case, as will be shown in Corollary 1 at the end of this section. More importantly, for more complicated models, one can at best expect to have approximations of the type (7) rather than the sharp asymptotic formula (6). In this case, using (8) to perform importance sampling usually does not yield a good Monte Carlo estimate because unlike the situation in (5), (7) does not imply good bounds for $\prod_{k=1}^n [p_k(y_k | \mathbf{y}_{k-1})/q_k(y_k | \mathbf{y}_{k-1})]$ on A_n . On the other hand, using (8) for the importance sampling component of an SISR procedure, whose resampling weights are proportional to $w_{k-1}(\mathbf{y}_{k-1})$, can result in a Monte Carlo estimate $\hat{\alpha}_B$ that has a bound similar to (5), which can be used to establish efficiency of the SISR procedure, as we now proceed to show.

Following [9], let E^* denote the expectation with respect to the probability measure from which the $\tilde{Y}_k^{(i)}$ and $Y_k^{(i)}$ are drawn; this differs from E_Q for importance sampling from the measure Q since it involves both importance sampling and resampling. A key tool for the analysis of the SISR estimate $\hat{\alpha}_B$ is the following martingale representation of $m(\hat{\alpha}_B - \alpha)$; see Section 2 of [9].

Lemma 1. *Let $f_k(y_k) = P(Y_n \in \Gamma \mid Y_k = y_k)$ for $1 \leq k \leq n - 1$, $f_0 = \alpha$, and $f_n(y_n) = \mathbf{1}_{\{y_n \in \Gamma\}}$. Let*

$$\begin{aligned} \mathcal{F}_{2k-1} &= \sigma(\{(\tilde{Y}_t^{(j)}, Y_t^{(j)}): 1 \leq j \leq m, 1 \leq t \leq k - 1\} \cup \{\tilde{Y}_k^{(j)}: 1 \leq j \leq m\}), \\ \mathcal{F}_{2k} &= \sigma(\{(\tilde{Y}_t^{(j)}, Y_t^{(j)}): 1 \leq j \leq m, 1 \leq t \leq k\}). \end{aligned}$$

Let $\#_k^{(j)}$ be the number of copies of $\tilde{Y}_k^{(j)}$ generated during the k th resampling stage. Define

$$\begin{aligned} \eta_{2k-1}^{(j)} &= (g_0 \bar{w}_1 \cdots \bar{w}_{k-1}) \\ &\quad \times \left[\frac{f_k(\tilde{Y}_k^{(j)})}{g_n(\tilde{Y}_k^{(j)}, n - k)} - \frac{f_{k-1}(Y_{k-1}^{(j)})}{w_{k-1}(Y_{k-1}^{(j)})g_n(Y_{k-1}^{(j)}, n - k + 1)} \right], \quad 1 \leq k \leq n, \quad (12) \\ \eta_{2k}^{(j)} &= (\#_k^{(j)} - m w_k^{(j)}) (g_0 \bar{w}_1 \cdots \bar{w}_k) \frac{f_k(\tilde{Y}_k^{(j)})}{w_k(\tilde{Y}_k^{(j)})g_n(\tilde{Y}_k^{(j)}, n - k)}, \quad 1 \leq k \leq n - 1. \end{aligned}$$

Then $\{(\eta_t^{(1)}, \dots, \eta_t^{(m)}): 1 \leq t \leq 2n - 1\}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t, 1 \leq t \leq 2n - 1\}$. Moreover,

$$m(\hat{\alpha}_B - \alpha) = \sum_{t=1}^{2n-1} \sum_{j=1}^m \eta_t^{(j)}. \quad (13)$$

Proof. By (8),

$$\begin{aligned} E^* \left[\frac{f_k(\tilde{Y}_k^{(j)})}{g_n(\tilde{Y}_k^{(j)}, n - k)} \mid \mathcal{F}_{2k-2} \right] &= E_Q \left[\frac{f_k(\tilde{Y}_k^{(j)})}{g_n(\tilde{Y}_k^{(j)}, n - k)} \mid Y_{k-1}^{(j)} \right] \\ &= \frac{E[f_k(\tilde{Y}_k^{(j)}) \mid Y_{k-1}^{(j)}]}{w_{k-1}(Y_{k-1}^{(j)})g_n(Y_{k-1}^{(j)}, n - k + 1)} \\ &= \frac{f_{k-1}(Y_{k-1}^{(j)})}{w_{k-1}(Y_{k-1}^{(j)})g_n(Y_{k-1}^{(j)}, n - k + 1)}. \end{aligned} \quad (14)$$

Since $g_0 \bar{w}_1 \cdots \bar{w}_{k-1}$ is measurable with respect to \mathcal{F}_{2k-2} , $E^*[\eta_{2k-1}^{(j)} \mid \mathcal{F}_{2k-2}] = 0$ by (14). Moreover, note that $E^*[\#_k^{(j)} \mid \mathcal{F}_{2k-1}] = m w_k^{(j)}$ and that $g_0 \bar{w}_1 \cdots \bar{w}_k$ and $\tilde{Y}_k^{(j)}$ are measurable with respect to \mathcal{F}_{2k-1} . Therefore, $E^*[\eta_{2k}^{(j)} \mid \mathcal{F}_{2k-1}] = 0$.

Theorem 1. *If (7) holds then $\hat{\alpha}_B$ is unbiased and*

$$m \text{var}^*(\hat{\alpha}_B^2) \leq n(\kappa_n^4 + \kappa_n^6) \left(1 + \frac{\kappa_n^2}{m} \right)^{n-2} \alpha^2.$$

Hence, the SISR estimate $\hat{\alpha}_B$ of α is $n\kappa_n^6$ -efficient.

Proof. Since $\eta_t^{(i)}$ is a martingale difference sequence by Lemma 1, it follows from (13) that $\hat{\alpha}_B$ is unbiased. Moreover, as shown in Example 1 of [9], all terms on the right-hand side of (13) are either uncorrelated or negatively correlated with each other, and, therefore,

$$m^2 \text{var}^*(\hat{\alpha}_B^2) \leq \sum_{t=1}^{2n-1} \sum_{j=1}^m \text{var}^*(\eta_t^{(j)}). \tag{15}$$

Since $f_k(\mathbf{y}_k) = P(A_n | \mathbf{Y}_k)$ with $A_n = \{Y_n \in \Gamma\}$, $f_k(\mathbf{y}_k)/g_n(\mathbf{y}_k, n-k) \leq c'_n$ by (7); moreover, $g_0 \leq \alpha/c_n$. Hence,

$$\frac{g_0 f_k(\mathbf{y}_k)}{g_n(\mathbf{y}_k, n-k)} \leq \kappa_n \alpha. \tag{16}$$

By (9), (12), and (14),

$$\text{var}^*(\eta_{2k-1}^{(j)}) \leq E^*(\eta_{2k-1}^{(j)})^2 \leq \kappa_n^2 \alpha^2 E^*[\bar{w}_1^2 \cdots \bar{w}_{k-1}^2]. \tag{17}$$

Similarly, since $E^*[(\#_k^{(j)})^2 | \mathcal{F}_{2k-1}] = m w_k^{(j)}$ and $\sum_{j=1}^m w_k^{(j)} = 1$,

$$\sum_{j=1}^m \text{var}^*(\eta_{2k}^{(j)}) \leq m \kappa_n^4 \alpha^2 E^*[\bar{w}_1^2 \cdots \bar{w}_k^2]. \tag{18}$$

By (8), for $0 \leq s \leq k$,

$$\prod_{\ell=s}^k q_{\ell+1}(y_{\ell+1} | \mathbf{y}_\ell) = \frac{\{\prod_{\ell=s}^k p_{\ell+1}(y_{\ell+1} | \mathbf{y}_\ell)\} g_n(\mathbf{y}_{k+1}, n-k-1)}{\{\prod_{\ell=s}^k w_\ell(\mathbf{y}_\ell)\} g_n(\mathbf{y}_s, n-s)},$$

and, therefore, by (7),

$$E_Q \left[\prod_{\ell=s}^k w_\ell(\mathbf{Y}_\ell) \mid \mathbf{Y}_s \right] \leq \kappa_n \quad \text{for } 0 \leq s \leq k. \tag{19}$$

Theorem 1 follows from (15)–(18) and Lemma 2 below.

Lemma 2. *If (19) holds then $E^*[\bar{w}_1^2 \cdots \bar{w}_k^2] \leq \kappa_n^2 (1 + m^{-1} \kappa_n^2)^{k-1}$.*

Proof. By (9),

$$\bar{w}_k^2 \leq m^{-2} \sum_{u \neq v} w_k(\tilde{\mathbf{Y}}_k^{(u)}) w_k(\tilde{\mathbf{Y}}_k^{(v)}) + m^{-1} \kappa_n^2.$$

Hence, by the independence of $\tilde{\mathbf{Y}}_k^{(u)}$ and $\tilde{\mathbf{Y}}_k^{(v)}$ conditioned on \mathcal{F}_{2k-2} ,

$$E^*[\bar{w}_k^2 | \mathcal{F}_{2k-2}] \leq m^{-2} \sum_{u \neq v} E_Q[w_k(\mathbf{Y}_k) | \mathbf{Y}_{k-1}^{(u)}] E_Q[w_k(\mathbf{Y}_k) | \mathbf{Y}_{k-1}^{(v)}] + m^{-1} \kappa_n^2. \tag{20}$$

Since $\mathbf{Y}_{k-1}^{(u)}$ is sampled from $\tilde{\mathbf{Y}}_{k-1}^{(i)}$ with probability $w_{k-1}^{(i)} = w_{k-1}(\tilde{\mathbf{Y}}_{k-1}^{(i)})/(m\bar{w}_{k-1})$,

$$\begin{aligned} E^*[E_Q[w_k(\mathbf{Y}_k) | \mathbf{Y}_{k-1}^{(u)}] | \mathcal{F}_{2k-3}] &= \sum_{i=1}^m w_{k-1}^{(i)} E_Q[w_k(\mathbf{Y}_k) | \tilde{\mathbf{Y}}_{k-1}^{(i)}] \\ &= \frac{1}{m\bar{w}_{k-1}} \sum_{i=1}^m E_Q \left[\prod_{\ell=k-1}^k w_\ell(\mathbf{Y}_\ell) \mid \tilde{\mathbf{Y}}_{k-1}^{(i)} \right]. \end{aligned} \tag{21}$$

By (19)–(21),

$$\begin{aligned}
 & E^*[\bar{w}_1^2 \cdots \bar{w}_k^2 \mid \mathcal{F}_{2k-3}] \\
 & \leq \bar{w}_1^2 \cdots \bar{w}_{k-2}^2 \left\{ m^{-1} \sum_{i=1}^m E_Q \left[\prod_{\ell=k-1}^k w_\ell(\mathbf{Y}_\ell) \mid \tilde{\mathbf{Y}}_{k-1}^{(i)} \right] \right\}^2 \\
 & \quad + m^{-1} \kappa_n^2 E^*[\bar{w}_1^2 \cdots \bar{w}_{k-1}^2 \mid \mathcal{F}_{2k-3}] \\
 & \leq \bar{w}_1^2 \cdots \bar{w}_{k-2}^2 m^{-2} \sum_{u \neq v} E_Q \left[\prod_{\ell=k-1}^k w_\ell(\mathbf{Y}_\ell) \mid \tilde{\mathbf{Y}}_{k-1}^{(u)} \right] E_Q \left[\prod_{\ell=k-1}^k w_\ell(\mathbf{Y}_\ell) \mid \tilde{\mathbf{Y}}_{k-1}^{(v)} \right] \\
 & \quad + m^{-1} \kappa_n^2 E^*[\bar{w}_1^2 \cdots \bar{w}_{k-1}^2 + \bar{w}_1^2 \cdots \bar{w}_{k-2}^2 \mid \mathcal{F}_{2k-3}].
 \end{aligned}$$

Conditioning successively on $\mathcal{F}_{2k-4}, \mathcal{F}_{2k-5}, \dots$ then yields

$$\begin{aligned}
 E^*[\bar{w}_1^2 \cdots \bar{w}_k^2] & \leq \left\{ E_Q \left[\prod_{\ell=1}^k w_\ell(\mathbf{Y}_\ell) \right] \right\}^2 + m^{-1} \kappa_n^2 E^*[\bar{w}_1^2 \cdots \bar{w}_{k-1}^2 + \cdots + \bar{w}_1^2] \\
 & \leq \kappa_n^2 + m^{-1} \kappa_n^2 E^*[\bar{w}_1^2 \cdots \bar{w}_{k-1}^2 + \cdots + \bar{w}_1^2],
 \end{aligned}$$

from which the desired conclusion follows by induction.

We next give an application of Theorem 1 in which $\kappa_n = O(1)$. A distribution function F is said to be *regularly varying* with index $\gamma > 0$ if

$$\bar{F}(x) \sim x^{-\gamma} L(x) \quad \text{as } x \rightarrow \infty \tag{22}$$

for some slowly varying function L , that is, $\lim_{x \rightarrow \infty} L(tx)/L(x) = 1$ for all $t > 0$. Suppose that $E[X] = \mu$ and $\text{var}(X) = \sigma^2 < \infty$. Let

$$g^*(b, n) = n \bar{F}(b - (n - 1)\mu) \mathbf{1}_{\{b - n\mu \geq \sigma \sqrt{n}\}} + \bar{\Phi} \left(\frac{b - n\mu}{\sigma \sqrt{n}} \right), \tag{23}$$

in which Φ denotes the standard normal distribution. Rozovskiĭ [18] showed that if F is regularly varying then

$$P(S_n \geq b) \sim g^*(b, n) \quad \text{as } n \rightarrow \infty \text{ uniformly over } b \in \mathbb{R}. \tag{24}$$

By (24), (6) holds with $g = g^*$. The usefulness of weakening (6) to (7) which only requires bounds is that g_n can be chosen to be considerably simpler than g^* . In particular, we can discretize g^* and define

$$g_n(\mathbf{Y}_n, n - k) = g^*(\zeta_i, n - k) \quad \text{if } \zeta_i \leq b - S_k < \zeta_{i+1}, \tag{25}$$

where $\zeta_{-2} = -\infty, \zeta_{-1} = 0$, and $\zeta_i = \eta^i$ for some $\eta > 1$ and all $i \geq 0$.

Corollary 1. *Let X_1, X_2, \dots be i.i.d. with distribution function F satisfying (22) and such that $\text{var}(X_1) < \infty$. Consider the SISR procedure with importance density (8), in which g_n is given by (25), and with resampling weights (10). Then the SISR estimate of $\alpha_n = P(S_n \geq b)$ is linearly efficient.*

Proof. From (22) and (24), it follows that (7) holds with $c_n = c < 1 < c' = c'_n$. Putting $\kappa_n = c'/c$ in Theorem 1 yields Corollary 1.

3. SISR schemes via truncation and tilting for heavy-tailed random walks

Let X, X_1, X_2, \dots be i.i.d. with a common distribution function F . Let $S_n = \sum_{k=1}^n X_k$ and $M_n = \max_{1 \leq k \leq n} X_k$. Let $\tau_b = \inf\{n : S_n \geq b\}$. Assume that

$$\bar{F}(x) [= 1 - F(x)] = e^{-\Psi(x)}, \quad \text{with} \quad \psi(x) = \Psi'(x) \rightarrow 0 \quad \text{as } x \rightarrow \infty.$$

Then $\Psi(x) = o(x)$ and F is heavy tailed, with density function

$$f(x) = \psi(x)e^{-\Psi(x)}.$$

We use Ψ to develop general SISR procedures for simulating the probabilities

$$p = P(S_n \geq b), \quad \alpha = P\left(\max_{1 \leq j \leq n} S_j \geq b\right) = P(\tau_b \leq n).$$

These algorithms are shown to be linearly efficient in Section 4 as $b = b_n$ approaches ∞ with n , under certain conditions for which asymptotic approximations to p and α have been developed. Unlike the SISR procedures in Section 2 that are based on (6) or its relaxation (7), the SISR procedures based on Ψ do not make explicit use of the asymptotic approximations to p and α . On the other hand, these approximations guide the choice of importance measure and the truncation in the SISR procedure.

3.1. Truncation and tilting measures for evaluating p by SISR

To evaluate p , we express it as the sum of probabilities of two disjoint events

$$A_1 = \{S_n \geq b, M_n \leq c_b\}, \quad A_2 = \{S_n \geq b, M_n > c_b\},$$

for which the choice of c_b (which tends to ∞ as $b \rightarrow \infty$) will be discussed in Theorem 2 and in Sections 4 and 5. Juneja [16] applied a similar decomposition in the special case of nonnegative regularly varying random walks, and efficiency was achieved with $c_b = b$ and with fixed n . However, the rare events considered herein involve $n \rightarrow \infty$, which requires a more elaborate method to evaluate $P(A_1)$.

Let $\theta_b = \Psi(b)/b$ and $\pi_b = \int_1^{c_b} x^{-2} dx (\leq 1)$, $0 < r < 1$, and define the mixture density

$$q(x) = rf(x) + \frac{1-r}{\pi_b x^2} \mathbf{1}_{\{1 \leq x \leq c_b\}}. \tag{26}$$

Let \hat{p}_1 be the SISR estimate of $P(A_1)$, with importance density (26) and resampling weights

$$w_k(\mathbf{X}_k) = \frac{e^{\theta_b X_k} f(\mathbf{X}_k)}{q(\mathbf{X}_k)} \mathbf{1}_{\{X_k \leq c_b\}}. \tag{27}$$

Specifically, instead of using (8) to define $q_k(\cdot | \mathbf{y}_{k-1})$, we define $q_k(\cdot | \mathbf{y}_{k-1})$ by (26) for the importance sampling step at stage k in the third paragraph of Section 2. Moreover, we now use (27) instead of (9) to define the resampling weights and perform resampling even at stage n . The counterpart of (11) now takes the simple form

$$\hat{p}_1 = (\bar{w}_1 \cdots \bar{w}_n) m^{-1} \sum_{j=1}^m e^{-\theta_b S_n^{(j)}} \mathbf{1}_{\{S_n^{(j)} \geq b\}}, \tag{28}$$

where $\bar{w}_k = m^{-1} \sum_{j=1}^m w_k(\tilde{X}_k^{(j)})$; see Equations (2.3) and (2.4) of [9]. As in Equations (2.4) and (2.5) of [9], define

$$Z_k(\mathbf{x}_k) = \left[\prod_{t=1}^k \frac{f(x_t)}{q(x_t)} \right] P(A_1 | \mathbf{x}_k), \quad h_k(\mathbf{x}_k) = \prod_{t=1}^k \frac{\bar{w}_t}{w_t(\mathbf{x}_t)}, \quad w_k^{(j)} = \frac{w_k(\tilde{X}_k^{(j)})}{m\bar{w}_k}, \tag{29}$$

with $Z_0 = \alpha$ and $h_0 = 1$. Then Equation (2.10) of [9] gives the martingale decomposition

$$m[\hat{p}_1 - P(A_1)] = \sum_{t=1}^{2n} \xi_t, \tag{30}$$

where

$$\xi_{2k-1} = \sum_{j=1}^m [Z_k(\tilde{X}_k^{(j)}) - Z_{k-1}(\mathbf{X}_{k-1}^{(j)})] h_{k-1}(\mathbf{X}_{k-1}^{(j)})$$

and

$$\xi_{2k} = \sum_{j=1}^m (\#_k^{(j)} - m w_k^{(j)}) Z_k(\tilde{X}_k^{(j)}) h_k(\tilde{X}_k^{(j)}),$$

and $\#_k^{(j)}$ is the number of copies of $\tilde{X}_k^{(j)}$ in the k th resampling step.

Theorem 2. Let $\zeta_b = E_Q[w_1(X_1)]$. Suppose that one of the following conditions is satisfied:

(C) $\int_1^{c_b} \psi^2(x) x^2 e^{2[\theta_b x - \Psi(x)]} dx = O(1)$,

(C') $\int_{-\infty}^{c_b} \psi(x) e^{2\theta_b x - \Psi(x)} dx = O(1)$,

as $b \rightarrow \infty$. Then there exists a constant $K > 0$ such that, for all large b ,

$$\text{var}(\hat{p}_1) \leq \frac{Kn}{m} \zeta_b^{2n} e^{Kn/m} P^2(X > b).$$

Proof. We will show that

$$P(S_t \geq x, M_t \leq c_b) \leq \zeta_b^t e^{-\theta_b x} \quad \text{for all } t \geq 1, x \in \mathbb{R}. \tag{31}$$

Let G be the distribution function with density

$$g(x) = \zeta_b^{-1} e^{\theta_b x} f(x) \mathbf{1}_{\{x \leq c_b\}}.$$

Let E_G denote the expectation under which X_1, \dots, X_t are i.i.d. with distribution G . Then

$$P(S_t \geq x, M_t \leq c_b) = E_G \left[\left[\prod_{k=1}^t \frac{f(X_k)}{g(X_k)} \right] \mathbf{1}_{\{S_t \geq x\}} \right] = \zeta_b^t E_G [e^{-\theta_b S_t} \mathbf{1}_{\{S_t \geq x\}}],$$

and (31) indeed holds.

In the martingale decomposition (30), the summands are either uncorrelated or negatively correlated with each other, as shown in Example 1 of [9]. Therefore,

$$\begin{aligned} E^*[\hat{p}_1 - P(A_1)]^2 &\leq m^{-1} \sum_{k=1}^n E^*[Z_k^2(\tilde{X}_k^{(1)}) h_{k-1}^2(\mathbf{X}_{k-1}^{(1)})] \\ &\quad + m^{-1} \sum_{k=1}^n E^*[(\#_k^{(1)} - m w_k^{(1)})^2 Z_k^2(\tilde{X}_k^{(1)}) h_k^2(\tilde{X}_k^{(1)})]. \end{aligned} \tag{32}$$

Let $s_k = x_1 + \dots + x_k$. Since

$$P(A_1 \mid \mathbf{x}_k) = P(S_{n-k} \geq b - s_k, M_{n-k} \leq c_b) \mathbf{1}_{\{\max(x_1, \dots, x_k) \leq c_b\}},$$

it follows from (27), (29), and (31) that

$$\begin{aligned} & E^*[Z_k^2(\tilde{\mathbf{X}}_k^{(1)})h_{k-1}^2(\mathbf{X}_{k-1}^{(1)}) \mid \mathbf{X}_{k-1}^{(1)} = \mathbf{x}_{k-1}] \\ &= \bar{w}_1^2 \dots \bar{w}_{k-1}^2 E_Q \left[\frac{f^2(X)e^{-2\theta_b s_{k-1}} P^2(A_1 \mid \mathbf{X}_k = (\mathbf{x}_{k-1}, X))}{q^2(X)} \right] \\ &\leq \bar{w}_1^2 \dots \bar{w}_{k-1}^2 \zeta_b^{2n-2k} e^{-2\theta_b b} E_Q \left[\frac{f^2(X)e^{2\theta_b X}}{q^2(X)} \mathbf{1}_{\{X \leq c_b\}} \right] \\ &= \bar{w}_1^2 \dots \bar{w}_{k-1}^2 \zeta_b^{2n-2k} e^{-2\theta_b b} E_Q[w_1^2(X_1)]. \end{aligned} \tag{33}$$

By the independence of the X_k in (27),

$$\begin{aligned} E^*[\bar{w}_1^2 \dots \bar{w}_{k-1}^2] &= [E_Q(\bar{w}_1^2)]^{k-1} \\ &= \left(\zeta_b^2 + \frac{\text{var}_Q(w_1(X_1))}{m} \right)^{k-1} \\ &\leq \zeta_b^{2k-2} \exp\left(\frac{(k-1) E_Q[w_1^2(X_1)]}{m \zeta_b^2}\right). \end{aligned} \tag{34}$$

Since $c_b \rightarrow \infty$ as $b \rightarrow \infty$, $\zeta_b \geq 1 + o(1)$. Moreover, $e^{-2\theta_b b} = P^2(X > b)$. Hence, it follows from (33), (34), and Lemma 3 below that there exists $K_1 > 0$ such that

$$m^{-1} \sum_{k=1}^n E^*[Z_k^2(\tilde{\mathbf{X}}_k^{(1)})h_{k-1}^2(\mathbf{X}_{k-1}^{(1)})] \leq \frac{K_1 n}{m} \zeta_b^{2n} \exp\left(\frac{K_1 n}{m}\right) P^2(X > b). \tag{35}$$

By (31),

$$Z_k^2(\tilde{\mathbf{X}}_k^{(j)})h_k^2(\tilde{\mathbf{X}}_k^{(j)}) = \bar{w}_1^2 \dots \bar{w}_k^2 e^{-2\theta_b \tilde{S}_k^{(j)}} P^2(A_1 \mid \tilde{\mathbf{X}}_k^{(j)}) \leq \bar{w}_1^2 \dots \bar{w}_k^2 \zeta_b^{2n-2k} e^{-2\theta_b b}. \tag{36}$$

Since

$$\text{var}(\#_k^{(j)} \mid \mathcal{F}_{2k-1}) \leq m w_k^{(j)} \quad \text{and} \quad \sum_{j=1}^m w_k^{(j)} = 1,$$

by (36),

$$\begin{aligned} & E^*[(\#_k^{(1)} - m w_k^{(1)})^2 Z_k^2(\tilde{\mathbf{X}}_k^{(1)})h_k^2(\tilde{\mathbf{X}}_k^{(1)})] \\ &= m^{-1} \sum_{j=1}^m E^*[(\#_k^{(j)} - m w_k^{(j)})^2 Z_k^2(\tilde{\mathbf{X}}_k^{(j)})h_k^2(\tilde{\mathbf{X}}_k^{(j)})] \\ &\leq \zeta_b^{2n-2k} e^{-2\theta_b b} E^* \left[\left(\sum_{j=1}^m w_k^{(j)} \right) \bar{w}_1^2 \dots \bar{w}_k^2 \right] \\ &= \zeta_b^{2n-2k} e^{-2\theta_b b} E^*(\bar{w}_1^2 \dots \bar{w}_k^2). \end{aligned} \tag{37}$$

Combining (34) with (37) and applying (35), we then obtain Theorem 2 from (32).

Lemma 3. *Under the assumptions of Theorem 2,*

$$E_Q[w_1^2(X_1)] = \dots = E_Q[w_n^2(X_n)] = O(1) \quad \text{as } b \rightarrow \infty. \tag{38}$$

Proof. First assume that (C) holds. Then

$$\begin{aligned} E_Q[w_1^2(X_1)] &= \int_{-\infty}^{c_b} \frac{e^{2\theta_b x} f^2(x)}{q(x)} dx \\ &\leq \frac{e^{2\theta_b}}{r} \int_{-\infty}^1 f(x) dx + \frac{1}{1-r} \int_1^{c_b} \psi^2(x) x^2 e^{2[\theta_b x - \Psi(x)]} dx. \end{aligned}$$

As $b \rightarrow \infty$, $\theta_b = \Psi(b)/b \rightarrow 0$ and, therefore, the first summand in the above inequality converges to $F(1)/r$. Moreover, by (C), the integral in the second summand is $O(1)$, proving (38) in this case.

Next assume that (C') holds. Since $E_Q[w_1^2(X_1)] \leq r^{-1} \int_{-\infty}^{c_b} e^{2\theta_b x} f(x) dx$ and $f(x) = \psi(x)e^{-\Psi(x)}$, (38) follows similarly. In fact, under (C'), (38) still holds when $r = 1$ in (26), i.e. when q is the original density f . Therefore, if (C') holds then Theorem 2 still holds with $q = f$.

We next evaluate $P(A_2)$ by using importance sampling that draws \mathbf{X}_n from a measure \tilde{Q} for which

$$\frac{d\tilde{Q}}{dP}(\mathbf{X}_n) = \frac{\#\{i : X_i > c_b\}}{n P(X > c_b)} \quad \text{on } \{M_n > c_b\}. \tag{39}$$

Letting $F(x | X > c) = P(c < X \leq x) / P(X > c)$, we carry out m simulation runs, each using the following procedure.

1. Choose an index $k \in \{1, \dots, n\}$ at random.
2. Generate $X_k \sim F(\cdot | X > c_b)$ and $X_i \sim F$ for $i \neq k$.

This sampling procedure indeed draws from the measure \tilde{Q} as the factor $\#\{i : X_i > c_b\}$ in the likelihood ratio (39) corresponds to assigning equal probability to each component X_i of \mathbf{X}_n that exceeds c_b to be the maximum M_n on $\{M_n > c_b\}$. We estimate $P(A_2)$ by the average \hat{p}_2 of the m independent realizations of

$$\frac{n P(X > c_b)}{\#\{i : X_i > c_b\}} \mathbf{1}_{\{S_n \geq b\}} \tag{40}$$

given by the m simulation runs. Note that \hat{p}_2 is an importance sampling estimate and is therefore unbiased. Since the denominator in (40) is at least 1 under the measure \tilde{Q} , (40) $\leq n P(X > c_b)$, yielding the variance bound

$$\text{var}(\hat{p}_2) \leq \frac{n^2 P^2(X > c_b)}{m}. \tag{41}$$

3.2. Truncations and tilting measures for SISR estimates of α

We are interested here in the Monte Carlo evaluation of $P(\max_{1 \leq j \leq n} S_j \geq b)$ as $b, n \rightarrow \infty$, when $E[X] \leq 0$. It is technically easier to consider the equivalent case of evaluating $P(S_j \geq b + ja$ for some $1 \leq j \leq n)$ in the case where $E[X] = 0$ and $a \geq 0$. More generally, consider the evaluation of $P(\tau_b \leq n)$, where $\tau_b = \inf\{j : S_j \geq b(j)\}$ and $b(j)$ is monotone increasing,

e.g. $b(j) = b + ja$. Let c_b be monotone increasing in b , and let $n_i^* = \min\{j : b(j) \geq 2^i\}$ and $n_i = \min(n_i^*, n)$. Let

$$\begin{aligned} \mathcal{A}_{1,i} &= \{n_i \leq \tau_b < n_{i+1}, X_k \leq c_{b(k)} \text{ for all } 1 \leq k \leq \tau_b\}, \\ \mathcal{A}_2 &= \{\tau_b \leq n, X_k > c_{b(k)} \text{ for some } 1 \leq k \leq \tau_b\}. \end{aligned}$$

Let $\theta_i = \Psi(2^i)/2^i$. Let $\hat{\alpha}_{1,i}$ be the SISR estimate of $P(\mathcal{A}_{1,i})$, with importance density for X_k of the form

$$q_k(x) = rf(x) + \frac{1-r}{\pi_{b(k)}x^2} \mathbf{1}_{\{1 \leq x \leq c_{b(k)}\}}, \tag{42}$$

and with resampling weights

$$w_{k,i}(X_k) = \begin{cases} \frac{e^{\theta_i X_k} f(X_k)}{q_k(X_k)} \mathbf{1}_{\{X_k \leq c_{b(k)}\}} & \text{for } 1 \leq k \leq \tau_b, \\ 1 & \text{otherwise.} \end{cases} \tag{43}$$

Note the similarity between (26)–(27) and (42)–(43). In fact, the latter just replaces c_b, θ_b , and q in (26)–(27) by $c_{b(k)}, \theta_i$, and q_k . Using an argument similar to the proof of Theorem 2, we can extend (28) to obtain a similar variance bound for $\hat{\alpha}_{1,i}$ in the following.

Theorem 3. Let $\zeta_i^* = \max\{1, \int_{-\infty}^{c_{2^{i+1}}} e^{\theta_i x} f(x) dx\}$. Suppose that $\hat{\alpha}_{1,i}$ is based on m_i SISR samples. Suppose that one of the following conditions is satisfied:

- (A) $\int_1^{c_{2^{i+1}}} \psi^2(x)x^2 e^{2[\theta_i x - \Psi(x)]} dx = O(1)$,
- (A') $\int_{-\infty}^{c_{2^{i+1}}} \psi(x)e^{2\theta_i x - \Psi(x)} dx = O(1)$,

as $i \rightarrow \infty$. Then there exists a constant $K > 0$ such that, for all large i ,

$$\text{var}(\hat{\alpha}_{1,i}) \leq \frac{Kn_{i+1}}{m_i} (\zeta_i^*)^{n_{i+1}} e^{Kn_{i+1}/m_i} P^2(X > 2^i).$$

To evaluate $P(\mathcal{A}_2)$, we perform m simulations, each using the following procedure.

1. Choose an index $k \in \{1, \dots, n\}$ with probability $\bar{F}(c_{b(k)}) / \sum_{j=1}^n \bar{F}(c_{b(j)})$.
2. Generate $X_k \sim F(\cdot \mid X > c_{b(k)})$ and $X_j \sim F$ for $j \neq k$.

We estimate $P(\mathcal{A}_2)$ by the average $\hat{\alpha}_2$ of m independent realizations of

$$\frac{[\sum_{k=1}^n \bar{F}(c_{b(k)})] \mathbf{1}_{\mathcal{A}_2}}{\#\{k : X_k > c_{b(k)}\}}$$

given by the m simulation runs. Analogous to (41), we have the following variance bound for $\hat{\alpha}_2$.

Lemma 4. Suppose that $\bar{F}(c_{b(k)}) = O(\bar{F}(b(k)))$ as $b \rightarrow \infty$, uniformly in $1 \leq k \leq n$. Then

$$m\text{var}(\hat{\alpha}_2) \leq \left[\sum_{k=1}^n \bar{F}(c_{b(k)}) \right]^2 = O\left(\left[\sum_{k=1}^n \bar{F}(b(k)) \right]^2 \right).$$

4. Efficiency of SISR schemes via truncation and tilting

In this section we apply the bounds in Theorems 1–3 to show that the SISR procedures in Section 3 give efficient estimates of p and α when we have asymptotic lower bounds to these quantities for certain classes of heavy-tailed random walks.

4.1. Regularly varying tails

In Section 3.1 we proposed an alternative SISR procedure that involves a truncation scheme and established in Theorem 2 and (41) upper bounds for $\text{var}(\hat{p}_1) + \text{var}(\hat{p}_2)$, which can be used to prove linear efficiency of the procedure, in the case of b being some power of n . This is the content of the following corollary, which gives a stronger result than linear efficiency.

Corollary 2. *Assume that (22) holds and that there exists $J > 0$ for which*

$$\psi(x) = \Psi'(x) \leq \frac{J}{x} \quad \text{for all large } x. \tag{44}$$

Assume that, for some $0 < \beta < \gamma$ with $\beta \leq 2$, $n = O(b^\beta / (\log b)^\beta)$ and $E(X^-)^\beta < \infty$. For the case $\beta > 1$, also assume that $E X = 0$. Then the estimate $\hat{p}_1 + \hat{p}_2$ of p is linearly efficient if $c_b = \rho b$ for some $0 < \rho < \min\{(\gamma - \beta)/\gamma, \frac{1}{2}\}$. In fact,

$$\text{var}(\hat{p}_1 + \hat{p}_2) = O\left(\frac{p^2}{m}\right) [= o(p^2)] \quad \text{when } \liminf\left(\frac{m}{n}\right) > 0.$$

Proof. Recall that $f(x) = \psi(x)e^{-\Psi(x)}$ is the density of X . With ζ_b defined in Theorem 2, we will show that

$$\zeta_b = \int_{-\infty}^{\rho b} e^{\theta_b x} f(x) dx \leq 1 + O(\theta_b^\beta) = 1 + O(n^{-1}), \tag{45}$$

$$\int_{-\infty}^{\rho b} e^{2\theta_b x} f(x) dx = O(1), \tag{46}$$

i.e. (C') holds. From (45), it follows that $\zeta_b^{2n} = O(1)$. Moreover, it will be shown that

$$n P(X \geq b) = O(P(S_n \geq b)). \tag{47}$$

Since $P(X > \rho b) = O(P(X > b))$ by (22), Corollary 2 follows from Theorem 2, (41), and (47).

To prove (47), note that in the case $\gamma > 2$, $\text{var}(X) < \infty$ and (47) follows from (24). For the case $\gamma \leq 2$, we use the inclusion–exclusion principle to obtain

$$\begin{aligned} P(S_n \geq b) &\geq P\left(\bigcup_{i=1}^n B_i\right) \quad (\text{where } B_i = \{X_i \geq 2b, S_n - X_i \geq -b\}) \\ &\geq n P(S_{n-1} \geq -b) P(X \geq 2b) - n^2 P^2(X \geq 2b). \end{aligned} \tag{48}$$

Note that $n P(X \geq b) \rightarrow 0$ under (22) and $n = O(b^\beta / (\log b)^\beta)$ for $0 < \beta < \gamma$. For the case $\beta \leq 1$, $P(S_{n-1} < -b) \leq E(S_{n-1}^-)^\beta / b^\beta \leq n E(X^-)^\beta / b^\beta \rightarrow 0$. For the case $1 < \beta < \gamma \leq 2$, (22) and the assumption $E(X^-)^\beta < \infty$ imply that $E|X|^\beta < \infty$. Therefore, by the Marcinkiewicz–Zygmund law of large numbers [12, p. 125], $S_n = o(n^{1/\beta})$ almost surely and, hence, $P(S_{n-1} < -b) \rightarrow 0$ as $n^{1/\beta} = o(b)$. Since $P(X \geq 2b) \sim 2^{-\gamma} P(X \geq b)$ by (22), (47) follows from (48).

We next prove (45) and (46) when $0 < \beta \leq 1$. Since $e^x \leq 1 + 2x^\beta$ for $0 < x \leq 1$ and $e^x \leq 1$ for $x \leq 0$,

$$\zeta_b \leq 1 + 2\theta_b^\beta \int_0^{1/\theta_b} x^\beta f(x) dx + \int_{1/\theta_b}^{\rho b} e^{\theta_b x} f(x) dx. \tag{49}$$

By (22), $E(X^+)^\beta < \infty$ and, therefore,

$$2\theta_b^\beta \int_0^{1/\theta_b} x^\beta f(x) dx = O(\theta_b^\beta). \tag{50}$$

Let $0 < \delta < 1$. Since $\Psi(x) \sim \gamma \log x$, $\Psi(x) \geq \gamma\delta \log x$ for large x . Moreover, $1/\theta_b \geq b^\delta$ for all large b and, therefore, by selecting $\delta \geq \sqrt{\beta/\gamma}$,

$$e^{-\Psi(2J/\theta_b)} \leq e^{-\Psi(1/\theta_b)} = O(e^{-\gamma\delta \log b^\delta}) = O(b^{-\gamma\delta^2}) = O(\theta_b^\beta). \tag{51}$$

By (44) and (51),

$$\int_{1/\theta_b}^{2J/\theta_b} e^{\theta_b x} f(x) dx \leq \left(\frac{2J}{\theta_b}\right) e^{2J} \sup_{1/\theta_b \leq x \leq 2J/\theta_b} [\psi(x)e^{-\Psi(x)}] = O(\theta_b^\beta). \tag{52}$$

Integration by parts yields

$$\int_{2J/\theta_b}^{\rho b} e^{\theta_b x} \psi(x)e^{-\Psi(x)} dx \leq e^{2J-\Psi(2J/\theta_b)} + \theta_b \int_{2J/\theta_b}^{\rho b} e^{\theta_b x - \Psi(x)} dx. \tag{53}$$

For $x \geq 2J/\theta_b$, $[\theta_b x - \Psi(x)]' = \theta_b - \psi(x) \geq \theta_b/2$ by (44) and, therefore, $\theta_b x - \Psi(x) \leq \theta_b \rho b - \Psi(\rho b) + \frac{1}{2}\theta_b(x - \rho b)$ if $x \leq \rho b$. The change of variable $y = x - \rho b$ then yields

$$\theta_b \int_{2J/\theta_b}^{\rho b} e^{\theta_b x - \Psi(x)} dx \leq \theta_b e^{\theta_b(\rho b) - \Psi(\rho b)} \int_{-\infty}^0 e^{y\theta_b/2} dy = O(e^{\rho\Psi(b) - \Psi(\rho b)}). \tag{54}$$

By (22), $\Psi(\rho b) = \Psi(b) + O(1)$ and, therefore, $\rho\Psi(b) - \Psi(\rho b) = (\rho - 1)\Psi(b) + O(1)$. Since $\rho - 1 < -\beta/\gamma$ and $\Psi(b) \sim \gamma \log b$, it follows that $e^{\rho\Psi(b) - \Psi(\rho b)} = O(b^{-\beta}) = O(\theta_b^\beta)$. Combining this with (51)–(54) yields

$$\int_{1/\theta_b}^{\rho b} e^{\theta_b x} f(x) dx = O(\theta_b^\beta). \tag{55}$$

Substituting (50) and (55) into (49) proves (45). To prove (46), we make use of the inequality

$$\int_{-\infty}^{\rho b} e^{2\theta_b x} f(x) dx \leq e^{2J} \int_{-\infty}^{J/\theta_b} f(x) dx + \int_{J/\theta_b}^{\rho b} e^{2\theta_b x} f(x) dx. \tag{56}$$

Since $[2\theta_b x - \Psi(x)]' \geq \theta_b$ for $x \geq J/\theta_b$ and $\rho \leq \frac{1}{2}$, it follows from integration by parts and the bounds in (51)–(55) that

$$\begin{aligned} \int_{J/\theta_b}^{\rho b} e^{2\theta_b x} \psi(x)e^{-\Psi(x)} dx &= -e^{2\theta_b x - \Psi(x)} \Big|_{J/\theta_b}^{\rho b} + 2\theta_b \int_{J/\theta_b}^{\rho b} e^{2\theta_b x - \Psi(x)} dx \\ &= O(\theta_b^\beta) + O(e^{(2\rho-1)\Psi(b)}) \\ &= O(1). \end{aligned} \tag{57}$$

By (56) and (57), (46) holds.

To prove (45) and (46) for the case $1 < \beta \leq 2$, $E[X] = 0$ and $E[|X|^\beta] < \infty$, we start with the bound $e^x \leq 1 + 2\beta x^{\beta-1}$ for $0 \leq x \leq 1$ from which it follows by integration that

$$e^x \leq 1 + x + 2|x|^\beta \tag{58}$$

for $0 \leq x \leq 1$. We next show that (58) in fact holds for all $x \leq 1$, by noting that the left-hand side of (58) is less than or equal to 1, whereas the right-hand side (RHS) of (58) is greater than or equal to $1 + |x|$ for $x \leq -1$, and that

$$\text{RHS of (58)} \geq 1 + x + 2x^2 \begin{cases} \geq 1 + x + x^2 + x^4 + x^6 + \dots \geq e^x & \text{for } -\frac{1}{2} \leq x \leq 0, \\ = 2(x + \frac{1}{4})^2 + \frac{7}{8} \geq e^x & \text{for } -1 \leq x \leq -\frac{1}{2}. \end{cases}$$

It follows from (58) that

$$\begin{aligned} \zeta_b &\leq 1 + \theta_b \int_{-\infty}^{1/\theta_b} x f(x) dx + 2\theta_b^\beta \int_{-\infty}^{1/\theta_b} x^\beta f(x) dx + \int_{1/\theta_b}^{\rho b} e^{\theta_b x} f(x) dx \\ &\leq 1 + O(\theta_b^\beta), \end{aligned} \tag{59}$$

since

$$\begin{aligned} E[X] = 0 &\implies \theta_b \int_{-\infty}^{1/\theta_b} x f(x) dx \leq 0, \\ E[|X|^\beta] < \infty &\implies 2\theta_b^\beta \int_{-\infty}^{1/\theta_b} x^\beta f(x) dx = O(\theta_b^\beta), \end{aligned}$$

and (52)–(54) can still be applied to show that (55) holds. Using arguments similar to (56) and (57), we can prove (46) in this case.

Similarly, we can prove the following analog of Corollary 2 for the SISR algorithm in Section 3.2 to simulate α .

Corollary 3. Assume that (22) with $\gamma > 1$ and (44) hold. Suppose that $n = O(b^\beta / (\log b)^\beta)$, $E[X] = 0$, and $E(X^-)^\beta < \infty$ for some $1 < \beta < \gamma$ with $\beta \leq 2$. Let $b(j) = b$ for all $1 \leq j \leq n$, and suppose that $2^i \leq b < 2^{i+1}$. Assign all m simulations to evaluate $P(\mathcal{A}_{1,i})$. Then $\hat{\alpha}_1 + \hat{\alpha}_2$ is linearly efficient when $c_b = \rho b$ for some $0 < \rho < \frac{1}{2} \min\{(\gamma - \beta)/\gamma, \frac{1}{2}\}$. In fact,

$$\text{var}(\hat{\alpha}_1 + \hat{\alpha}_2) = O\left(\frac{\alpha^2}{m}\right) \quad \text{when } \liminf\left(\frac{m}{n}\right) > 0. \tag{60}$$

Proof. By Theorem 3 and Lemma 4,

$$\begin{aligned} \text{var}(\hat{\alpha}_1) &= O\left(\frac{n}{m} P^2(X > 2^i)\right) = O\left(\frac{n}{m} P^2(X > b)\right) \quad \text{when } \liminf\left(\frac{m}{n}\right) > 0, \\ \text{var}(\hat{\alpha}_2) &= O\left(\frac{n^2}{m} P^2(X > b)\right). \end{aligned}$$

By (47), $n P(X > b) = O(P(S_n \geq b)) = O(\alpha)$ and, therefore, (60) holds.

Corollary 4. Assume that (22) with $\gamma > 1$ and (44) hold. Suppose that $E[X] = 0$ and $E(X^-)^\beta < \infty$ for some $1 < \beta < \gamma$. Let $b(j) = b + ja$ for some $a > 0$, and let $\ell(k) = \lfloor \log_2 b(k) \rfloor$, where \log_2 denotes the logarithm to base 2. Assign m_i simulation runs for the

estimation of $P(\mathcal{A}_{1,i})$ such that

$$m_i \sim \frac{m[i - \ell(1) + 1]^{-2}}{\sum_{\ell=1}^{\ell(n) - \ell(1) + 1} \ell^{-2}} \quad \text{uniformly over } \ell(1) \leq i \leq \ell(n) \text{ as } b \rightarrow \infty, \tag{61}$$

where $m = \sum_{i=\ell(1)}^{\ell(n)} m_i$ is the total number of simulation runs. Let $\hat{\alpha}_1 = \sum_{i=\ell(1)}^{\ell(n)} \hat{\alpha}_{1,i}$. Then the estimate $\hat{\alpha}_1 + \hat{\alpha}_2$ is $n(\log_2 n)^2$ -efficient if $c_b = \rho b$ for some $0 < \rho < \frac{1}{2} \min\{(\gamma - \beta)/\gamma, \frac{1}{2}\}$. In fact,

$$\text{var}(\hat{\alpha}_1 + \hat{\alpha}_2) = m^{-1} O(\alpha^2) [= o(\alpha^2)] \quad \text{whenever } \liminf \left(\frac{m}{n(\log_2 n)^2} \right) > 0.$$

Proof. We can proceed as the proofs of (45) and (46) to show that $\zeta_i^* \leq 1 + O(\theta_i^\beta)$ and (A') holds in Theorem 3. Noting that

$$\liminf \left(\frac{m}{n \log^2 n} \right) > 0 \implies \liminf \left[\inf_{\ell(1) \leq i \leq \ell(n)} \frac{m}{n(i - \ell(1) + 1)^2} \right] > 0,$$

we obtain, from (61),

$$\liminf \left(\frac{\inf_{\ell(1) \leq i \leq \ell(n)} m_i}{n} \right) > 0. \tag{62}$$

Since $n_{i+1} \theta_i^\beta \rightarrow 0$, it then follows from Theorem 3 and (62) that

$$\text{var}(\hat{\alpha}_{1,i}) = O \left(\frac{n_{i+1}}{m_i} P^2(X > 2^i) \right) \quad \text{uniformly over } \ell(1) \leq i \leq \ell(n),$$

and, hence, by (61),

$$\text{var}(\hat{\alpha}_1) = m^{-1} O \left(\sum_{i=\ell(1)}^{\ell(n)} [i - \ell(1) + 1]^2 n_{i+1} P^2(X > 2^i) \right). \tag{63}$$

Since $2^i \leq b(j) \leq 2^{i+1}$ for $n_i \leq j < n_{i+1}$ and $n_{i+1} - n_i \geq n_{i+1}/2$, (22) implies that, for some positive constants C_1 and C_2 ,

$$\left[\sum_{j=n_i}^{n_{i+1}-1} P(X > b(j)) \right]^2 \geq C_1 (n_{i+1} - n_i)^2 P^2(X > 2^i) \geq C_2 [i - \ell(1) + 1]^2 n_{i+1} P^2(X > 2^i).$$

Substituting this into (63) yields

$$\text{var}(\hat{\alpha}_1) = m^{-1} O \left(\left[\sum_{j=1}^n P(X > b(j)) \right]^2 \right) = m^{-1} O(\alpha^2);$$

see [15, Theorem 5.5(i)]. A similar bound can be derived for $\text{var}(\hat{\alpha}_2)$ by applying Lemma 4, completing the proof of Corollary 4.

4.2. More general heavy-tailed distributions

A distribution function F is said to be (right) heavy tailed if $\int_{-\infty}^{\infty} e^{\lambda x} F(dx) = \infty$ for all $\lambda > 0$. It is said to be *long tailed* if its support is not bounded above and, for all fixed $a > 0$, $\bar{F}(x+a)/\bar{F}(x) \rightarrow 1$, or, equivalently, $\Psi(x+a) - \Psi(x) \rightarrow 0$ as $x \rightarrow \infty$; see [15, Section 3.5]. To simulate $p = P(S_n \geq b)$, we have shown in Section 4.1 that the truncation method described in Section 3.1 is linearly efficient in the case of regularly varying tails. For other long-tailed distributions, such as the Weibull and log-normal distributions, some modification of the truncation method is needed for efficiency. It is based on representing $P(S_n \geq b)$ as a sum of four probabilities that can be evaluated by SISR or importance sampling.

Let $c_b < b$, $V_n = \#\{k: c_b < X_k \leq b\}$,

$$\begin{aligned} \mathcal{A}_1 &= \{S_n \geq b, M_n \leq c_b\}, & \mathcal{A}_2 &= \{S_n \geq b, M_n > b\}, \\ \mathcal{A}_3 &= \{S_n \geq b, V_n = 1, M_n \leq b\}, & \mathcal{A}_4 &= \{S_n \geq b, V_n \geq 2, M_n \leq b\}. \end{aligned}$$

The Monte Carlo estimate \hat{p}_1 of $P(\mathcal{A}_1)$ is described in Section 3.1, using SISR with mixture density (26) and resampling weights (27). The Monte Carlo estimate \hat{p}_2 of $P(\mathcal{A}_2)$ uses the importance sampling scheme described in Section 3.1, with b taking the place of c_b . To simulate $P(\mathcal{A}_3)$, we retain the simulation results $\{X_{n-1}^{(j)}: 1 \leq j \leq m\}$ in \hat{p}_1 after the $(n-1)$ th resampling step. The corresponding SISR estimate is

$$\begin{aligned} \hat{p}_3 &= (\bar{w}_1 \cdots \bar{w}_{n-1})m^{-1} \\ &\times \sum_{j=1}^m ne^{-\theta_b S_{n-1}^{(j)}} [\min(e^{-\Psi(b-S_{n-1}^{(j)})}, e^{-\Psi(c_b)}) - e^{-\Psi(b)}] \mathbf{1}_{\{S_{n-1}^{(j)} \geq 0, M_{n-1}^{(j)} \leq c_b\}}. \end{aligned}$$

To evaluate $P(\mathcal{A}_4)$, we perform m simulations such that, for the j th simulation run, $k_1^{(j)}$ and $k_2^{(j)}$ are selected at random without replacement from $\{1, \dots, n\}$, and $X_k^{(j)} \sim F(\cdot | c_b < X \leq b)$ for $k = k_1^{(j)}, k_2^{(j)}$, while $X_k^{(j)} \sim F$ for $k \neq k_1^{(j)}, k_2^{(j)}$. The Monte Carlo estimate of $P(\mathcal{A}_4)$ is

$$\hat{p}_4 = [\bar{F}(c_b) - \bar{F}(b)]^2 \binom{n}{2} m^{-1} \sum_{j=1}^m \binom{V_n^{(j)}}{2}^{-1} \mathbf{1}_{\{S_n^{(j)} \geq b, M_n^{(j)} \leq b\}}.$$

Theorem 4. *The Monte Carlo estimate \hat{p}_i of $P(\mathcal{A}_i)$ is unbiased for $i = 1, 2, 3, 4$. Let $\lambda_b = \min_{0 \leq x \leq b-c_b} [\theta_b x + \Psi(b-x)]$. Assume that either (C) or (C') holds. Then there exists $K > 0$ such that*

$$\begin{aligned} \text{mvar}(\hat{p}_1) &\leq Kn\zeta_b^{2n} e^{Kn/m} P^2(X > b), & \text{mvar}(\hat{p}_2) &\leq n^2 \bar{F}^2(b), \\ \text{mvar}(\hat{p}_3) &\leq K(n-1)\zeta_b^{2n-2} e^{K(n-1)/m} n^2 e^{-2\lambda_b}, & \text{mvar}(\hat{p}_4) &\leq n^4 \bar{F}^4(c_b). \end{aligned} \tag{64}$$

Proof. As noted above, \hat{p}_1 is an SISR estimate of $P(\mathcal{A}_1)$ and \hat{p}_2 is an importance sampling estimate of $P(\mathcal{A}_2)$. By exchangeability,

$$P(\mathcal{A}_3) = n P(S_n \geq b, M_{n-1} \leq c_b, c_b < X_n \leq b). \tag{65}$$

Let $\mathcal{A} = \{S_n \geq b, c_b < X_n \leq b\}$. In view of (65), $P(\mathcal{A}_3)$ can be evaluated by Monte Carlo using the SISR estimate

$$n\bar{w}_1 \cdots \bar{w}_{n-1} m^{-1} \sum_{j=1}^n e^{-\theta_b S_{n-1}^{(j)}} P(\mathcal{A} | S_{n-1}^{(j)}) \mathbf{1}_{\{M_{n-1}^{(j)} \leq c_b\}}. \tag{66}$$

Note that $P(\mathcal{A} \mid S_{n-1}^{(j)}) = 0$ if $S_{n-1}^{(j)} < 0$. For $s > 0$,

$$P(\mathcal{A} \mid S_{n-1}^{(j)} = s) = \begin{cases} P(b - s \leq X_n \leq b) = e^{-\Psi(b-s)} - e^{-\Psi(b)} & \text{if } b - s > c_b, \\ P(c_b < X_n \leq b) = e^{-\Psi(c_b)} - e^{-\Psi(b)} & \text{if } b - s \leq c_b. \end{cases}$$

Hence, \hat{p}_3 is the same as the SISR estimate (66) of $P(\mathcal{A}_3)$ and is therefore unbiased. The estimate \hat{p}_4 is also unbiased. In fact, it is an importance sampling estimate that draws X_n from a measure \mathcal{Q} for which

$$\frac{d\mathcal{Q}}{dP}(X_n) = \binom{V_n}{2} / \left\{ \binom{n}{2} P^2(c_b < X \leq b) \right\} \quad \text{on } \{V_n \geq 2, M_n \leq b\},$$

which is an extension of (39) to the present problem.

We next prove the variance bounds (64) for the unbiased estimates \hat{p}_3 and \hat{p}_4 ; those for \hat{p}_1 and \hat{p}_2 have already been shown in Section 3.1. Consider the martingale decomposition $m[\hat{p}_3 - P(\mathcal{A}_3)] = \sum_{t=1}^{2(n-1)} \xi_t$, where ξ_t is given in the display after (30) with

$$Z_k(\mathbf{x}_k) = \left[\prod_{t=1}^k \frac{f(x_t)}{q(x_t)} \right] n P(S_n \geq b, M_{n-1} \leq c_b, c_b < X_n \leq b \mid \mathbf{X}_k = \mathbf{x}_k) \quad (67)$$

in view of (65), noting that \hat{p}_3 is based on the simulations used in \hat{p}_1 up to the $(n - 1)$ th resampling step. The change-of-measure argument used to prove (31) can be modified to show that, for all $t \geq 1$ and $x \in \mathbb{R}$,

$$P(S_t \geq x, M_{t-1} \leq c_b, c_b < X_t \leq b) \leq \zeta_b^{t-1} e^{\theta_b(b-x)} \max_{0 \leq y \leq b-c_b} e^{-\theta_b y - \Psi(b-y)}. \quad (68)$$

Making use of (67) and (68), we can proceed as in the proof of Theorem 2 to prove the upper bound for $\text{var}(\hat{p}_3)$ in (64). The bound for $\text{var}(\hat{p}_4)$ follows from $(\bar{F}(c_b) - \bar{F}(b))^2 \binom{n}{2} \leq n^2 \bar{F}^2(c_b)$, thus completing the proof of Theorem 4.

The following corollary of Theorem 4 establishes linear efficiency of the Monte Carlo method to evaluate $P(S_n \geq b)$ for heavy-tailed distributions satisfying certain assumptions. Examples 1 and 2 in Section 5.1 show that these assumptions are satisfied in particular by Weibull and log-normal X .

Corollary 5. *Let X be heavy tailed with $E[X] = 0$ and $\text{var}(X) < \infty$, and let $n = O(b^2/\Psi^2(b))$. Assume that either (C) or (C') holds. If $\theta_b \geq \psi(x)$ for all $c_b \leq x \leq b$, $\zeta_b = 1 + O(\theta_b^2)$, and*

$$b e^{-2\Psi(c_b)} = O(e^{-\Psi(b)}), \quad (69)$$

then $\sum_{i=1}^4 \hat{p}_i$ is linearly efficient for estimating $P(S_n \geq b)$.

Proof. Since $n = o(b^2)$, $n P(X > b) \rightarrow 0$ by Chebyshev's inequality. Therefore, it follows from the inclusion-exclusion principle and the central limit theorem that

$$P(S_n \geq b) \geq n P(S_{n-1} \geq 0) P(X > b) - n^2 P^2(X > b) \geq \frac{1}{2} [1 + o(1)] n P(X > b).$$

Hence, it suffices to show that, for any $\varepsilon > 0$, there exists $m = O(n)$ such that

$$\text{var}(\hat{p}_i) \leq \varepsilon n^2 \bar{F}^2(b) \quad \text{for } 1 \leq i \leq 4. \quad (70)$$

We will assume that $\liminf m/n > 0$. Since $n P(X > b) \rightarrow 0$, (70) holds for $i = 2$. Since $n\theta_b^2 = O(1)$ and $\zeta_b = 1 + O(\theta_b^2)$, $\zeta_b^{2n} = O(1)$ and (70) holds for $i = 1$. Since

$$[\theta_b x + \Psi(b - x)]' = \theta_b - \psi(b - x) \geq 0 \quad \text{for all } 0 \leq x \leq b - c_b,$$

the minimum of $\theta_b x + \Psi(b - x)$ over $0 \leq x \leq c_b$ is attained at $x = 0$ and, therefore, $\lambda_b = \Psi(b)$, proving (70) for $i = 3$. Finally, by (69), $n^3 \bar{F}^4(c_b) = O(n^3 e^{-2\Psi(b)}/b^2) = O(n^2 e^{-2\Psi(b)})$, proving (70) for $i = 4$.

5. Examples and discussion

In this concluding section we first give examples of heavy-tailed distributions satisfying the assumptions of Corollary 5. We also give numerical examples to illustrate the performance of the proposed Monte Carlo methods. In this connection we describe in Section 5.2 some implementation details such as the use of occasional resampling to speed up the SISR procedure and the estimation of standard errors for the SISR estimates of rare-event probabilities. Finally, we discuss in Section 5.4 related works in the literature and compare our approach with importance sampling and IPS (interacting particle system) methods.

5.1. Weibull and log-normal increments

Example 1. (Weibull.) A long-tailed distribution is Weibull if $\Psi(x) = x^\gamma \mathbf{1}_{\{x>0\}}$ for some $0 < \gamma < 1$. Let $Y \sim F$, where $\bar{F}(x) = e^{-\Psi(x)}$, and let $X = Y - EY$. Then

$$P(X > x) = e^{-\Psi(x+\mu)} = \exp(-(x + \mu)^\gamma) \quad \text{for } x + \mu > 0,$$

where $\mu = EY$. Moreover, for $x > -\mu$, $\Psi'(x + \mu) = \gamma(x + \mu)^{\gamma-1}$. Therefore, $\theta_b = \Psi(b + \mu)/b \sim b^{\gamma-1}$ and $\Psi'(x + \mu) \leq \theta_b$ for all $b/2 \leq x \leq b$ when b is sufficiently large, noting that $\gamma 2^{1-\gamma} < 1$ for $0 < \gamma < 1$. Let $c_b = b/2$ and $n = O(b^{2(1-\gamma)})$. It is easy to check that (69) holds. By (58) with $\beta = 2$ and (59),

$$\zeta_b \leq 1 + O(\theta_b^2) + \int_{1/\theta_b}^{b/2} f(x)e^{\theta_b x} dx \leq 1 + O(\theta_b^2),$$

where $f(x) = \gamma(x + \mu)^{\gamma-1} \exp(-(x + \mu)^\gamma)$. Moreover, applying (58) with $\beta = 2$ to the range $2\theta_b x \leq 1$ and using the bound $f(x) \leq 1$ for $x \geq 1$,

$$\begin{aligned} \int_1^{b/2} x^2 e^{2\theta_b x} f^2(x) dx &\leq e \int_1^{1/2\theta_b} x^2 f(x) dx \\ &\quad + \left(\frac{b}{2}\right)^3 \max_{1/2\theta_b \leq x \leq b/2} \exp(2[\theta_b x - (x + \mu)^\gamma]), \end{aligned} \tag{71}$$

in which the last term is an upper bound of $\int_{1/2\theta_b}^{b/2} x^2 e^{2\theta_b x} f^2(x) dx$, noting that $f(x) \leq \exp(-(x + \mu)^\gamma)$ for $x \geq 1$. The maximum of the convex function $\theta_b x - (x + \mu)^\gamma$ over $1/2\theta_b \leq x \leq b/2$ is attained at $1/2\theta_b$ and is equal to $-(1/2)^\gamma + o(1)$ for all large b . Since $E[X^2] < \infty$, (71) implies that (C) holds. Hence, all the conditions of Corollary 5 hold in this case.

Example 2. (Log normal.) Let ϕ and Φ be the standard normal density and distribution functions, respectively. Let $X = e^Z$, where Z is standard normal. Then X is log normal

TABLE 1: Monte Carlo estimates of $P(S_n \geq (5 + \mu)n)$ for log-normal increments, with estimated standard errors (after the \pm sign).

n	Direct method	Truncation method
10	$(4.80 \pm 0.69) \times 10^{-4}$	$(5.02 \pm 0.04) \times 10^{-4}$
50	0 ± 0	$(8.78 \pm 0.06) \times 10^{-7}$
100	0 ± 0	$(2.61 \pm 0.02) \times 10^{-8}$
500	0 ± 0	$(1.27 \pm 0.01) \times 10^{-12}$
1000	0 ± 0	$(8.61 \pm 0.07) \times 10^{-15}$

and has distribution function $F(x) = 1 - e^{-\Psi(x)}$, where $\Psi(x) = |\log \bar{\Phi}(\log x)|\mathbf{1}_{\{x>0\}}$. Since $\bar{\Phi}(z) \sim (2\pi z^2)^{-1/2}e^{-z^2/2}$ as $z \rightarrow \infty$, it follows that

$$\Psi(x) = \frac{(\log x)^2}{2} + \log \log x + \frac{\log(2\pi)}{2} + o(1) \quad \text{as } x \rightarrow \infty,$$

$$f(x) (= \psi(x)e^{-\Psi(x)}) = \frac{\phi(\log x)}{x} \implies \psi(x) \sim \frac{\log x}{x} \quad \text{as } x \rightarrow \infty.$$

Let $\mu = E[X] = E[e^Z] = \sqrt{e}$ and $p = P(S_n \geq b + n\mu)$, where $n = O(b^2/\Psi^2(b))$. Let $c_b = b/2$. By using arguments similar to those in Example 1, it can be shown that all the assumptions of Corollary 5 again hold in this case.

To illustrate the performance of the truncation method in Section 4.2 to estimate $p = P(S_n \geq (5 + \mu)n)$, which is shown to be linearly efficient in Corollary 5, we consider $n = 10, 50, 100, 500$, and 1000 and use the procedure described in the next subsection to implement the SISR estimates \hat{p}_1 and \hat{p}_3 with 10 000 sample paths and the importance density (26) in which $r = 0.8$. Recall that \hat{p}_3 uses the SISR sample paths for \hat{p}_1 up to the $(n - 1)$ th resampling step. The importance sampling estimates \hat{p}_2 and \hat{p}_4 are each based on 100 000 simulations. For comparison, we also apply direct Monte Carlo with 100 000 runs to evaluate the probability. The results are given in Table 1, which shows about 300-fold variance reduction for a probability of order 10^{-4} . For probabilities of order 10^{-7} or smaller, the results given in Table 1 show that direct Monte Carlo is not feasible whereas the truncation method does not seem to deteriorate in performance.

5.2. Standard errors and occasional resampling

The SISR procedure carries out importance sampling sequentially within each simulated trajectory and performs resampling across the m trajectories. Instead of implementing this procedure directly, we use the modification in [9, Section 3.3] to reduce computation time for resampling, which increases with m , and also to obtain standard error estimates easily. Dividing the m sample paths into r subgroups of size ν so that $m = r\nu$, we perform resampling within each subgroup of ν sample paths, independently of the other subgroups. This method also has the advantage of providing a direct estimate of the standard error of the Monte Carlo estimate $\bar{\alpha} := r^{-1} \sum_{i=1}^r \hat{\alpha}(i)$, where $\hat{\alpha}(i)$ denotes the SISR estimate of the rare-event probability α based on the i th subgroup of simulated sample paths. Due to resampling, the SISR samples are no longer independent and one cannot use the conventional estimate of the standard error for Monte Carlo estimates. On the other hand, since the r subgroups are independent and yield the independent estimates $\hat{\alpha}(1), \dots, \hat{\alpha}(r)$ of α , we can estimate the standard error of $\bar{\alpha}$ to be $\hat{\sigma}/\sqrt{r}$, where $\hat{\sigma}^2 = (r - 1)^{-1} \sum_{i=1}^r (\hat{\alpha}(i) - \bar{\alpha})^2$. In Example 2 above and Example 3 below, we use $\nu = r = 100$, corresponding to a total of $m = 10\,000$ SISR sample paths.

An additional modification that can be used to further reduce the resampling task is to carry out resampling at stage k only when the coefficient of variation (CV) of the resampling weights $w_k^{(j)}$ exceeds some threshold. As pointed out by Kong *et al.* [17], the purpose of resampling is to help prevent the weights $w_k^{(j)}$ from becoming heavily skewed (e.g. nearly degenerate) and the effective sample size for ν sequentially generated sample paths is $\nu/(1+CV^2)$. Therefore, Kong *et al.* [17] recommended resampling when CV exceeds a threshold. Choosing the threshold to be 0 is tantamount to resampling at every step, and a good choice in many applications is in the range from 1 to 2.

5.3. Positive increments with regularly varying tails

Example 3. Let $X = \Lambda Y$, where $P(Y > x) = \min(x^{-4}, 1)$ and $\Lambda \sim \text{Laplace}(1)$ is independent of Y . Blanchet and Liu [7] in their Example 1 showed that X has tail probability

$$\bar{F}(x) = 2x^{-4}[6 - e^{-x}(6 + 6x + 3x^2 + x^3)].$$

Let X, X_1, \dots, X_n be i.i.d., and let $S_n = X_1 + \dots + X_n$. In [7], $P(S_n \geq n)$ is simulated for $n = 100, 500, \text{ and } 1000$ by using

- (I) state-dependent importance sampling (IS) that approximates the h -transform,
- (II) time-varying mixtures for IS introduced by Dupuis *et al.* [14].

We compare their results in [7], each of which is based on 10 000 simulations, with those of 10 000 SISR sample paths generated by the following methods:

- (III) SISR using (25) with

$$\zeta_i = \begin{cases} -\infty & \text{for } i = -1, \\ \frac{bi}{180} & \text{for } 0 \leq i \leq 90, \\ \frac{b}{2} + \frac{b(i - 90)}{20} & \text{for } 91 \leq i \leq 100, \\ \infty & \text{for } i = 101, \end{cases} \tag{72}$$

and resampling conducted at every step,

- (IV) SISR using (25) and (72) with resampling only when CV exceeds 2,
- (V) the truncation method in Section 3.1 with $c_b = 2b/5$, importance density (26) in which $r = 0.9$, and resampling weights (27) in which $\theta_b = 4b^{-1} \log b$.

For (V), besides using 10 000 SISR sample paths to estimate $P(S_n \geq n, M_n \leq 2n/5)$, we also use 10 000 IS simulations to estimate $P(S_n \geq n, M_n > 2n/5)$. As shown in Table 2, the standard errors of (I) and (III)–(V) are comparable and are all smaller than that of (II) when $n = 500$ and 1000, whereas, for $n = 100$, the standard errors of (III)–(V) are substantially smaller than those of (I) and (II). Although Blanchet and Liu [7, Theorem 4] showed (II) to be strongly efficient, their parametric mixtures are based on a single large jump since the effect of two or more large jumps is asymptotically negligible when the tail probability is of the order 10^{-7} or smaller. For larger tail probabilities, the effect of two or more jumps may be significant, and the results given in Table 2 show that (V) can provide substantial improvement by taking this effect into consideration.

TABLE 2: Monte Carlo estimates of $P(S_n \geq n) \pm$ standard errors.

Method	n		
	100	500	1000
I	$(2.37 \pm 0.23) \times 10^{-5}$	$(1.02 \pm 0.01) \times 10^{-7}$	$(1.23 \pm 0.01) \times 10^{-8}$
II	$(2.09 \pm 0.10) \times 10^{-5}$	$(1.11 \pm 0.04) \times 10^{-7}$	$(1.16 \pm 0.05) \times 10^{-8}$
III	$(2.21 \pm 0.06) \times 10^{-5}$	$(1.04 \pm 0.01) \times 10^{-7}$	$(1.25 \pm 0.01) \times 10^{-8}$
IV	$(2.26 \pm 0.03) \times 10^{-5}$	$(1.05 \pm 0.01) \times 10^{-7}$	$(1.24 \pm 0.01) \times 10^{-8}$
V	$(2.16 \pm 0.03) \times 10^{-5}$	$(1.05 \pm 0.02) \times 10^{-7}$	$(1.24 \pm 0.02) \times 10^{-8}$

5.4. Other methods, related works, and discussion

Asmussen *et al.* [3] introduced several methods for importance sampling of tail probabilities of sums of heavy-tailed random variables and showed that these importance sampling methods are strongly efficient for fixed n as $b \rightarrow \infty$. One such method involves simulating i.i.d. X_1, \dots, X_n from a distribution H that has a heavier tail than F . This method cannot be extended to the case $n \rightarrow \infty$ because the likelihood ratio statistic has exponentially increasing variance with n . Noting that

$$P(S_n \geq b) = n E[P(S_n \geq b, X_n \geq \max(X_1, \dots, X_{n-1}) \mid X_1, \dots, X_{n-1})],$$

Asmussen and Kroese [2] introduced the conditional Monte Carlo method that estimates $P(S_n \geq b)$ by the average of m independent realizations of

$$\bar{F}(\max\{b - (X_1 + \dots + X_{n-1}), X_1, \dots, X_{n-1}\}),$$

and showed that it is strongly efficient for fixed n as $b \rightarrow \infty$, when F is regularly varying. This approach, however, breaks down if n also approaches ∞ .

Blanchet *et al.* [8] also introduced a truncation method to simulate tail probabilities of a random walk S_n with log-normal increments, and showed that it is strongly efficient as $b \rightarrow \infty$ for fixed n . Their truncation method used $c_b = b$ and importance sampling to estimate $P(S_n \geq b, M_n \leq b)$, and their argument heavily depended on fixed n . By using SISR instead, we can control the variances of the likelihood ratio statistics associated with sequential importance sampling and of the resampling steps, as shown in Theorems 2 and 4 and Corollaries 3 and 5.

The truncation scheme in Sections 3 and 4 can be regarded as a Monte Carlo implementation of a similar truncation method for the analysis of tail probabilities of random walks whose i.i.d. increments have mean 0 and finite variance. Chow and Lai [10], [11] used the truncation method to prove that, for $\alpha > \frac{1}{2}$ and $p > 1/\alpha$,

$$\sum_{n=1}^{\infty} n^{p\alpha-2} P\left(\max_{1 \leq k \leq n} S_k \geq n^\alpha\right) \leq C_{p,\alpha} \{E(X^+)^p + (E X^2)^{(p\alpha-1)/(2\alpha-1)}\}, \tag{73}$$

where $C_{p,\alpha}$ is a universal constant depending only on p and α . This inequality is sharp in the sense that there is a corresponding lower bound for the two-sided tail probability in the case $p \geq 2$:

$$\begin{aligned} \sum_{n=1}^{\infty} n^{p\alpha-2} P\left(\max_{1 \leq k \leq n} |S_k| \geq n^\alpha\right) &\geq \sum_{n=1}^{\infty} n^{p\alpha-2} P(|S_n| \geq n^\alpha) \\ &\geq B_{p,\alpha} \{E|X|^p + (E X^2)^{(p\alpha-1)/(2\alpha-1)}\}. \end{aligned}$$

The proof of (73) makes use of the bound

$$\mathbb{P}\left(\max_{1 \leq k \leq n} S_k \geq n^\alpha\right) \leq \mathbb{P}(M_n > \varepsilon n^\alpha) + \mathbb{P}\left(\max_{1 \leq k \leq n} S_k \geq n^\alpha, M_n \leq \varepsilon n^\alpha\right),$$

with $\varepsilon = 1/2\nu$ for some positive integer ν . In fact, the term $\mathbb{E}(X^+)^p$ in (73) comes from the bound

$$\sum_{n=1}^{\infty} n^{p\alpha-2} \mathbb{P}(M_n > \varepsilon n^\alpha) \leq \sum_{n=1}^{\infty} n^{p\alpha-1} \mathbb{P}(X > \varepsilon n^\alpha) \leq A_{p,\alpha} \mathbb{E}(X^+)^p,$$

and is associated with the ‘large jump’ probability of an increment for heavy-tailed random walks. In this connection, note that $b = n^\alpha$ satisfies the assumption that $n = O(b^2/\Psi^2(b))$ in Corollary 2 and Examples 1 and 2 when $\alpha > \frac{1}{2}$ and $\mathbb{E} X^2 < \infty$.

Although we have focused on one-dimensional random walks, the SISR procedures can be readily extended to the multivariate setting in which the X_i are i.i.d. d -dimensional random vectors such that $\|X\|$ is heavy tailed, satisfying $\mathbb{P}(\|X\| > x) = e^{-\Psi(x)}$ such that $\psi(x) = \Psi'(x) \rightarrow 0$. Here $p = \mathbb{P}(g(S_n/n) \geq b)$ and $\alpha = \mathbb{P}(\max_{n_1 \leq j \leq n} jg(S_j/j) \geq b_n)$, as considered in [9] for the light-tailed case. Another extension, also considered in [9] for the light-tailed case, is to heavy-tailed Markov random walks for which $\Psi(x)$ above is replaced by $\Psi_u(x)$, where u is a generic state of the underlying Markov chain.

Approximating the h -transform closely is crucial for the sequential (state-dependent) importance sampling methods of Blanchet and Glynn [4] and Blanchet and Liu [6], [7] to be strongly efficient. This requires sharp and easily computable analytic approximations of α and p , provided by the Pakes–Veraverbeke theorem [1, p. 296] in [4] and provided by Rozovskiĭ’s theorem [18] in [7]. In addition, an elaborate acceptance–rejection scheme is needed to sample from the state-dependent importance measure at every stage. If less accurate approximations to the h -transform are used, e.g. using (7) instead of (6) because either (6) is not available or because the g_n in (7) is much simpler to compute, then the likelihood ratios associated with the corresponding sequential importance sampling scheme would eventually have very large variances that approach ∞ as $n \rightarrow \infty$. This was first pointed out by Kong *et al.* [17], who proposed using resampling to address this difficulty. While these SISR schemes, also called *particle filters* or *interacting particle systems*, were used primarily for filtering in nonlinear state-space models and more general hidden Markov models, Del Moral and Garnier [13] recognized that they could be used to simulate probabilities of rare events of the form $\{V(U_n) \geq a\}$ for a possibly nonhomogeneous Markov chain U_n , with large a but fixed n . Chan and Lai [9] recently developed a comprehensive theory of SISR for simulating large deviation probabilities of $g(S_n/n)$ for large n in the case of light-tailed multivariate random walks. This paper continues the development for the heavy-tailed case, which provides new insights into the SISR approach to rare-event simulation.

References

- [1] ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd edn. Springer, New York.
- [2] ASMUSSEN, S. AND KROESE, D. P. (2006). Improved algorithms for rare event simulation with heavy tails. *Adv. Appl. Prob.* **38**, 545–558.
- [3] ASMUSSEN, S., BINSWANGER, K. AND HOJGAARD, B. (2000). Rare events simulation for heavy-tailed distributions. *Bernoulli* **6**, 303–322.
- [4] BLANCHET, J. AND GLYNN, P. (2008). Efficient rare-event simulation for the maximum of heavy-tailed random walks. *Ann. Appl. Prob.* **18**, 1351–1378.
- [5] BLANCHET, J. AND LAM, H. (2012). State-dependent importance sampling for rare-event simulation: an overview and recent advances. *Surveys Operat. Res. Manag. Sci.* **17**, 38–59.

- [6] BLANCHET, J. AND LIU, J. (2006). Efficient simulation for large deviation probabilities of sums of heavy-tailed increments. In *Proc. 2006 Winter Simul. Conf.*, IEEE, pp. 757–764.
- [7] BLANCHET, J. H. AND LIU, J. (2008). State-dependent importance sampling for regularly varying random walks. *Adv. Appl. Prob.* **40**, 1104–1128.
- [8] BLANCHET, J., JUNEJA, S. AND ROJAS-NANDAYAPA, L. (2008). Efficient tail estimation for sums of correlated lognormals. In *Proc. 2008 Winter Simul. Conf.*, IEEE, pp. 607–614.
- [9] CHAN, H. P. AND LAI, T. L. (2011). A sequential Monte Carlo approach to computing tail probabilities in stochastic models. *Ann. Appl. Prob.* **21**, 2315–2342.
- [10] CHOW, Y. S. AND LAI, T. L. (1975). Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossings. *Trans. Amer. Math. Soc.* **208**, 51–72.
- [11] CHOW, Y. S. AND LAI, T. L. (1978). Paley-type inequalities and convergence rates related to the law of large numbers and extended renewal theory. *Z. Wahrscheinlichkeitsth.* **45**, 1–19.
- [12] CHOW, Y. S. AND TEICHER, H. (1988). *Probability Theory: Independence, Interchangeability, Martingales*, 2nd edn. Springer, New York.
- [13] DEL MORAL, P. AND GARNIER, J. (2005). Genealogical particle analysis of rare events. *Ann. Appl. Prob.* **15**, 2496–2534.
- [14] DUPUIS, P., LEDER, K. AND WANG, H. (2007). Importance sampling for sums of random variables with regularly varying tails. *ACM Trans. Model. Comput. Simul.* **17**, 21pp.
- [15] FOSS, S., KORSHUNOV, D. AND ZACHARY, S. (2009). *An Introduction to Heavy-tailed and Subexponential Distributions*. Springer, New York.
- [16] JUNEJA, S. (2007). Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error. *Queueing Systems* **57**, 115–127.
- [17] KONG, A., LIU, J. S. AND WONG, W. H. (1994). Sequential imputations and Bayesian missing data problems. *J. Amer. Statist. Assoc.* **89**, 278–288.
- [18] ROZOVSKIĬ, L. V. (1989). Probabilities of large deviations of sums of independent random variables with common distribution function in the domain of attraction of the normal law. *Theory Prob. Appl.* **34**, 625–644.