State-independent Importance Sampling for Estimating Large Deviation Probabilities in Heavy-tailed Random Walks

Karthikey A. M.
Tata Institute of Fundamental Research
Mumbai
Email: kamurthy@tifr.res.in

Sandeep Juneja
Tata Institute of Fundamental Research
Mumbai
Email: juneja@tifr.res.in

Abstract—Efficient simulation of rare events involving sums of heavy-tailed random variables has been an active research area in applied probability over the last fifteen years. These rare events arise in many applications including telecommunications, computer and communication networks, insurance and finance. These problems are viewed as challenging, since large deviations theory inspired and exponential twisting based importance sampling methods that work well for rare events involving sums of light tailed random variables fail in these settings. Moreover, there exist negative results suggesting that state-independent importance sampling methods that work well in light-tailed settings fail for certain rare events involving sums of heavy-tailed random variables. This has led to the development of growing literature for efficiently simulating such events using more nuanced, and in many cases, computationally demanding state-dependent importance sampling methods. In this article we shed new light on this issue by observing that simpler state-independent exponential twisting based importance sampling methods, suitably adjusted in the tails, can provide strongly efficient algorithms to estimate such rare event probabilities. Specifically, we develop strongly efficient state-independent importance sampling algorithms for the classical large deviations probability that sums of independent, identically distributed random variables with regularly varying tails (roughly speaking, this means that the right tails decay as a power law; see II-A for a precise definition). For an introduction to the use of importance sampling in rare event simulation see, e.g., [1] Chapters V and VI, and [2].

I. INTRODUCTION

In this article we consider the problem of efficient simulation via importance sampling of large deviation probabilities \( P\{S_n > b\} \) for \( b \geq c n^{2+\epsilon} \) given \( c \) and \( \epsilon > 0 \) as \( n \uparrow \infty \). Here \( S_n = \sum_{i=1}^n X_i \) is the sum of independent identically distributed (i.i.d) zero mean random variables with regularly varying right tails (roughly speaking, this means that the right tails decay as a power law; see II-A for a precise definition). For an introduction to the use of importance sampling in rare event simulation see, e.g., [1] Chapters V and VI, and [2]. Over the last few years, elegant state-dependent importance sampling algorithms have been proposed in [3], [4] and [5] to efficiently simulate these large-deviation probabilities; state-dependence essentially means that the importance sampling distribution for generating \( X_i \) depends on the realized values of \( X_1, \ldots, X_{i-1} \) (typically through \( S_{i-1} \)); state-independence on the other hand implies that such dependence does not exist and samples of \( X_1, \ldots, X_n \) can be drawn independently.

State-independent methods enjoy obvious advantages over state-dependent ones in terms of complexity of generating samples and ease of implementation. When the running time requirement is stringent, samples can even be generated offline before hand, which may be difficult for state-dependent methods. Our key contribution is that we show that proposed simpler state-independent importance sampling algorithms can also provide strongly efficient algorithms for computing such large deviation probabilities involving random variables (rv) with regularly varying tails. The general folklore in the rare event simulation community is that exponentially twisted distributions may be used to estimate such probabilities when the right tails of the constituent \( X_i \)'s are light*, but these methods fail when the underlying tails are instead heavy. Our another contribution is that we show that by suitably adjusting along the right tail, state-independent exponential twisting can be used to efficiently estimate \( P\{S_n > b\} \) when \( X_i \)’s have regularly varying right tails.

The problem of efficient estimation of \( P\{S_n > b\} \) as \( n \uparrow \infty \) when \( X_i \)'s are light-tailed as well as heavy-tailed has received considerable attention in the literature (see [6], [7], [8], [9], and [10] for literature on light-tailed random walks, [4] and [5] for the right heavy-tailed ones), mainly because this forms a building block to many more complex rare event problems involving combination of renewal processes. For examples in queueing, see [11], and in financial credit risk modeling, see [12] and [13].

This problem was first addressed in [6] where they considered \( X_i \)'s that were light-tailed. They used an importance sampling measure obtained by exponentially twisting the original distribution of each \( X_i \) appropriately to arrive at a weakly efficient algorithm for estimating \( P(S_n > na) \) for \( a > 0 \) (the notions of weakly efficient and strongly efficient estimation methods for rare event probabilities are standard; these are

* A rv \( X \) is said to have a heavy right (resp., left) tail if \( \mathbb{E}[e^{\theta X}] = \infty \) for all positive (negative) values of \( \theta \). Otherwise it is said to have a light right (resp., left) tail. For presentation convenience, in this paper unless otherwise specified, by light-tailed rv we mean right light-tailed rv, and by heavy-tailed rv we mean right heavy-tailed rv.
reviewed later in Section II-B). This methodology does not carry over to heavy-tailed random variables as arriving at appropriate exponentially twisted distributions assumes the existence of the moment generating function of the original random variable in a right neighborhood of zero. Reference [14] provides an account of why large deviations based methods for approximating zero-variance measure may fail in a heavy-tailed setting.

In the heavy-tailed setting, while state-independent importance sampling distributions have been proposed for estimating probabilities such as \( P(S_n > b) \) in the asymptotic regime where \( n \) is fixed and \( b \rightarrow \infty \) (see [15], [14], [16], [17], [18] and [19]), for problems that involve large number of heavy-tailed random variables, increasingly researchers resort to using state-dependent importance sampling (see, e.g., [18], [20], [4] and [5]). This approach is also supported by [21], where for certain rare probabilities involving heavy-tailed random walks, it is shown that state-independent importance sampling algorithms cannot be weakly efficient. We refer the reader to [22] for an excellent survey of state-dependent importance sampling methods.

As mentioned earlier, our key aim is to show that state-independent importance sampling algorithm based on the proposed ‘tail-adjusted’ exponential twisting provides strongly efficient algorithms for estimating \( P(S_n > b) \) when component \( X_i \)'s have regularly varying right tails, for \( b \geq \tilde{c} n^{\frac{1}{\alpha} + \epsilon} \) as \( n \rightarrow \infty \). We believe that this approach has much wider applicability. To illustrate this, we further show that similar importance sampling distribution can also be used to develop strongly efficient algorithms for estimating \( P(S_n > b) \) when \( n \) is fixed and \( b \) increases to infinity. Note that algorithms proposed in [15], [14] and [16] for this problem were weakly efficient. [17] and [19] exploit alternative representations for \( P(S_n > b) \) to arrive at faster algorithms. Our approach can be potentially adapted to those representations to provide further performance improvement.

The organization of the rest of the paper is as follows: In Section II we discuss the preliminary concepts relevant to the problems addressed. We introduce our state-independent importance sampling algorithms for estimating \( P(S_n > b) \), \( b \geq \tilde{c} n^{\frac{1}{\alpha} + \epsilon} \) as \( n \rightarrow \infty \) in Section III and for \( P(S_n > b) \) as \( b \rightarrow \infty \) with fixed \( n \) in Section IV. We discuss a numerical simulation example in Section V, where we compare the performance of the proposed algorithm, which runs in \( O(n) \) time with that of the state-dependent importance sampling algorithm in [3], which runs in \( O(n^2) \) time; there we note that the observed values of relative error are comparable for large values of \( n \). We end with a brief conclusion in Section VI. Some of the more technical proofs are presented in the appendix.

II. PRELIMINARIES

In this section we first specify the exceedance probabilities considered in this paper for efficient simulation. We then briefly review rare event simulation and the use of importance sampling for estimating rare event probabilities. We also review the relevant efficiency notions that asymptotically quantify the performance improvement offered by successful importance sampling algorithms. We then review the asymptotics in the existing literature for the tail probabilities that we need in proving efficiency results for the proposed algorithms. Finally we review some of the importance sampling techniques proposed in the literature for efficient simulation of large deviation probabilities associated with random walks involving light as well as regularly varying heavy-tailed random variables.

A. Problem setup

Let \( \{X_n : n \geq 1\} \) denote a collection of i.i.d. random variables with distribution function \( F \), mean 0 and finite variance \( \sigma^2 \). The right tail of \( F \) is taken to be regularly-varying, so that \( \tilde{F}(x) = 1 - F(x) = L(x)/x^\alpha \) for some slowly varying function \( L \) and \( \alpha > 2 \); as is well known, function \( L \) is said to be slowly varying if for all \( t > 0 \),

\[
L(tx)/\tilde{L}(x) \rightarrow 1
\]

as \( x \rightarrow \infty \). Regularly varying distributions form an important class of heavy-tailed distributions. The random walk associated with the collection \( \{X_n : n \geq 1\} \) is given by \( S_0 = 0 \),

\[
S_n = X_1 + \ldots + X_n
\]

for each \( n \geq 1 \). We let \( M(\theta) = E[e^{\theta X_1}] \) and \( \Lambda(\theta) = \ln\{M(\theta)\} \)

 denote the moment and cumulant generating functions of \( X_1 \), respectively.

The following assumption is imposed on the random variables \( X_i \) throughout the paper.

Assumption 1: The left tail of \( F \) is light, that is, there exists \( \phi > 0 \) such that \( M(-\phi) < \infty \).

In Sections III and IV, respectively, we provide strongly efficient algorithms for the following two probabilities:

P1: computation of \( P(S_n > b) \) as \( n \rightarrow \infty \), for \( b \geq \tilde{c} n^{\frac{1}{\alpha} + \epsilon} \) given \( \tilde{c}, \epsilon > 0 \), and

P2: computation of \( P(S_n > b) \) as \( b \rightarrow \infty \) for fixed \( n \).

B. Rare event simulation and importance sampling

Let \( A \) denote a rare event on the probability space \( (\Omega, \mathcal{F}, P) \), i.e., \( z := P(A) > 0 \) is small (in our setup \( A \) corresponds to the event \( \{S_n > b\} \)). Suppose that we are interested in obtaining an estimator \( \hat{z} \) for \( z \) such that the relative error \( |\hat{z} - z|/z < \delta \), with probability at least \( 1 - \epsilon \), for given \( \epsilon \) and \( \delta > 0 \).

Naive simulation for estimating \( z \) involves drawing \( N \) independent samples of the indicator \( I_A \) and taking their sample mean as the estimator. For a different measure \( \tilde{P} \) such that the Radon-Nikodym derivative \( \frac{dP}{d\tilde{P}} \) is well defined on \( A \), we get:

\[
P(A) = \int_A \frac{dP}{d\tilde{P}}(\omega) d\tilde{P}(\omega) = \tilde{E}\left[L I_A\right],
\]

where \( L := \frac{dP}{d\tilde{P}} \) and \( \tilde{E} \) is the expectation associated with \( \tilde{P} \). Define \( Z := L I_A \); then \( Z \) is an unbiased estimator of \( z \) under measure \( \tilde{P} \). If \( N \) i.i.d samples \( Z_1, \ldots, Z_N \) of \( Z \) can be drawn from \( \tilde{P} \), then by strong law of large numbers we have:

\[
\hat{z}_N := \frac{Z_1 + \ldots + Z_N}{N} \rightarrow z \text{ a.s.,}
\]
as $N \nearrow \infty$. This method of generating an estimator is called importance sampling (IS). The measure $\tilde{P}$ is called the importance sampling measure and $Z$ is called an importance sampling estimator.

Using Chebyshev’s inequality allows us to find an upper bound on the required number of samples $N$ to achieve the desired relative precision:

$$
\mathbb{P}\left( \frac{|\hat{z}_n - z|}{z} > \delta \right) \leq \frac{\text{Var}(\hat{z}_n)}{z^2\delta^2} = \frac{CV^2(Z)}{N\delta^2}.
$$

Here $CV(Z) = \sqrt{\text{Var}(Z)}/z$ is the coefficient of variation of $Z$. This enables us to conclude that if we generate at least

$$
N = \frac{CV^2(Z)}{\epsilon^2\delta^2}
$$

i.i.d. samples of $Z$, we can guarantee the desired relative precision.

In naive simulation we use the measure $\bar{P}$ itself and have $Z = 1_{\Lambda}|A$ as the estimator; so the number of samples required in (1) grows (roughly proportional to $z^{-1}$) to infinity if $z \searrow 0$.

As is well known, the choice $P^*(\cdot) := P(\cdot|A)$ as an importance sampling measure yields zero variance for the associated estimator $Z = z1_{\Lambda}$. Then, in simulation every sample equals $z$ with $P^*$ probability 1. However, the explicit dependence of $Z$ on $z$, the quantity which we want to estimate makes this method impractical.

**Efficiency notions of algorithms:** Consider a family of events $\{A_n : n \geq 1\}$ such that $z_n := P(A_n) \searrow 0$ as $n \nearrow \infty$. For an importance sampling algorithm to compute $\{z_n : n \geq 1\}$, we come up with a sequence of changes of measures $\{\tilde{P}_n : n \geq 1\}$ and estimators $\{Z_n : n \geq 1\}$ such that $\mathbb{E}_n Z_n = z_n$, where $\mathbb{E}_n$ denotes the expectation operator under $\tilde{P}_n$.

**Definition 1:** The sequence $\{Z_n : n \geq 1\}$ of unbiased importance sampling estimators of $\{z_n : n \geq 1\}$, is said to be strongly efficient if,

$$
\lim_{n \to \infty} \frac{\mathbb{E}_n (Z_n^2)}{z_n^2} < \infty. \tag{2}
$$

It is said to be weakly efficient if for each $\epsilon > 0$,

$$
\lim_{n \to \infty} \frac{\mathbb{E}_n (Z_n^2)}{z_n^2} < \infty. \tag{3}
$$

From (1), we can see that if an algorithm is strongly efficient, the number of simulation runs required to guarantee desired relative precision stays bounded as $n \nearrow \infty$. It also follows that strong efficiency implies weak efficiency and that naive simulation is not even weakly efficient.

**C. Tail asymptotics for i.i.d sums**

(a) The well-known asymptotics

$$
\mathbb{P}\{S_n > b\} \sim n\bar{F}(b), \quad \text{as } n \nearrow \infty \tag{4}
$$

for $b > \sqrt{n \log n}$ can be found in [23], [24] or [25]. Let $M_n := \max_{k \leq n} X_k$, then it is easily verified that

$$
\mathbb{P}\{M_n > b\} \sim n\bar{F}(b).
$$

Additionally, the following asymptotics can be found in [24]: as $n \nearrow \infty$,

$$
\mathbb{P}\{S_n > b, M_n < b\} = o(n\bar{F}(b))
$$

$$
\sup_{b \geq \sqrt{n \log n}} \mathbb{P}\{N(n,b) = 1|S_n \geq b\} \to 0,
$$

$$
\sup_{b \geq \sqrt{n \log n}} \mathbb{P}\{M_{n-1} \leq b, S_n \geq b|X_n > b\} \to 0,
$$

where $N(n,b)$ denotes the cardinality of $\{1 \leq i \leq n : X_i > b\}$. These large deviations asymptotics reveal that with the number of summands growing to infinity, with high probability, the sum becomes large because of one of the components becomes large.

(b) The following tail probability asymptotic of $\mathbb{P}\{S_n > b\}$ for fixed $n$ can be found, e.g., in [26]:

$$
\mathbb{P}\{S_n > b\} \sim n\bar{F}(b), \quad \text{as } b \nearrow \infty. \tag{5}
$$

**D. Review of existing IS techniques**

1) For light-tails: Suppose that $\{Y_i\}$ are i.i.d., zero mean, light-tailed random variables (that is, their moment generating function exists in a neighborhood of zero) with distribution function $F_Y$. Sadowsky and Bucklew [6] introduce a weakly-efficient method for estimating $\mathbb{P}\{\sum_{i=1}^n Y_i > na\}$, $a > 0$ that involves generating i.i.d. samples $\{Y_i\}$ using the distribution function obtained by exponentially twisting the original distribution function, that is, using

$$
F_\theta(dx) = e^{\theta x - \Lambda_Y(\theta)} F_Y(dx), \tag{6}
$$

where $\Lambda_Y(\theta) = \ln E[e^{\theta Y_i}]$ denotes the log-moment generating function of $Y_i$. [6] propose that $\theta > 0$ be chosen such that $\Lambda_Y'(\theta) = a$, (this is assumed to exist). Note that the resulting estimator has the form:

$$
\hat{Z}_n = \exp \left( - \left( \theta \sum_{i=1}^n Y_i - n\Lambda_Y(\theta) \right) \right) \mathbb{I}_{\{\sum_{i=1}^n Y_i > na\}}
$$

2) For heavy-tails: For regularly varying right tailed random variables $\{X_i\}$, the associated log-moment generating function $\Lambda(\theta) = \infty$, for each $\theta > 0$; therefore the above method of exponential twisting is not directly applicable for simulation of $P(S_n > b)$. Blanchet and Liu [4] propose two efficient state-dependent importance sampling algorithms for solving $P1$. The better of the two involves drawing samples for increment $X_{k+1}$ based on $S_k$ from an appropriately parameterized family of density functions. In particular, they recommend using

$$
\mathbb{P}\{X_{k+1} \in dx|S_k = s\} = p_n F(dx) \frac{F_1(dx)1_{\{x > a(b-s)\}}}{F(a(b-s))} + (1 - p_n) F(dx) \frac{F_2(dx)1_{\{x \leq a(b-s)\}}}{F(a(b-s))},
$$

\footnote{\text{We say that $\{a_n\} \sim \{b_n\}$ if $\frac{a_n}{b_n} \to 1$ as $n \to \infty$}}
where \( p_n \) is the suitably chosen mixture probability and \( a \in (0, 1) \). Asymptotic forms of \( \mathbb{P}\{S_n > b\} \) given in [23] are used to construct a family of functions that satisfy a certain Lyapunov inequality; relevant parameters such as \( p_n \) and \( a \) are chosen in a way that the Lyapunov inequality holds. The state-dependent scheme used in very recent paper [5] avoids using asymptotic approximations of \( \mathbb{P}\{S_n > b\} \) by employing sequential importance sampling and resampling methods.

III. PROPOSED ALGORITHM FOR P1

The proposed importance sampling distribution for drawing samples of \( X_1, \ldots, X_n \) to estimate \( \mathbb{P}\{S_n > b\} \) is:

\[
\tilde{F}_n(dx) := c_n e^{\theta_n(x_n b)} F(dx),
\]

with \( \theta_n := \theta_n(b) \) given by,

\[
\theta_n(b) = \frac{1}{b} \log \left( \frac{1}{nF(b)} \right).
\]

In (7) \( c_n := c_n(b) \) denotes the appropriate normalizing constant, which is well-defined because:

\[
\frac{1}{c_n(b)} = \int_{-\infty}^{b} e^{\theta_n x} F(dx) + e^{\theta_n b} \tilde{F}(b) \leq e^{\theta_n b} = \frac{1}{nF(b)}.
\]

Though we do not emphasize in notation, it can be seen that all \( \tilde{F}_n, \theta_n \) and \( c_n \) defined above depend on both \( n \) and \( b \). Since we suitably modify the tail and circumvent the problem of normalizing constant \( c_n \) becoming 0, we refer to \( \{\tilde{F}_n : n \geq 1\} \) as tail-adjusted exponentially twisted distributions corresponding to \( F \). It is easy to check that \( \theta_n \searrow 0 \) as \( n \to \infty \).

Let \( \tilde{P}_n \) be the probability measure induced when \( \{X_i : i \geq n\} \) are independently distributed according to \( \tilde{F}_n \). Then under the importance sampling change of measure \( \tilde{P}_n \), the unbiased estimator for \( \mathbb{P}\{S_n > b\} \) takes the form:

\[
Z_1(n, b) = \frac{1}{c_n} \exp \left( -\theta_n \sum_{i=1}^{n} (X_i \land b) \right) \mathbb{I}_{\{S_n > b\}}.
\]

One simulation run of the corresponding state-independent importance sampling algorithm is given below:

Algorithm 1:

Given parameters \( b \geq \tilde{c} n^{1+\epsilon}, \tilde{c}, \epsilon > 0 \)

STEP 1: Draw \( n \) i.i.d. samples \( x_1, \ldots, x_n \) from \( \tilde{F}_n \)

STEP 2: \( s \leftarrow x_1 + \ldots + x_n \)

STEP 3: \( L \leftarrow \frac{1}{c_n} e^{\theta_n \sum_{i=1}^{n} (x_i \land b)} \)

STEP 4: RETURN \( L \mathbb{I}_{\{s > b\}} \)

STOP

Doing many independent simulation runs and taking the sample average of the returned values gives an unbiased estimator of \( \mathbb{P}\{S_n > b\} \).

Observe that the IS measure \( \tilde{P}_n \) induced by \( \tilde{F}_n \) is such that the event \( \{X_k > b\} \) receives a considerable mass of \( c_n/b \) under \( \tilde{P}_n \). We provide asymptotic bounds for the integral \( \int_{-\infty}^{b} e^{\theta_n x} F(dx) \) in the appendix and there we prove the following lemma:

Lemma 1: \( \lim_{n \to \infty} c_n = 1 \).

We now argue that the proposed measure \( \tilde{P}_n \) assigns similar probability to events \( \{X_j > x\} \) for each \( x \geq b \) as the zero-variance measure asymptotically as \( n \to \infty \).

Recall that \( \mathbb{P}\{X_j > x|S_n > b\} \) denotes the probability assigned to \( \{X_j > x\} \) under the zero-variance measure. We have,

\[
\mathbb{P}\{X_j > x|S_n > b\} = \frac{\mathbb{P}\{X_j > x\}}{\mathbb{P}\{S_n > b\}}.
\]

Since \( \mathbb{P}\{S_n > b|X_j > x\} \sim 1 \) as \( n \to \infty \), we get

\[
\mathbb{P}\{X_j > x|S_n > b\} \sim \frac{\tilde{F}(x)}{nF(b)}.
\]

It can be verified that \( \tilde{P}_n \{X_j > x\} = c_n \tilde{F}(x) / nF(b) \). Therefore from Lemma 1,

\[
\mathbb{P}\{X_j > x|S_n > b\} \sim \tilde{P}_n \{X_j > x\}, \text{ as } n \to \infty.
\]

A. Strong efficiency of \( Z_1 \)

For proving the strong efficiency of \( Z_1 \) we bound its second moment separately over the elements of the partition \( \{A_k : 0 \leq k \leq n\} \), with \( A_k \) denoting the event of exactly \( k \) of \( X_1, X_2, \ldots, X_n \) crossing \( b \). Recall that \( \theta_n \) and \( c_n \) depend on both \( n \) and \( b \), though in notation we often make only the dependence on \( n \) explicit. The following lemma, useful for establishing strong efficiency of the above algorithm, is proved in the appendix.

Lemma 2: Under Assumption 1, for any given \( \tilde{c}, \epsilon > 0 \), there exists a constant \( D_1 \) such that:

(a) \( \frac{1}{c_n(b)} \leq D_1 \),

(b) \( \left( \frac{\tilde{P}_n \{\exp(-2\theta_n X_1 \mathbb{I}_{\{X_1 < b\}}\}) \}}{c_n} \right)^n \leq D_1 \)

are satisfied \( \forall n \) and \( b \geq \tilde{c} n^{1+\epsilon} \).

The strong efficiency of Algorithm 1 is established in the following theorem.

Theorem 1: Under Assumption 1, Algorithm 1 is strongly efficient when \( b \geq \tilde{c} n^{1+\epsilon} \), that is,

\[
\lim_{n \to \infty} \frac{\tilde{E}_n[Z_1^2(n, b)]}{\mathbb{P}\{S_n > b\}}^2 < \infty,
\]

given any \( b \geq \tilde{c} n^{1+\epsilon}, \forall \tilde{c}, \epsilon > 0 \).

Proof: Recall that \( A_k = \{\omega : \sum_{i=1}^{n} \mathbb{I}_{\{X_i > b\}} = k\} \).

Let \( B_k := A_k \cap \{S_n > b\} \), for \( k = 0, \ldots, n \) and let

\[
I_k(n) := \frac{1}{c_n^{2n}} \tilde{P}_n \left[ \exp \left( -2\theta_n \sum_{i=1}^{n} (X_i \land b) \right) : B_k \right],
\]

then \( \tilde{E}_n[Z_1^2(n, b)] = \sum_{k=0}^{n} I_k(n) \). On the set \( B_0 \), none of the \( \{X_1, \ldots, X_n\} \) cross \( b \) but \( S_n > b \). Therefore,

\[
c_n^{2n} I_0(n) = \tilde{E}_n \left[ e^{-2\theta_n \sum_{i=1}^{n} X_i} : B_0 \right] \leq e^{-2\theta_n b} \tilde{P}_n (B_0) \leq e^{2 \log(nF(b))}.
\]

Thus, \( I_0(n) \leq \frac{(nF(b))^2}{c_n^{2n}} \).

(11)
We observe that $X_i \land b = X_i 1_{\{X_i < b\}} + b 1_{\{X_i \geq b\}}$, and on $B_k, k = 1, \ldots, n \sum_{i=1}^{n} 1_{\{X_i > b\}} = k$. Therefore,
\[
c_n^2 I_k(n) = \frac{\hat{E}_n}{c_n} \left[ e^{-2\theta_n X_i 1_{\{X_i < b\}}} e^{-2\theta_n X_i 1_{\{X_i \geq b\}}}; B_k \right] = \frac{\hat{E}_n}{c_n} \left[ e^{-2\theta_n X_i 1_{\{X_i < b\}}} e^{-2k\theta_n b}; B_k \right] \leq (nF(b))^{2k} \hat{E}_n \left[ e^{-2\theta_n X_i 1_{\{X_i < b\}}} \right],
\]
where in the last line we have used the relation $\theta_n b = -\log nF(b)$, and the fact that $\hat{E}_n[Y; B_k] \leq \hat{E}_n[Y]$ for $Y \geq 0$. Therefore, for $k = 1, \ldots, n$,
\[
I_k(n) \leq \frac{(nF(b))^{2k}}{c_n^2} \left( \frac{\hat{E}_n \left[ e^{-2\theta_n X_i 1_{\{X_i < b\}}} \right]}{c_n} \right)^n. \tag{12}
\]
Now combining the above expression with (11),
\[
\frac{\hat{E}_n \left[ Z_1^2(n, b) \right]}{(nF(b))^2} \leq \frac{1}{c_n^2} + \frac{1}{c_n} \left( \frac{\hat{E}_n \left[ e^{-2\theta_n X_i 1_{\{X_i < b\}}} \right]}{c_n} \right)^n \times \left( \frac{1 - (nF(b))^{2n}}{1 - (nF(b))^2} \right).
\]
From (4), $\mathbb{P}\{S_n > b\} \sim nF(b)$ as $n \rightarrow \infty$ uniformly for $b > \sqrt{n \log n}$. Therefore application of Lemma 2 results in:
\[
\lim_{n \rightarrow \infty} \frac{\hat{E}_n \left[ Z_1^2(n, b) \right]}{(\mathbb{P}\{S_n > b\})^2} \leq D_1^2 + D_1.1 < \infty,
\]
thus establishing the strong efficiency of Algorithm 1.

**Theorem 2:** Under Assumption 1, estimator $Z_2(b)$ is strongly efficient under measure $\hat{P}_b$; that is,
\[
\lim_{b \rightarrow \infty} \frac{\hat{E}_b \left[ Z_2^2(b) \right]}{(\mathbb{P}\{S_n > b\})^2} < \infty. \tag{16}
\]

**Proof:** As before for $k = 1, \ldots, n$, let
\[
B_k := \left\{ \omega : \sum_{i=1}^{n} 1_{\{X_i > b\}} = k \right\} \cap \{S_n > b\}, \text{ and}
\]
\[
I_k(b) := \frac{1}{c_b^2} \hat{E}_b \left[ e^{-2\theta_n X_i 1_{\{X_i < b\}}}; B_k \right].
\]
So we have $\hat{E}_b[Z_2^2(b)] = \sum_{k=1}^{n} I_k(b)$. As in the proof of Lemma 2, we can show that:
\[
\lim_{b \rightarrow \infty} \frac{1}{c_b^2} \hat{E}_b \left[ e^{-2\theta_n X_i 1_{\{X_i < b\}}} \right] \leq \frac{1}{n}. \tag{17}
\]
For $I_k, k = 1, \ldots, n$ we just follow the same procedure to get,
\[
I_0(b) \leq \frac{(nF(b))^2}{c_b^2},
\]
\[
I_k(b) \leq \frac{(nF(b))^{2k}}{c_b^2} \left( \frac{\hat{E}_b \left[ e^{-2\theta_n X_i 1_{\{X_i < b\}}} \right]}{c_b} \right)^n.
\]
Therefore,
\[
\frac{\hat{E}_b \left[ Z_2^2(b) \right]}{(nF(b))^2} \leq \frac{1}{c_b^2} + \frac{1}{c_b} \left( \frac{\hat{E}_b \left[ e^{-2\theta_n X_i 1_{\{X_i < b\}}} \right]}{c_b} \right)^n \times \left( \frac{1 - (nF(b))^{2n}}{1 - (nF(b))^2} \right).
\]
Using the asymptotics $\mathbb{P}\{S_n > b\} \sim nF(b)$ as $b \rightarrow \infty$, (17) and (18) we get:
\[
\lim_{b \rightarrow \infty} \frac{\hat{E}_b \left[ Z_2^2(b) \right]}{(\mathbb{P}\{S_n > b\})^2} \leq \left( 1 + \frac{2}{n} \right)^{2n} + \left( 1 + \frac{2}{n} \right)^n \frac{1}{n^2} < \infty.
\]
This completes the proof.

**V. Numerical examples**

Here we explain the results of simulation experiments for computing $\mathbb{P}\{S_n > na\}$ for different values of $a$ and $n$. Suppose that $\{X_n : n \geq 1\}$ have a Pareto tail distribution in the right side and exponential tail distribution in the left. Specifically, the probability distribution function for $X_i$ is:
\[
F(x) := \begin{cases} \frac{2x}{\alpha}, & \text{if } x < 0, \\ 1 - \frac{1}{2(1+x)^2}, & \text{otherwise} \end{cases}
\]

The random variable $X_i$ has zero mean. We use $N = 1000$ simulation runs to estimate $\mathbb{P}\{S_n > na\}$ for $n = 50, 75, 100$ and $a = 1, 2, 3$. In the implementation of Algorithm 1, we draw samples from the density given by (7). For $a = 1$, we take $\theta_{50} = 0.0745, \theta_{75} = 0.0542, \theta_{100} = 0.0431$ as dictated by (8); $\theta_n$ is chosen similarly for other choices of $a$ as well. Table I provides a comparison of the
performance of Algorithm 1 which runs in $O(n)$ time with that of the $O(n^2)$ algorithm proposed in [3], which is identified as Algorithm BL in the table. In Table I, VR denotes the ratio of the variance of naive simulation estimator to the variance of the estimator corresponding to the specified algorithm. Variance of the naive simulation estimator is taken to be $\hat{\gamma}(1 - \hat{\gamma})$, where $\hat{\gamma}$ denotes the estimate of the probability $\mathbb{P}\{S_n > na\}$ obtained using the specified algorithm. Furthermore, RE denotes the ratio of empirical standard deviation to the estimate $\hat{\gamma}$.

The relative error of the estimates under Algorithm 1 appear to be bounded as $n$ increases, thus suggesting strong efficiency. Further it can be noted that the variance reduction by Algorithm 1 is comparable with that of the Algorithm BL, as $n$ increases. Using 1000 simulation runs, for large values of $n$, it has been observed that Algorithm 1 runs at least 20 times faster than Algorithm BL, thus underscoring the computational advantage offered by the $O(n)$ running time of Algorithm 1 over the $O(n^2)$ running time of Algorithm BL.

VI. CONCLUSION

In this paper we revisited the problem of estimating $P\{S_n > b\}$ in the settings where both $n$ and $b$ increase to infinity (P1) as well as when $n$ is fixed and $b$ increases to infinity (P2) when the constituent random variables have heavy regularly varying right tails. We showed that state-independent exponential twisting based algorithms can be developed to efficiently estimate these probabilities. The current literature on the other hand focuses on estimating P1 using more nuanced state-dependent methods. In many settings, the proposed algorithms offer implementation and computational benefits over state-dependent algorithms. Also, our work re-examines the widely held belief that exponential twisting is applicable only when light-tailed random variables are involved.

We believe that the proposed methods are applicable in greater generality. For instance, in our ongoing effort we attempt to efficiently estimate probabilities such as

$$P\left\{ \max_{k \leq n} S_k \geq b \right\}$$

as both $n$ and $b$ increase to infinity, when the constituent variables have a negative mean and heavy right-tails. Such probabilities have wide application in insurance and in queueing (see, e.g. [27] and [28]).

APPENDIX

We aim to provide proofs of Lemmas 1 and 2 making use of Lemmas 3 and 4, which are stated and proved below. The proof of the Lemma 3 below, which gives asymptotic upper bound for the integral $\int_{-\infty}^{b} e^{\theta_n x} F(dx)$, follows the approach in [25] for bounding similar integral while deriving large deviations asymptotics for $\mathbb{P}\{S_n > b\}$.

Lemma 3: For any pair of sequences $\{x_n\}, \{\phi_n\}$ satisfying $x_n \nearrow \infty$ and $\phi_n x_n \nearrow \infty$, the integral,

$$\int_{-\infty}^{x_n} e^{\phi_n x} F(dx) \leq 1 + \frac{e^{2\alpha} \sigma^2 \phi_n^2}{2} + e^{2\alpha} \hat{F}\left(\frac{2\alpha}{\phi_n}\right) + e^{\phi_n x_F}(x_n(1 + \epsilon_n),$$

where $\epsilon_n \geq 0$ is such that $\epsilon_n \downarrow 0$ as $n \nearrow \infty$.

Proof: We split the region of integration into $(-\infty, \gamma/\phi_n]$, and $(\gamma/\phi_n, x_n]$ for some constant $\gamma > 0$; the partition is such that the integrand stays bounded in the former despite its growth to $(-\infty, \infty)$. Let $I_1 := \int_{-\infty}^{x_n} e^{\phi_n x} F(dx)$ and $I_2 := \int_{\gamma/\phi_n}^{x_n} e^{\phi_n x} F(dx)$. Since $e^{\phi_n x} \leq 1 + \phi_n x + \frac{\phi_n^2 x^2}{2} e^{\phi_n x}$,

$$I_1 \leq \int_{-\infty}^{\gamma/\phi_n} F(dx) + \phi_n \int_{-\infty}^{\gamma/\phi_n} x F(dx) + \frac{\phi_n^2}{2} \int_{-\infty}^{\gamma/\phi_n} x^2 e^{\phi_n x} F(dx)$$

$$\leq \int_{-\infty}^{\infty} F(dx) + \phi_n \int_{-\infty}^{\infty} x F(dx) + \frac{\phi_n^2 e^{\gamma}}{2} \int_{-\infty}^{\infty} x^2 F(dx)$$

$$= 1 + \frac{e^{\gamma} \phi_n^2 \sigma^2}{2},$$

(19)

<table>
<thead>
<tr>
<th>n</th>
<th>ALGORITHM 1</th>
<th>ALGORITHM BL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a=1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.905×10^{-4}</td>
<td>6.591×10^{-1}</td>
</tr>
<tr>
<td>75</td>
<td>1.037×10^{-4}</td>
<td>1.145×10^{-1}</td>
</tr>
<tr>
<td>100</td>
<td>4.887×10^{-5}</td>
<td>3.213×10^{-1}</td>
</tr>
<tr>
<td>$a=2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>2.833×10^{-5}</td>
<td>7.677×10^{-1}</td>
</tr>
<tr>
<td>75</td>
<td>1.098×10^{-5}</td>
<td>2.336×10^{-1}</td>
</tr>
<tr>
<td>100</td>
<td>6.641×10^{-6}</td>
<td>6.369×10^{-1}</td>
</tr>
<tr>
<td>$a=3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>6.947×10^{-6}</td>
<td>3.725×10^{-1}</td>
</tr>
<tr>
<td>75</td>
<td>3.489×10^{-6}</td>
<td>8.575×10^{-1}</td>
</tr>
<tr>
<td>100</td>
<td>1.785×10^{-6}</td>
<td>1.617×10^{-1}</td>
</tr>
</tbody>
</table>

As is apparent from the VR column of Algorithm 1, the proposed IS scheme gives large variance reduction compared to naive simulation; the RE column appears to be bounded, thus suggesting strong efficiency; also for large $n$, the variance reduction offered by Algorithm 1 is comparable with that of Algorithm BL. Recall that the computational effort for sample generation in Algorithm 1 and Algorithm BL scale like $O(n)$ and $O(n^2)$, respectively.
because $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = \sigma^2$. Integrating by parts for the second integral,

$$I_2 = -\int_{\gamma/\phi_n}^{x_n} e^{\phi_n x} \tilde{F}(dx)$$

$$= e^{\phi_n \gamma/\phi_n} \tilde{F}(\gamma/\phi_n) - e^{\phi_n x_n} \tilde{F}(x_n) + \phi_n \int_{\gamma/\phi_n}^{x_n} e^{\phi_n x} \tilde{F}(dx)$$

$$\leq e^{\gamma} \tilde{F}(\gamma/\phi_n) + I_2', \ (20)$$

where, $I_2' := \phi_n \int_{\gamma/\phi_n}^{x_n} e^{\phi_n x} \tilde{F}(dx)$. Now the change of variable $u = \phi_n (x_n - x)$ results in:

$$I_2' = e^{\phi_n x_n} \int_{0}^{\phi_n x_n - \gamma} e^{-u \tilde{F}(x_n - u/\phi_n)} \ du$$

$$= e^{\phi_n x_n} \tilde{F}(x_n) \int_{0}^{\phi_n x_n - \gamma} e^{-u} g_n(u) \ du, \ (21)$$

where,

$$g_n(u) := \frac{\tilde{F}(x_n - u/\phi_n)}{\tilde{F}(x_n)} = \frac{\tilde{F}(x_n (1 - u/\phi_n x_n))}{\tilde{F}(x_n^\gamma) \phi_n}.$$ 

Since $L$ is slowly varying and $\phi_n x_n \to \infty$, for all $n$ large enough we have:

$$(1-\delta) \left(1 - \frac{u}{\phi_n x_n}\right)^{-\alpha + \delta} \leq g_n(u) \leq (1+\delta) \left(1 - \frac{u}{\phi_n x_n}\right)^{-\alpha - \delta}.$$ (22)

This preliminary fact about slowly varying functions, which is Theorem 1.1.4 of [25], helps in concluding that $g_n(u) \sim 1$ as $n \to \infty$ for any $u \in [0, \phi_n x_n - \gamma]$.

Now fix $\delta = \frac{\alpha}{2}$. Then for large enough $n$,

$$g_n(u) \leq \left(1 + \frac{\alpha}{2}\right) \left(1 - \frac{u}{\phi_n x_n}\right)^{-\frac{2\alpha}{\gamma}}.$$ (22)

To show that $g_n(u)$ is uniformly bounded in $n$ and $u \in [0, \phi_n x_n - \gamma]$ by an integrable function, so that we can apply dominated convergence, we call the term involving $u$ in the upper bound in (22) as $h(u)$, that is $h(u) = (1 - u/\phi_n x_n)^{-\frac{2\alpha}{\gamma}}$.

Since $\log h(0) = 0$ and $\frac{du}{du} (\log(h(u))) \leq \frac{2\alpha}{\gamma}$ for $0 \leq u \leq \phi_n x_n - \gamma$, we have $h(u) \leq e^{-\frac{2\alpha}{\gamma}}$ for the same interval.

Therefore if we choose $\gamma = 2\alpha$, the integrand in $I_2'$ is bounded for large enough $n$ by an integrable function as below:

$$|e^{-u} g_n(u) 1_{\{0 \leq u \leq \phi_n x_n - \gamma\}}| \leq |e^{-u} \left(1 + \frac{\alpha}{2}\right) \left(1 - \frac{u}{\phi_n x_n}\right)^{-\frac{2\alpha}{\gamma}} h(u) 1_{\{0 \leq u \leq \phi_n x_n - \gamma\}}|$$

$$\leq \left(1 + \frac{\alpha}{2}\right) e^{-u + \frac{2\alpha}{\gamma}}$$

$$= \left(1 + \frac{\alpha}{2}\right) e^{-u - \frac{2\alpha}{\gamma}}.$$ (22)

Applying dominated convergence theorem, we get

$$\int_{0}^{\phi_n x_n - \gamma} e^{-u} g_n(u) \ du \sim 1 \ as \ n \to \infty.$$ (22)

Since $\int_{-\infty}^{x_n} e^{\phi_n x} \tilde{F}(dx) = I_1 + I_2$, combining this result with (19), (20) and (21), completes the proof.  

**Lemma 4:** For $b \geq \tilde{c} n^{2+\epsilon}$, given $\epsilon$ and $\geq 0$, if $\theta_n$ is given by (8), then the following hold true as $n \to \infty$:

(a) $\theta_n = o\left(\frac{1}{\sqrt{n}}\right)$,

(b) $\tilde{F}\left(\frac{2\alpha}{\theta_n}\right) = o\left(\frac{1}{n}\right)$.

**Proof:** (a) We have $\tilde{F}(x) = \frac{L(x)}{x^2}$. Since $L$ is slowly varying, given any $\delta > 0$ for sufficiently large values of $b$, we have $b^{-\delta} \leq L(b) \leq b^\delta$, thus yielding $L(b) = b^{\sigma(1)}$ as $b \to \infty$.

Since $b \geq \tilde{c} n^{2+\epsilon}$, $n/b^2 \leq \tilde{c}^{-1} n^{-2\epsilon}$. Then,

$$n\theta_n^2 = \frac{n}{b^2} \log^2 \left(\frac{1}{n\tilde{F}(b)}\right) \leq \frac{n}{b^2} \log^2 \left(\frac{b^\delta}{n L(b)}\right)$$

$$= O\left(\frac{\log^2 n}{n^{2\epsilon}}\right) \to 0, \ as \ n \to \infty.$$ (22)

(b) Since $\tilde{F}$ is regularly varying. By Theorem 1.1.4 (iii) of [25], given any $\delta > 0$, for $n$ large enough,

$$n\tilde{F}\left(\frac{2\alpha}{\theta_n}\right) = n\tilde{F}\left(-\log(n\tilde{F}(b))\right) \leq n\left(-\log(n\tilde{F}(b))\right)^{\alpha + \delta}$$

$$= O\left((\log n)^{\alpha + \delta}\right) \frac{n L(b)}{b^\alpha} \to 0, \ as \ n \to \infty.$$ (22)

because $\alpha > 2$ and $L$ is slowly varying.  

Now using the above two lemmas, we provide proofs for the results left unproved in previous sections.

**Proof of Lemma 1**

Since $c_n(b)$ is the normalizing constant in (7),

$$\frac{1}{c_n(b)} = \int_{-\infty}^{b} e^{\theta_n x} \tilde{F}(dx) + e^{\theta_n b} \tilde{F}(b)$$

$$\leq 1 + e^{2\alpha} \sigma^2 \theta_n^2 + e^{2\alpha} \tilde{F}\left(\frac{2\alpha}{\theta_n}\right) + \frac{1}{n}(2 + \epsilon_n), \ (23)$$

where we have used the Lemma 3 and the relationship, $\theta_n b = -\log(n\tilde{F}(b))$. Usage of Lemma 3 is justified because $\theta_n b \sim \infty$ as $n \to \infty$. Since the terms involved are of $o(1)$ as noted in Lemma 4, $\lim_{n \to \infty} c_n \geq 1$. Now for the other side,

$$\frac{1}{c_n} = \int_{-\infty}^{b} e^{\theta_n x} \tilde{F}(dx) + \frac{1}{n}$$

$$\geq \tilde{F}(b) e^{\theta_n b} \tilde{F}(b) + \frac{1}{n},$$

where the last inequality is due to Jensen’s inequality. Since $\int_{-\infty}^{b} x \tilde{F}(dx) \to 0$, $\theta_n \to 0$ and $\tilde{F}(b) \to 1$, as $n \to \infty$ we get $\lim_{n \to \infty} c_n \leq 1$.  

**Proof of Lemma 2**

(a) In (23) if we use $1 + x \leq e^x$ for positive $x$, then

$$\frac{1}{c_n^2(b)} \leq e^{2\alpha x} \sigma^2 \theta_n^2 + \epsilon_n \tilde{F}\left(\frac{2\alpha}{\theta_n}\right) + n \frac{1}{(2 + \epsilon_n)}$$

$$= e^{o(1)+1.2(2+\epsilon_n)} \to e^2,$$
where the final equality is due to application of Lemma 4 for the first two terms in the exponent. Thus we have established that $c_n^{-n}(b)$ stays bounded.

(b) Recollect that $\Lambda$ is the cumulant generating function of $X_1$ under $\mathbb{P}$. Now consider,

$$
\mathbb{E}_n \left[ e^{-2\theta_n X_1 1_{\{X_1 < b\}}} \right] = \int_{-\infty}^{b} e^{-2\theta_n x} \bar{F}_n(x) \, dx + 1 - \bar{F}_n(b) = c_n \int_{-\infty}^{b} e^{-\theta_n x} F(dx) + c_n \frac{1}{n} \leq c_n \left( e^{\Lambda(-\theta_n)} + \frac{1}{n} \right),
$$

(24)

where the second equality follows from (7) and using $\tilde{n}^{-1}$. The continuity of $\Lambda$ in $(-\phi, 0)$, Taylor expansion of $\Lambda$ around 0 gives

$$
\Lambda(-\theta_n) = 0 + \frac{\theta_n^2 \sigma^2}{2} \tilde{n} \tilde{\sigma}^2,
$$

for some $\tilde{\sigma}$ between $-\theta_n$ and 0. Since $\Lambda(0) = \sigma^2$, given any $\delta > 0$, we use continuity of $\Lambda''$ to write $\Lambda''(\theta_n) \leq \sigma^2 (1 + \delta)$, for all $n$ large enough. Thus we conclude that, given any $\delta > 0$, for sufficiently large values of $n$, $\Lambda(-\theta_n) \leq \frac{\theta_n^2 \sigma^2 (1 + \delta)}{2}$. Therefore from (24), we can write:

$$
\frac{\mathbb{E}_n \left[ e^{-2\theta_n X_1 1_{\{X_1 < b\}}} \right]}{c_n} \leq e^{\theta_n^2 \sigma^2 (1 + \delta)} \frac{1}{n}.
$$

Then,

$$
\left( \frac{\mathbb{E}_n \left[ e^{-2\theta_n X_1 1_{\{X_1 < b\}}} \right]}{c_n} \right)^n \leq e^{n \theta_n^2 \sigma^2 (1 + \delta)} \left( 1 + \frac{1}{n e^{\theta_n^2 \sigma^2 (1 + \delta)}} \right)^n \leq e^{n \theta_n^2 \sigma^2 (1 + \delta) + e^{-\theta_n^2 \sigma^2 (1 + \delta)}},
$$

(26)

holds true for all $n$ large enough. Here we have used $1 + x \leq e^x$ for $x$ positive. Since we have $n \theta_n^2 = o(1)$ from Lemma 4, (26) results in,

$$
\lim_{n \to \infty} \left( \frac{\mathbb{E}_n \left[ e^{-2\theta_n X_1 1_{\{X_1 < b\}}} \right]}{c_n} \right)^n \leq e^{0 + e^0} = e.
$$

Therefore we can find an appropriate constant $D_1$ such that both the inequalities stated in Lemma 2 hold true.

REFERENCES


[23] L. V. Rozovski{(k˘ı)}, “Probabilities of large deviations of sums of independent random variables with a common distribution function that belongs to the domain of attraction of a normal law,” *Teor. Verojatnost.*


