Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

EFFICIENT SIMULATION FOR LARGE DEVIATION PROBABILITIES OF SUMS OF HEAVY-TAILED INCREMENTS

Jose H. Blanchet

Department of Statistics
Harvard University
Cambridge, MA 02138, U.S.A.

Jingchen Liu

Department of Statistics
Harvard University
Cambridge, MA 02138, U.S.A.

ABSTRACT

Let $(X_n : n \geq 0)$ be a sequence of iid rv's with mean zero and finite variance. We describe an efficient state-dependent importance sampling algorithm for estimating the tail of $S_n = X_1 + ... + X_n$ in a large deviations framework as $n \nearrow \infty$. Our algorithm can be shown to be strongly efficient basically throughout the whole large deviations region as $n \nearrow \infty$ (in particular, for probabilities of the form $P(S_n > \kappa n)$ as $\kappa > 0$). The techniques combine results of the theory of large deviations for sums of regularly varying distributions and the basic ideas can be applied to other rare-event simulation problems involving both light and heavy-tailed features.

1 INTRODUCTION

We are interested in efficient simulation of large deviation probabilities for sums of heavy-tailed iid (independent and identically distributed) rv's (random variables). In particular, our main goal is to provide a framework, based on state-dependent importance sampling, to simulate such probabilities with bounded relative error. In order to put our results in perspective, let us briefly discuss the problem at hand in the setting of light-tailed rv's. One of the most basic examples that arises in rare-event simulation considers large deviation probabilities for sums of iid rv's with finite (and steep) moment generating function (in particular, with light-tailed increments). This example may constitute the most natural starting point of any introductory discussion on rare-event simulation, not only because sums of iid rv's arise in a huge number of application settings but also because, methodologically, this example illustrates key ideas behind the design of importance samplers that take advantage of the theory of large deviations (see Bucklew 1990 and Dembo and Zeitouni 1998). The connection between importance sampling and large deviations theory has been thoroughly studied in substantial generality (see, for instance, Bucklew 1990, Heidelberger 1995, Asmussen et al. 1994, Sadowsky 1996, Dieker and Mandjes 2005 and Dupuis and Wang 2004). As a consequence, understanding the ideas behind rare-event simulation of large deviation probabilities for sums of iid rv's in light-tailed systems has helped substantially in the development of general rare-event simulation methodology based on large deviation theory.

In the presence of heavy-tailed features, the connection between large deviations theory and importance sampling has been studied substantially in recent years. The work by Asmussen et al. (2000) provides a number of examples that indicates some of the problems that arise when trying to apply importance sampling ideas based on large deviations theory for systems with heavy-tailed characteristics; see also Bassamboo, Juneja and Zeevi (2005) for additional issues arising in rare-event simulation for systems with heavy tails. Perhaps the simplest way to see the types of issues that arise in rare-event simulation of systems with heavy-tailed features steams from the fact that often, in the heavy-tailed setting, the large deviations behavior of the system is caused by one or few components with "extreme" behavior, while the rest of the system is operating in "normal" circumstances. Implementing a change-of-measure with such characteristics is may be non-trivial, the reason is that – in general – the contribution of the sample paths that lie in the large deviations domain for which none or very few components exhibit "extreme" behavior is non-negligible (for the purpose of controlling the variance of the estimator). It is then important (and often not easy) to have a good understanding of the contribution of the sample paths that will add to the variance significantly but are negligible in terms of the most likely asymptotic behavior of the system.

Motivated by the problem of providing direct means for connecting large deviations and importance sampling for systems with heavy-tailed characteristics, we revisit large deviation probabilities for sums of iid rv's – although, in our case, in a heavy-tailed context. More precisely, set $S_0 = 0$ and define $S_n = X_1 + ... + X_n$ where the X_i's have regularly varying right tails with index $\alpha > 2$ (i.e., $P(X > xt)/P(X > t) \longrightarrow x^{-\alpha}$ as $t \nearrow \infty$ for each

$x \in (0, \infty)$). Further, assume that the X_i's have mean zero and finite variance (precise conditions are given in Section 4 below). Our goal is to develop an efficient importance sampling algorithm for estimation of tail probabilities of the form $P(S_n > b)$ when both n and b are large in a large deviations regime. In particular, assuming some technical regularity conditions on the density of the increments, we shall develop an efficient importance sampling algorithm that:

1. Is based on a state-dependent change-of-measure.
2. Takes advantage of large deviations estimates that are used to approximate the zero-variance change-of-measure – which can be characterized by a system of linear equations.
3. Is based on basic principles that can be applied to many other rare-event simulation problems involving systems with heavy-tailed or light-tailed characteristics.
4. Provides a strongly efficient estimator for large deviations probabilities of the form $P(S_n > b)$ whenever $b \geq cn^{1/2+\varepsilon}$ for $\varepsilon > 0$. In other words, the algorithm is efficient basically throughout the whole large deviations domain.

Item 3 is particularly convenient because in practical situations the tail parameter b is not explicitly given as a function of n, so, it is desirable to have an algorithm that is guaranteed to exhibit good performance over a wide region of the space.

Going back to our discussion in the context of light-tailed increments. It is important to recall that the importance sampling algorithm based on the standard exponential change-of-measure, although logarithmically efficient, is not strongly efficient (i.e., the squared coefficient of variation is not bounded as a function of n but grows at a subexponential rate, typically is of order $O\left(n^{1/2}\right)$ as $n \nearrow \infty$). However, the ideas behind the state-dependent algorithm that we propose here can be adapted in light-tailed situations to provide strongly efficient algorithms (see Blanchet and Glynn 2006b). The robustness characteristic discussed in item 3 is a feature that is also shared by the standard exponential change-of-measure in the context of light-tailed increments but one needs to impose the extra condition $b \leq c_1 n$ for some $c_1 > 0$ (of course, as we indicated before, the estimator remains weakly efficiency).

Rare-event simulation for sums of heavy-tailed increments has been studied in the literature. However, most of the previous work in this line has considered the case in which n is fixed (n is often allowed to be a random variable with exponentially decaying tails) and $b \nearrow \infty$. This situation is motivated by the problem of efficient estimation of the tail distribution of the steady-state waiting time of the $M/G/1$ queue, which – by virtue of the Pollaczek-Khintchine representation – can be expressed as a sum of a random number of positive iid rv's with an explicit distribution (in fact, a Geometric number of such positive rv's). Several efficient algorithms have been developed for the situation in which n is fixed, the first one which applies to the case of regularly varying increments and is based on conditional Monte Carlo was proposed by Asmussen et al. (2000), Juneja and Shahabuddin (2002) proposed an importance sampling algorithm based on Hazard rate twisting. Recent work of Asmussen and Kroese (2006) proposed improved estimators that are strongly efficient and have excellent practical performance. Dupuis, Leder, and Wang (2006) proposed algorithms that are based on state-dependent importance sampling, their change-of-measure, although motivated from a control-theoretic perspective, is perhaps the closest in nature to the one discussed here – although is not designed to cover the case in which n is large. An efficient rare-event simulation algorithm for the tail of the steady-state waiting time in the $GI/G/1$ queue was recently proposed by Blanchet and Glynn (2006). The method proposed in this paper shares important similarities to the algorithm given in Blanchet and Glynn (2006) for the $GI/G/1$ queue. In particular, the basic idea here also involves the use of asymptotic results that approximate the zero-variance change-of-measure, however, the fact that the large deviations domain in our current context is parameterized by both *time* and *space* introduces an interesting twist to the method studied in Blanchet and Glynn (2006). In particular, the proposed change-of-measure suggested here should control the behavior of the estimator on the domain of both large and moderate deviations.

The general strategy proposed here involves three steps. First, we need to characterize the zero-variance change-of-measure by means of a system of linear equations – this step provides the structure of the importance sampler that will be proposed using approximations based on large deviations theory. Second, we must study the large deviations theory that applies to the problem at hand, in this case large deviations for sums of regularly varying random variables, and analyze changes-of-measure suggested by the large deviations theory. This step involves studying the behavior of the estimators suggested (in combination with step 1) by theory of large deviations. The last step involves a careful analysis of the likelihood ratio on a large deviations domain of interest and the use of this analysis in the estimates of the second moment of the likelihood ratio.

The rest of the paper is organized as follows. In Section 2 we review basic concepts related to strong efficiency of rare-event simulation estimators and basic notions on importance sampling. Section 3 provides a characterization of the optimal change-of-measure. In Section 4 we review basic notions of large deviations theory for sums of regularly varying distributions and propose a convenient change-of-measure; this part corresponds to step 2 in the outline given

in the previous paragraph. In Section 5 we provide an outline of a proposed algorithm and analyze its efficiency.

2 ON IMPORTANCE SAMPLING AND EFFICIENT RARE-EVENT SIMULATION

Suppose that we are interested in estimating $P(Z \in A) > 0$, for a given random object taking values on a space $\mathcal{X}$ with a σ-field $\mathcal{B}$. We define the probability distribution $F_Z(dz)$ on $(\mathcal{X}, \mathcal{B})$ via $F_Z(dz) = P(Z \in dz)$, so that $P(Z \in A)$ can be expressed as

$$P(Z \in A) = \int_A F_Z(dz).$$

Let $G(dz)$ be any probability distribution on $(\mathcal{X}, \mathcal{B})$ and assume that the likelihood ratio $L(z) = I_A(z)(dF/dG)(z)$ is well defined. Then,

$$P(Z \in A) = E^G[L(Z)] = \int L(z) G(dz), \qquad (1)$$

note that we are using $E^G(\cdot)$ to denote an expectation that is computed under the distribution $G(\cdot)$ (similarly, we will use $Var_G(\cdot)$ for variances under $G(\cdot)$ and so on).

Importance sampling takes advantage of representation (1) in order to estimate $P(Z \in A)$. We can simulate m iid copies of Z using the distribution $G(\cdot)$ and output the importance sampling estimator

$$W_{IS}^{(m)} = \frac{1}{m} \sum_{k=1}^{m} L(Z_k).$$

By the LLN's and identity (1), this is a consistent estimator of $P(Z \in A)$ as $m \nearrow \infty$. Note that importance sampling can achieve zero variance. Indeed, if

$$G(dz) = P(Z \in dz | Z \in A)$$

we have that

$$G(dz) = \frac{P(Z \in dz) I(z \in A)}{P(Z \in A)}$$

and therefore, say, if $m = 1$ and $Z = z_1$

$$W_{IS} = L(z_1) = P(Z \in dz_1) \left[\frac{P(Z \in dz_1)}{P(Z \in A)} \right]^{-1} = P(Z \in A).$$

So, our estimate of $P(Z \in A)$ is exact (in particular, it has zero variance). Obviously, this importance sampling algorithm cannot be implemented in most applications because it requires knowledge of $P(Z \in A)$, which is the quantity of interest. However, this discussion indicates that a good importance sampling distribution should be similar to the conditional distribution of Z given that $Z \in A$.

In our context, $S_n = Z$ and $A = [b, \infty)$. Thus, to construct good importance sampling schemes, it is necessary to obtain a description of the distribution of the increments (X_i's) given that $S_n > b$ in a large deviations regime – developing a good understanding of this distribution is one of the main goals of large deviations theory. One of the goals of this paper is to translate this understanding into a simulation algorithm that satisfies good theoretical properties in terms of efficiency – properties that we review next.

We are interested in measuring the "efficiency" of a procedure for estimating $\beta \equiv P(Z \in A)$ via simulation in a setting in which $\beta \approx 0$. In order to be precise, we shall introduce a parameter n such that $\beta_n \equiv P(Z_n \in A_n) \longrightarrow 0$ as $n \nearrow \infty$ and perform our complexity analysis under this asymptotic regime.

We want to produce an estimate, $\widehat{\beta}_{n,r}$, such that for given $\varepsilon, \delta \in (0,1)$, $\left| \widehat{\beta}_{n,r} - \beta_n \right| \leq \beta_n \varepsilon$ with probability $(1-\delta)$. Here we use the subindex r to denote the number of iid replications required to produce $\widehat{\beta}_{n,r}$. Suppose that $L_{n,k}\left(Z_n^k\right)$ denotes the likelihood ratio obtained in the k-th replication using a given importance sampling estimator. We then consider the unbiased estimator

$$\widehat{\beta}_{n,r} = \frac{1}{r} \sum_{k=1}^{r} L_{n,k}\left(Z_n^k\right).$$

We can use the CLT to conclude that if $\sigma_n^2 = Var_G(L_{n,k}) < \infty$ (where $Var_G(\cdot)$ is the variance computed under the importance sampling distribution) and r is large,

$$P\left(\frac{r^{1/2}}{\sigma_n} \left| \widehat{\beta}_{n,r} - \beta_n \right| \geq z_{1-\delta/2} \right) \approx \delta,$$

where $P\left(N(0,1) \leq z_{1-\delta/2}\right) = 1 - \delta/2$. In other words, we must have that

$$\frac{\sigma_n z_{1-\delta/2}}{\beta_n \varepsilon} \approx r^{1/2}.$$

We therefore require $O(\varepsilon^{-2} z_{1-\delta/2}^2 (\sigma_n/\beta_n)^2)$ replications to achieve the required level of precision. The quantity $cv_n \equiv \sigma_n/\beta_n$ is known as the coefficient of variation of the random variable $L_n(Z_n)$ and (as indicated above) governs the number of replications required to generate an estimator with a controlled relative error.

The best performance (up to proportionality constants and for fixed δ and ε) is obtained by importance sampling

distributions such that

$$\overline{\lim_{n\to\infty}}\frac{\sigma_n}{\beta_n} < \infty. \tag{2}$$

Thus, such an importance sampling distribution is said to achieve *"strong efficiency"* (one can also find the terms "bounded efficiency" or "bounded relative error"). Other notion that appears often in the rare-event simulation literature is *"logarithmic efficiency"* which requires $cv_n = o\left(\log \beta_n^{-1}\right)$ as $n \nearrow \infty$.

3 ON THE OPTIMAL IMPORTANCE SAMPLER FOR $P\{S_n > b\}$

For $0 \leq k \leq n$ and $x \in \mathbb{R}$ we consider

$$u\left(n-k,x\right) = P\left(S_n > b \middle| S_k = x\right).$$

It is not hard to see that $u\left(\cdot\right)$ is satisfies the system of linear integral equations

$$\begin{aligned} u\left(n-k,x\right) &= E[\,u\left(n-k-1,S_{k+1}\right)\middle| S_k = x] \quad (3)\\ &= E[u\left(n-k-1,x+X\right)] \end{aligned}$$

subject to $u\left(0,x\right) = I\left(x > b\right)$. In particular, we note that equation (3) indicates that, for $k \in \{0,...,n-1\}$,

$$Q_{n-k}^{*}\left(x,y+dy\right) = \frac{P\left(x+X \in y+dy\right)u\left(n-k-1,y\right)}{u\left(n-k,x\right)} \tag{4}$$

is a well defined (time in-homogeneous) Markov transition kernel. If we use $Q_{n-k}^{*}\left(s_k,\cdot\right)$ as our importance sampling distribution to generate the sample $S_{k+1} = s_{k+1}$ given that $S_k = s_k$, the contribution to the likelihood ratio in our importance sampling estimator is given by the ratio

$$\frac{P\left(X \in s_{k+1}-s_k+dy\right)}{Q_{n-k}^{*}\left(s_k,s_{k+1}+dy\right)} = \frac{u\left(n-k,s_k\right)}{u\left(n-k-1,s_{k+1}\right)},$$

which corresponds to the generation of the k-th increment $s_{k+1}-s_k$. Thus, the combined likelihood ratio of the generated increments (ranging $k \in \{0,...,n-1\}$) takes the form (observing that we must have that $s_n > b$ and using the condition $u\left(0,x\right) = 1\left(x > b\right)$)

$$L^{*} = \prod_{k=0}^{n-1}\frac{u\left(n-k,s_k\right)}{u\left(n-k-1,s_{k+1}\right)} = \frac{u\left(n,s_0\right)}{u\left(0,s_n\right)} = u\left(n,s_0\right).$$

Therefore, $Q_{n-k}^{*}\left(\cdot\right)$ gives rise to the zero-variance importance sampling estimator and provides a description of the process $\left(S_k : 0 \leq k \leq n\right)$ conditioned on the event $\{S_n > b, S_0 = x_0\}$. The previous (somewhat heuristic) description of the zero-variance change-of-measure enhances intuitive ideas borrowed from importance sampling, a more precise description can be given in terms of Martingale theory – we shall not discuss this approach here many texts deal with this approach in one form or another, see for instance Meyn and Tweedie (1993) or Del Moral (2004).

4 LARGE DEVIATIONS FOR SUMS OF REGULARLY VARYING INCREMENTS

The following assumptions are imposed throughout the rest of the paper. Assume that $\left(X_n : n \geq 1\right)$ is a sequence of iid rv's that satisfy:

i) The X_n's possess a continuously differentiable density $f\left(\cdot\right)$ that satisfies $f'\left(t\right) \sim L\left(t\right)t^{-(\alpha+2)}$ as $t \nearrow \infty$ for $\alpha > 2$, where $L\left(\cdot\right)$ is a slowly varying function at infinity. That is, $L\left(tx\right)/L\left(t\right) \longrightarrow 1$ as $t \nearrow \infty$ for each $x \in (0,1)$.

ii) $EX_n = 0$, $EX_n^2 = 1$ and $E\left|X\right|^p < \infty$ for $p \geq 2$ (note that $\alpha \geq p$ must be satisfied).

We are interested in efficient rare-event simulation methodology for

$$u\left(n,0\right) = P\left(S_n > b \middle| S_0 = 0\right) \text{ for } b \geq cn^{1/2+\varepsilon} \text{ given } c,\varepsilon > 0,$$

when n is large.

The large deviations theory for sums of the form $S_n = X_1 + ... + X_n$ ($S_0 = 0$) has been studied extensively. In particular, a result developed by Rozovskii (1989) (see also Borovkov and Borovkov 2001) provides a convenient asymptotic description of the distribution of S_n for large n. In order to describe Rozovskii's result we shall introduce some notation, which we shall use throughout the rest of the paper, namely,

$$\begin{aligned} \overline{F}\left(b\right) &\equiv P\left(X > b\right),\\ \overline{\Phi}\left(z\right) &\equiv P\left(N\left(0,1\right) > z\right), \end{aligned}$$

where $N\left(0,1\right)$ denotes a standard Gaussian rv. We now state Rozovskii (1989).

Theorem 1 *Under assumptions i) and ii) above,*

$$P\left(S_n > b\right) = \left[n\overline{F}\left(b\right) + \overline{\Phi}\left(b/n^{1/2}\right)\right]\left(1+o\left(1\right)\right)$$

as $n \nearrow \infty$ uniformly over $b > n^{1/2}$.

Theorem 1 indicates that if define (for $0 \leq k \leq n-1$)

$$\begin{aligned} & v\left(n-k,x\right)\\ \equiv\ & \left[\left(n-k\right)\overline{F}\left(b-x\right) + \overline{\Phi}\left(\frac{b-x}{\sqrt{n-k}}\right)\right] I\left(b-x > \sqrt{n-k}\right)\\ & + I\left(b-x \leq \sqrt{n-k}\right), \end{aligned}$$

then $v\left(n-k,x\right)$ is approximately close to $u\left(n-k,x\right)$ in a suitable asymptotic sense (in particular, if $\left(b-x\right) >$

$(n-k)^{1/2}$ and n is much larger than k). Our discussion in Section 3 on the characterization of the zero-variance importance sampling distribution suggests to consider, for $k \in \{0, ..., n-1\}$ the (time in-homogeneous) Markov kernel

$$Q_{n-k}(x, y+dy) = \frac{P(X+x \in y+dy)\, v(n-k-1, y)}{w(n-k, x)},$$

where $w(n-k, x)$ is defined as

$$w(n-k, x) = E[v(n-k-1, x+X)].$$

Since $v(\cdot)$ is a suitable approximation to $u(\cdot)$ in a certain asymptotic sense, it is natural to expect that the "local" likelihood ratio $(w(n-k, x)/v(n-k-1, y))$ will be well behaved on a certain region, described in terms of time and space, that represents the domain under which the asymptotics provide a description that is close enough to the behavior of $u(\cdot)$. So, it is intuitively clear that the performance of our algorithm (in terms of the behavior of the local likelihood ratios and the ability to mimic the zero-variance change-of-measure) will be very good in the domain under which the asymptotics are guaranteed to be good enough. In order to describe a convenient region under which importance sampling can be successfully applied (i.e., the local likelihood ratio is suitably controlled) we define the stopping time (given $\varepsilon, c > 0$ and $n \in \mathbb{N}$)

$$\tau_{n,\varepsilon,c} = \inf\{k \geq 0 : S_k > b - c(n-k)^{1/2+\varepsilon}\}.$$

An important part of our analysis will involve identifying a region under which importance sampling can be successfully applied, a computation that will be done in the next section. In the mean time, let us state the following estimate which will be useful in the design (and analysis) of our efficient simulation algorithm. It concerns the distribution of the first passage time $\tau_n(\varepsilon, c)$ and follows as a direct consequence of Theorem 1 from Borovkov and Borovkov (2001).

Proposition 1 *If assumptions i) and ii) are in force,*

$$P\left(\tau_{n,\varepsilon,c/2} \leq n\right) = (1+o(1)) \sum_{j=1}^{n} \overline{F}\left(b - 2^{-1} c (n-j)^{1/2+\varepsilon}\right)$$

as $n \nearrow \infty$ uniformly over $b \geq cn^{1/2+\varepsilon}$ for fixed constants $\varepsilon, c > 0$.

Equipped with the previous two estimates, we are ready to describe our efficient importance sampling scheme.

5 A PROPOSED ALGORITHM AND EFFICIENCY ANALYSIS

As we indicated before, the natural strategy is to apply the importance sampling scheme suggested by $Q_{n-k}(\cdot)$ if we are well inside a domain where the asymptotic description of $u(\cdot)$ is guaranteed to hold – for purposes of rare-event simulation. In particular, we propose to apply importance sampling at time k as long as $k < \tau_{n,\varepsilon,c/2}$. Large deviations theory for heavy-tailed sums of iid rv's indicates that large excedences occur by means of a single large increment and the size of the overshoot over a given barrier is large. Consequently, it is plausible to expect that a good importance sampling algorithm that is applied only at steps $k < \tau_{n,\varepsilon,c/2}$ (whenever $\tau_{n,\varepsilon,c/2} \leq n$) should induce the occurrence of the event $\{S_n > b\}$ with relatively high probability. Therefore, it should be the case that a simple algorithm that just stops the importance sampling scheme by time $\tau_{n,\varepsilon,c/2}$ and outputs the corresponding likelihood ratio times the indicator $I\left(\tau_{n,\varepsilon,c/2} \leq n, S_n > b\right)$ should behave reasonably well. Of course, we also expect that a slightly better algorithm would be one that applies the change-of-measure whenever we are in the (time/space) region under which importance sampling can be guaranteed to apply – not only up to the first time the process leaves the region. However, as we will see, already the vanilla version algorithm outlined will be shown to be strongly efficient, so we shall just analyze the following simple version of the procedure.

Algorithm 1

Given parameters $b \geq cn^{1/2+\varepsilon}$ c and $\varepsilon > 0$.

Set $s = 0$, $m = 0$, $L = 1$

STEP 1. Sample X from the density $f_{m+1}(\cdot)$ defined via

$$f_{m+1}(x) = f(x) \frac{v(n-m-1, s+x)}{w(n-m, s)}.$$

Actualize

$$\begin{aligned} L &\longleftarrow L \frac{w(n-m, s)}{v(n-m-1, s+X)}, \\ s &\longleftarrow s+X, \\ m &\longleftarrow m+1. \end{aligned}$$

Repeat **STEP 1** until $s > b - 2^{-1} c (n-m)^{\varepsilon}$ or $m = n$.

STEP 2. IF $m = n$ then RETURN $L \cdot I(s > b)$, ELSE generate $n-m$ iid rv's $(X_1, ...X_{n-m})$ according to the density $f(\cdot)$, evaluate $s' = X_1 + ... + X_{n-m}$ and RETURN $L \cdot I(s' + s > b)$. STOP.

It follows that the estimator proposed is given by

$$L = \prod_{k=0}^{\tau_{n,\varepsilon,c/2}-1} \frac{w(n-k, S_k)}{v(n-k-1, S_{k+1})} I\left(\tau_{n,\varepsilon,c/2} \leq n\right) I(S_n > b). \tag{5}$$

The efficiency analysis of L involves estimating

$$\begin{aligned}
&E_0^Q L^2 \\
&= E_0^Q \left(\prod_{k=0}^{\tau_{n,\varepsilon,c/2}-1} \frac{w\left(n-k,S_k\right)^2}{v\left(n-k-1,S_{k+1}\right)^2}; \tau_{n,\varepsilon,c/2} \leq n; S_n > b \right) \\
&= E_0 \left(\prod_{k=0}^{\tau_{n,\varepsilon,c/2}-1} \frac{w\left(n-k,S_k\right)}{v\left(n-k-1,S_{k+1}\right)}; \tau_{n,\varepsilon,c/2} \leq n; S_n > b \right) \\
&= w\left(n,0\right) E_0 \left(\prod_{k=1}^{\tau_{n,\varepsilon,c/2}-1} \frac{w\left(n-k,S_k\right)}{v\left(n-k,S_k\right)} R; \tau_{n,\varepsilon,c/2} \leq n \right),
\end{aligned}$$

where

$$R \equiv \frac{u\left(n-\tau_{n,\varepsilon,c/2},S_{\tau_{n,\varepsilon,c/2}}\right)}{v\left(n-\tau_{n,\varepsilon,c/2},S_{\tau_{n,\varepsilon,c/2}}\right)}.$$

By virtue of Theorem 1, there exists a deterministic constant $\lambda_0 \in (0,\infty)$ (independent of n and b) such that $R \leq \lambda_0$. Consequently, we obtain

$$\begin{aligned}
E_0^Q L^2 \quad \leq \quad & \lambda_0 w\left(n,0\right) P_0\left(\tau_{n,\varepsilon,c/2} \leq n\right) \\
& \cdot E_0 \left(\prod_{k=1}^{\tau_{n,\varepsilon,c/2}-1} \frac{w\left(n-k,S_k\right)}{v\left(n-k,S_k\right)} R \middle| \tau_{n,\varepsilon,c/2} \leq n \right).
\end{aligned}$$

Now, it follows from Proposition 1 that there exists a constant $\lambda_1 \in (1,\infty)$ (independent of n and b) such that

$$P_0\left(\tau_{n,\varepsilon,c/2} \leq n\right) \leq \lambda_1 n \overline{F}\left(b - cn^{1/2}/2\right).$$

Consequently, it is possible to find a constant $\lambda \in (0,\infty)$ (independent of n and b) such that

$$\begin{aligned}
&E_0^Q L^2 \leq \lambda w\left(n,0\right) v\left(n,0\right) \\
&\quad \cdot E_0 \left(\prod_{k=1}^{\tau_{n,\varepsilon,c/2}-1} \frac{w\left(n-k,S_k\right)}{v\left(n-k,S_k\right)} \middle| \tau_{n,\varepsilon,c/2} \leq n \right). \qquad (6)
\end{aligned}$$

The efficiency analysis of our estimator depends on our ability to control the behavior of the local likelihood ratio $w\left(n-k,x\right)/v\left(n-k,x\right)$ on the region $b-x \geq 2^{-1}c\left(n-k\right)^{1/2+\varepsilon}$. The next lemma, which is proven in the appendix, provides the estimates that allow to control the behavior of the local likelihood ratio.

Lemma 1 *Under assumptions i), ii) and iii) we have that there exists a constant $\rho \in (0,\infty)$ such that*

$$\frac{w\left(n-k,x\right)}{v\left(n-k,x\right)} \leq 1+\rho\left[\frac{1}{\left(n-k\right)^{1/2}\cdot\left(b-x\right)} + \overline{F}\left(\sqrt{n-k}\right)\right]$$

uniformly over $\left(b-x\right) > c\left(n-k\right)^{1/2+\varepsilon}$ for $c \in (0,\infty)$ given.

Using the previous lemma, we are ready to state and prove our main result.

Theorem 2 *If assumptions i), ii) and iii) are in force, then the estimator L, provided by **Algorithm 1** (displayed in (5)), is strongly efficient.*

Proof. Equation (6) combined with Lemma 1 implies that

$$E_0^Q L^2 \leq \lambda w\left(n,0\right) v\left(n,0\right) \exp\left(\rho \sum_{k=1}^{n} \left(\frac{1}{k^{1+\varepsilon}} + \frac{1}{k^{\alpha/2}}\right)\right).$$

Since

$$\lim_{n\longrightarrow\infty} \frac{w\left(n,0\right) v\left(n,0\right)}{u\left(n,0\right)^2} = 1$$

and $\sum\left(k^{-(1+\varepsilon)} + k^{-\alpha/2}\right) < \infty$, strong efficiency of the estimator L follows. The proof of the theorem is complete. ■

Finally, we provide the proof of Lemma 1.

Proof of Lemma 1. We shall first introduce some notation. Set $\beta = \left(n-k\right)^{1/2}$, $a\beta = b-x$ and $\beta' = \left(\beta^2-1\right)^{1/2}$. Let us write

$$\widetilde{v}\left(\beta,a\right) = \left[\beta^2 \overline{F}\left(a\beta\right) + \overline{\Phi}\left(a\right)\right] I\left(a>1\right) + I\left(a \leq 1\right),$$

and

$$\begin{aligned}
\widetilde{w}\left(\beta,a\right) \quad &= \quad \left(\beta^2-1\right) E[\overline{F}\left(a\beta - X\right); X \leq a\beta - \beta'] \\
&+ \quad E[\overline{\Phi}\left(a - X/\beta'\right); X \leq a\beta - \beta'] + \overline{F}\left(\left(a-1\right)\beta'\right).
\end{aligned}$$

We have just expressed the functions $v\left(\cdot\right)$ and $w\left(\cdot\right)$ in more convenient terms. In particular, note that $\widetilde{v}\left(\beta,a\right) = v\left(n-k,x\right)$ and $\widetilde{w}\left(\beta,a\right) = w\left(n-k,x\right)$. We will show that

$$\frac{\widetilde{w}\left(\beta,a\right)}{\widetilde{v}\left(\beta,a\right)} = 1 + O\left(\frac{1}{a\beta^2}\right) + O\left(\overline{F}\left(\beta\right)\right).$$

as $\beta \nearrow \infty$. This will imply the conclusion of the lemma. First we estimate the ratio

$$E[\overline{F}\left(a\beta - X\right); X \leq a\beta - \beta']/\overline{F}\left(a\beta\right). \qquad (7)$$

Fix $\delta \in (0,1)$ and note that, by applying a Taylor expansion, using assumption i), and the fact that $EX = 0$, we obtain (as $\beta \nearrow \infty$)

$$\begin{aligned}
&E[\overline{F}\left(a\beta - X\right); |X| \leq \delta a\beta]/\overline{F}\left(a\beta\right) \\
&= \quad 1 + \frac{E[X^2 f'\left(\xi\right); |X| \leq \delta a\beta]}{\overline{F}\left(a\beta\right)} + o\left(a^{-2}\beta^{-2}\right),
\end{aligned}$$

where ξ is a rv such that $a\beta(1-\delta) \leq \xi \leq a\beta(1+\delta)$. The term $o\left(a^{-2}\beta^{-2}\right)$ above is justified because

$$\frac{f(a\beta)\, E[X; X < -\delta a\beta \cup X > \delta a\beta]}{\overline{F}(a\beta)} = o\left(a^{-2}\beta^{-2}\right)$$

as $\beta \nearrow \infty$. Therefore, we can conclude that

$$\begin{aligned} & E[\overline{F}(a\beta - X); |X| \leq \delta a\beta]/\overline{F}(a\beta) \\ = \;& 1 + O\left(\frac{1}{(a\beta)^2}\right). \end{aligned}$$

On the other hand,

$$\begin{aligned} & E[\overline{F}(a\beta - X); X \leq -\delta a\beta]/\overline{F}(a\beta) \\ \leq \;& \overline{F}(a\beta(1+\delta))\, P(X \leq -\delta a\beta)/\overline{F}(a\beta) = o\left(\frac{1}{a\beta^2}\right), \end{aligned}$$

where the previous estimate follows from assumption ii) ($E|X|^p < \infty$ for $p \geq 2$). Finally, note that (again by assumption i))

$$\begin{aligned} & E[\overline{F}(a\beta - X); \delta a\beta \leq X \leq a\beta - \beta']/\overline{F}(a\beta) \\ \leq \;& \frac{\overline{F}(\beta')}{\overline{F}(a\beta)} P(\delta a\beta \leq X \leq a\beta - \beta') = O\left(\overline{F}(\beta)\right). \end{aligned}$$

As a consequence, placing all these estimates together into (7), we obtain that

$$\begin{aligned} & E[\overline{F}(a\beta - X); X \leq a\beta - \beta']/\overline{F}(a\beta) \\ = \;& 1 + o\left(\frac{1}{a\beta^2}\right) + O\left(\overline{F}(\beta)\right) \end{aligned}$$

as $\beta \nearrow \infty$. Our next goal is to estimate the contribution of the terms involving $\overline{\Phi}(\cdot)$, this is done following a similar strategy as in the case of the ratio (7). First of all, observe that (under the assumption that $a \geq c\beta^{\varepsilon}$ for $\varepsilon, c > 0$) there exists $\beta_0 \in (0,\infty)$ such that, for fixed $\delta \in (0,1)$, $\left(\beta^2\overline{F}(a\beta)\right)^2 \geq a\overline{\Phi}(\delta a)$ uniformly over $\beta \geq \beta_0$. Now, combining the mean value theorem and the triangle inequality we obtain

$$\begin{aligned} & E[\overline{\Phi}(a - X/\beta); |X| \leq \delta a\beta]/\beta^2\overline{F}(a\beta) \\ \leq \;& \frac{\overline{\Phi}(a)}{\beta^2\overline{F}(a\beta)} + \frac{a\overline{\Phi}(a(1-\delta))}{\beta^2\overline{F}(a\beta)} \frac{E|X|}{\beta} = o\left(\frac{1}{a\beta^2}\right). \end{aligned}$$

The contribution of the expectations over the sets $\{X \leq -\delta a\beta\}$ and $\{X \in [\delta a\beta, a\beta - \beta']\}$ can be handled analogously. We then obtain

$$E[\overline{\Phi}(a - X/\beta); |X| \leq \delta a\beta]/\beta^2\overline{F}(a\beta) = o\left(\frac{1}{a\beta^2}\right).$$

Let us now put all our estimates together,

$$\begin{aligned} \frac{\widetilde{w}(\beta,a)}{\beta^2\overline{F}(a\beta)} &= \frac{E[\overline{F}(a\beta - X); X \leq a\beta - \beta']}{\overline{F}(a\beta)} \\ &- \frac{E[\overline{F}(a\beta - X); X \leq a\beta - \beta']}{\beta^2\overline{F}(a\beta)} + \frac{\overline{F}(a\beta - \beta')}{\beta^2\overline{F}(a\beta)} \\ &+ \frac{E[\overline{\Phi}(a - X/\beta); X \leq a\beta - \beta']}{\beta^2\overline{F}(a\beta)}. \end{aligned}$$

Our estimates for (7) indicate that the ratio in the first line of the previous display equals $1 + o\left(a^{-1}\beta^{-2}\right) + O\left(\overline{F}(\beta)\right)$; likewise for the integral in the third line of the display. On the other hand, we have that

$$\frac{\overline{F}(a\beta - \beta')}{\beta^2\overline{F}(a\beta)} = \frac{1}{\beta^2}[1 + O\left(\frac{1}{a}\right)].$$

Therefore, again making use of our previous estimates for (7), we conclude

$$\frac{\widetilde{w}(\beta,a)}{\beta^2 F(a\beta)} = 1 + O\left(\frac{1}{a\beta^2}\right) + O\left(\overline{F}(\beta)\right). \tag{8}$$

Finally, since $\widetilde{v}(\beta,a) = \beta^2\overline{F}(a\beta) + \overline{\Phi}(a)$ and (as discussed previously) $\overline{\Phi}(a) = o\left(\beta^4\overline{F}(a\beta)^2\right)$, (8) implies

$$\frac{\widetilde{w}(\beta,a)}{\widetilde{v}(\beta,a)} = 1 + O\left(\frac{1}{a\beta^2}\right) + O\left(\overline{F}(\beta)\right)$$

as $\beta \nearrow \infty$, which completes the proof of the result. ■

REFERENCES

Asmussen, S., K. Binswanger, and B. Hojgaard. 2000. Rare event simulation for heavy-tailed distributions. Bernoulli 6, Vol. 2, 303-322.

Asmussen, S., D.P. Kroese. 2006. Improved algorithms for rare event simulation with heavy tails. Advances in Applied Probability 38 (2), to appear.

Asmussen, S., R.Y. Rubinstein, and C.-L. Wang. 1994. Regenerative Rare Event Simulation via Likelihood Ratios, J. Appl. Prob. 31 (3), 797-815.

Bassamboo, A., S. Juneja, and A. Zeevi. 2005. Importance sampling simulation in the presence of heavy-tails. Proceedings WSC05.

Blanchet, J., and P. Glynn. 2006. Efficient Rare-event Simulation for the Maximum of a Random Walk with Heavy-tailed Increments. Preprint.

Blanchet, J., and P. Glynn. 2006b. Efficient simulation of light-tailed sums: an old-folk song sung to a faster new tune. Preprint.

Borovkov, A.A. 2000. Estimates for the Distribution of Sums and Maxima of Sums of Random Variables without the Cramer Condition. Siberian Math. J. 41, 811-848.

Borovkov, A.A., and K.A. Borovkov. 2001. On Probabilities of Large Deviations for Random Walks. I. Regularly Varying Distribution Tails. Theory Prob. Appl. 49, No. 2, 189-205.

Bucklew, J. 1990. *Large Deviation Techniques in Decision, Simulation, and Estimation.* John Wiley & Sons, New York.

Bucklew, J. 2004. *Introduction to Rare-event Simulation.* Springer, New York.

Del Moral, P. 2004. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Probability and its Applications series, Springer.

Dembo, A., and O. Zeitouni. 1998. *Large Deviations Techniques and Applications*: Second Edition, Springer-Verlag, New-York.

Dieker, A.B., and M. Mandjes. 2005. On asymptotically Efficient Simulation of Large Deviation Probabilities. Adv. in Appl. Probab. 37, no. 2, 539–552

Dupuis, P., K. Leder, and H. Wang. 2006. Importance Sampling for Sums of Random Variables with Regularly Varying Tails. Preprint.

Dupuis, P., and H. Wang. 2004. Importance Sampling, Large Deviations, and Differential Games. Stochastics and Stochastics Reports 76, 481-508.

Heidelberger, P. 1995. Fast Simulation of Rare Events in Queueing and Reliability Models, ACM Transactions on Modeling and Computer Simulation, vol.5, no.1 pp. 43-85.

Juneja, S., and P. Shahabuddin. 2002. Simulating heavy-tailed processes using delayed hazard rate twisting. ACM TOMACS, 12 - 2, p. 94 - 118.

Meyn, S., and R. Tweedie. 1993. *Markov Chains and Stochastic Stability*. Springer-Verlag. <http://decision.csl.uiuc.edu/~meyn/pages/book.html>

Rozovskii, 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 Probab. Appl., 34, pp. 625–644.

Sadowsky, J.S. 1996. On Monte Carlo estimation of large deviations probabilities. Ann. Appl. Prob., 6:399-422.

AUTHOR BIOGRAPHIES

JOSE H. BLANCHET is Assistant Professor of Statistics at Harvard University. Jose holds a M.Sc. in Engineering-Economic Systems and Operations Research and a Ph.D. in Management Science and Engineering, both from Stanford University. He also holds two B.Sc. degrees: one in Actuarial Science and another one in Applied Mathematics from ITAM (Mexico). Jose worked for two years as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability, computational finance, performance engineering, queueing theory, risk management, rare-event analysis, statistical inference, stochastic modelling, and simulation.

JINGCHEN LIU is currently a third year Ph.D. student in the Department of Statistics at Harvard University. Jingchen Liu graduated with BS in Mathematics from Beijing University, China.