Efficient simulation of light-tailed sums: an old-folk song sung to a faster new tune...

Blanchet, J., Leder, K., and Glynn, P.

Abstract

A classic problem in rare-event simulation concerns , namely, efficient estimation of the probability $\alpha_n = P(S_n/n \in A)$ for large n, where S_n is a mean zero d-dimensional random walk with finite moment generating function in a neighborhood of the origin and A is a suitable convex set such that $0 \notin A$. This paper provides the first estimator for α_n for which one can rigorously assert that the number of simulation runs needed to compute α_n to a given relative accuracy remains bounded as a function of n.

1 Introduction

Let $X, X_1, X_2, \ldots$ be a sequence of mean zero iid d-dimensional random vectors. Assume that A is a convex set, having a twice continuously differentiable boundary, for which $0 \notin A$. We revisit a fundamental problem in the theory of rare-event simulation, namely that of computing $\alpha_n = P(S_n/n \in A)$ for n large, where $S_n = X_1 + X_2 + \ldots + X_n$. In particular, we consider the setting in which X is light-tailed. The purpose of this paper is to provide the first simulation estimator for which the number of simulation runs needed to compute α_n to a given relative accuracy remains bounded as a function of the parameter n.

To fix ideas, consider the one dimensional case in which $A = (\beta, \infty)$ for $\beta > 0$. It is well known that the use of importance sampling, as implemented through optimal exponential tilting (OET), provides an estimator that is "logarithmically efficient" as $n \nearrow \infty$ (in the sense that the squared coefficient of variation grows subexponentially); [8]. Recall that OET involves an importance distribution in which each of the summands is independently sampled from that member of the natural exponential family having mean β.In fact, it can be further shown that OET provides the only iid importance sampling algorithm that achieves logarithmic efficiency (see [2]). This is not surprising, given that OET agrees with the conditional distribution of the S_k's ($k < n$) given $\{S_n > n\beta\}$ in a weak convergence sense as $n \nearrow \infty$.

However, it turns out that the squared coefficient of variation associated with OET does increase as $n \nearrow \infty$, so that the number of samples required to compute α_n to a given relative accuracy increases as a function of n. In fact, in Section 2, we prove that under mild conditions, the squared coefficient of variation grows at rate $O\left(n^{1/2}\right)$. The reason that OET becomes less efficient with the growth of n has to do with the fact that OET fails to agree with the conditional distribution at scales finer than that of the LLN. As one illustration of this phenomenon, Proposition 3 establishes that the "overshoot" $S_n - n\beta > 0$ is asymptotically exponentially distributed under the conditional distribution, whereas the OET produces an overshoot of order $n^{1/2}$. In other words, the OET tends to bias the increments excessively when the random walk is relatively close to reaching the boundary $n\beta$, thereby inducing a large overshoot.

An algorithm having the property that the sample size required to compute α_n to a given relative accuracy is bounded as a function of n is called a "strongly efficient" algorithm; see [2] and [7]. To produce a strongly efficient estimator, we introduce the concept of "optimal state-dependent exponential tilting" (OSDET) – closely related to the solution to the associated Isaacs equation introduced in [5]. In the one dimensional case, this corresponds to dynamically updating the OET at each step in the algorithm based on the current position S_k of the random walk $(S_k : k \geq 0)$. Our proof of strong efficiency, which relies on the analysis of several martingales that arise naturally from the description of the algorithm, is of independent interest. Not surprisingly, it turns out that OSDET induces a bounded overshoot as $n \nearrow \infty$.

A different approach based on direct approximation of the zero-variance change-of-measure is discussed in [4] for Gaussian random walks (we mention that a subtle technical error is present in the proof of Theorem 6 in [4]).

The rest of the paper is organized as follows. Sections 2 and 3 concentrate on the one dimensional case. Section 2 describes explicitly our assumptions and collects some needed results from the theory of large deviations. Section 3 analyzes the proposed algorithm in the one dimensional case and shows that the overshoot under OSDET remains bounded as $n \nearrow \infty$. Section 4 treats the multidimensional case and Section 5 includes a numerical experiment.

2 Large Deviations Results for Light Tailed Sums

In this section, we concentrate on the one dimensional case and present some auxiliary results from the theory of large deviations that will be useful for the description and analysis of our algorithm.

We start with listing the assumptions that will underlie our development in Sections 2 and 3.

i) $EX = 0$ and $Var\left(X\right) = \sigma^2$

ii) The logmoment generating function $\left(\psi\left(\theta\right) : \theta \in \mathbb{R}\right)$, defined as $\psi\left(\theta\right) = \log E \exp\left(\theta X\right)$, is assumed to be *steep* on the right. That is, for all $w > 0$ there exists $\theta_w > 0$ such that $\psi'(\theta_w) = w$ (by strict convexity the solution θ_w is unique).

iii) The rv X is nonlattice (i.e. the characteristic function has norm strictly less than one except at the origin).

We now are ready to describe some results from large deviations that will be useful in our development. The so-called rate function, $J(\cdot)$, plays a crucial role in the theory of large deviations. In our context, $J(\cdot)$ is defined (for $w \in \mathbb{R}$) via

$$J(w) \triangleq \max_{\theta \geq 0}[\theta w - \psi(\theta)]. \tag{1}$$

Assuming that $w \geq 0$ we have

$$J(w) \triangleq w\theta_w - \psi(\theta_w),$$

whereas $J(w) = 0$ for $w \leq 0$. Finally, the natural exponential family $(F_\theta : \theta \in \mathbb{R})$ generated by the distribution $F(\cdot) = P(X \leq \cdot)$ is defined via

$$dF_\theta = \exp(\theta x - \psi(\theta))\, dF. \tag{2}$$

The distribution F_θ is also said to be "exponentially tilted" by the parameter θ.Let $P_\theta(\cdot)$ be the product probability measure generated by F_θ (for $\theta \in \mathbb{R}$) under which the X_i's are iid and let $E_\theta(\cdot)$ be the corresponding expectation operator associated with $P_\theta(\cdot)$.

We shall need the following elementary properties of the rate function.

Proposition 1 *If assumption ii) is in force, then*

$$J(w) = \sigma^2 w^2/2 + O\left(w^3\right)$$

as $w \searrow 0$. *Moreover, for each* $w \in (0, \infty)$ *we have*

$$J(w + h) = J(w) + \theta_w h + O\left(h^2\right)$$

as $h \longrightarrow 0$ *(uniformly over* $w \in [\varepsilon, 1/\varepsilon]$ *for* $\varepsilon > 0$*)so that* $J'(w) = \theta_w$. *In fact, the rate function* $(J(w) : w \geq 0)$ *is infinitely differentiable at zero and its Taylor series converges in a neighborhood of the origin; similarly for* $J(w + \cdot)$.

Large deviations theory is intended to both address the question of how to compute asymptotics for rare event probabilities and to describe the conditional behavior of the underlying system given the occurrence of the rare event. The following result is a celebrated large deviations asymptotic approximation due to Bahadur and Rao (see [3]) that will be useful in our development.

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

$$P(S_n > n\beta) = \frac{\exp(-nJ(\beta))}{n^{1/2}(2\pi)^{1/2}\psi''(\beta)^{1/2}\theta_\beta}(1+o(1))$$

as $n \nearrow \infty$ *for fixed* $\beta > 0$.

The following theorem provides an asymptotic description of the conditional behavior of the process $(S_k : 0 \leq k \leq n)$ given that $S_n > \beta n$ (as $n \nearrow \infty$) and provides rigorous support for the claim that the asymptotic conditional distribution of the increments given $\{S_n > b\}$ is $P_{\theta_\beta}(\cdot)$.

Proposition 2 *Suppose that i) to iii) are in force. Then, any positive integers* $k_1 < k_2 < ... < k_m < \infty$ *we have that when each* $x_{k_i}, 1 \leq i \leq m$, *is a continuity point of* $F(\cdot)$

$$P(X_{k_1} \leq x_{k_1}, ..., X_{k_m} \leq k_m | S_n > n\beta) \longrightarrow F_{\theta_\beta}(x_{k_1}) ... F_{\theta_\beta}(x_{k_m}),$$

as $n \nearrow \infty$.

While the above result describes the behavior of a typical increment under the conditioning, the proposition below provides an asymptotic description of the limiting overshoot $S_n - n\beta > 0$.

Proposition 3 *Assume that i), ii) and iv) hold and put* $b = \beta n$ *for* $\beta > 0$. *Then, for all* $x > 0$

$$P(S_n - b > x | S_n > b) \longrightarrow \exp(-\theta_\beta x)$$

as $n \nearrow \infty$, *where* $\psi'(\theta_\beta) = \beta$.

Proof. Note that by introducing the change-of-measure

$$dF_{\theta_\beta}(x) = \exp(\theta_\beta x - \psi(\theta_\beta))\, dF,$$

(recall that $\psi'(\theta_\beta) = E_{\theta_\beta} X = \beta$ and $\psi''(\theta_\beta) = Var_{\theta_\beta}(X)$) we can write

$$\begin{aligned} & P(S_n - b > x | S_n > b) \\ & = \frac{P(S_n - b > x)}{P(S_n > b)} = \frac{E_{\theta_\beta}[\exp(-\theta_\beta(S_n - b)); S_n - b > x]}{E_{\theta_\beta}[\exp(-\theta_\beta(S_n - b)); S_n - b > 0]}. \end{aligned}$$

Applying the local CLT as in the classical proof of Theorem 1 (see, for instance, [1]) to each of the expectations in the previous ratio we obtain that as $n \nearrow \infty$

$$\frac{E_{\theta_\beta}[\exp(-\theta_\beta(S_n - b)); S_n - b > x]}{E_{\theta_\beta}[\exp(-\theta_\beta(S_n - b)); S_n - b > 0]} \longrightarrow \exp(-\theta_\beta x),$$

yielding the result. ■

The previous result makes clear that $P_{\theta_\beta}(\cdot)$ does not accurately describe the behavior of the random walk, conditioned on $\{S_n > n\beta\}$, at time n. Under $P_{\theta_\beta}(\cdot)$, the CLT implies that $(S_n - n\beta) \overset{D}{\approx} n^{1/2} N(0, \psi''(\beta))$ (where the symbol "$\overset{D}{\approx}$" is non-rigorous and is intended to suggest weak convergence of $n^{-1/2}(S_n - n\beta)$). On the other hand, Proposition 3 indicates that the conditional overshoot is of order $O(1)$ (in distribution). Therefore, $P_{\theta_\beta}(\cdot)$ may provide a poor description of the conditional distribution of the random walk at scales that are finer than linear (for instance at scales of order $n^{1/2}$). As a consequence, it is not surprising that the performance of $P_{\theta_\beta}(\cdot)$ as an importance sampling distribution degrades when measured at a fine enough scale. In particular, the estimator induced by $P_{\theta_\beta}(\cdot)$, namely

$$L = \exp(-\theta_\beta S_n + \psi(\theta_\beta)) I(S_n > \beta n), \tag{3}$$

is not strongly efficient (i.e. the squared coefficient of variation of the estimator is unbounded as $n \nearrow \infty$). More precisely, it follows that if $cv_n(L)$ denotes the coefficient of variation of L, then (under assumptions i) to iii))

$$cv_n(L)^2 \triangleq \frac{Var_{\theta_\beta}(L)}{E_{\theta_\beta}(L)^2} = \frac{Var_{\theta_\beta}(L)}{P(S_n > b)^2} = O\left(n^{1/2}\right) \tag{4}$$

as $n \nearrow \infty$. In our next section, we examine the form of the optimal change-of-measure and propose an importance sampling distribution that improves upon P_{θ_β} by achieving a bounded squared coefficient of variation.

3 A Proposed Algorithm and Efficiency Analysis

Let us start by describing the algorithm explicitly.

Algorithm 1

Set $w = \beta > 0$, $L = 1$, $s = 0$, $s' = 0$, $k = 0$, and $\lambda = 2\beta$

WHILE $w > 0$ AND $w \leq \lambda$

Sample X from $F_{J'(w)}$ and set

$$\begin{aligned} L &\longleftarrow \exp(-J'(w) X + \psi(J'(w))) L, \\ s &\longleftarrow s + X, \\ k &\longleftarrow k + 1, \\ w &\longleftarrow (n\beta - s)/(n - k). \end{aligned}$$

LOOP

Sample $X_1, ..., X_{n-k}$ iid rv's from $F_{J'(w)}$ and set

$$s' \longleftarrow X_1 + ... + X_{n-k},$$
$$L \longleftarrow \exp\left(-J'(w)\, s' + \psi\left(J'(w)\right)\right) L.$$

OUTPUT $Y_n = L \times I\left(s + s' > n\beta\right)$

END

The estimator Y_n obtained from the previous algorithm can be expressed as follows. First, define

$$W_j = (n\beta - S_j)/(n - j) \tag{5}$$

for $0 \leq j \leq n-1$. Next define the following stopping times

$$\begin{aligned} \tau_1^{(n)} &= \inf\{0 \leq k < n : W_k > \lambda\}, \\ \tau_0^{(n)} &= \inf\{k \geq 0 : n\beta - S_k \leq 0\}, \\ \tau^{(n)} &= \tau_0^{(n)} \wedge \tau_1^{(n)} \wedge n. \end{aligned} \tag{6}$$

Let $\theta_j = J'(W_j)$. Then we can define the OSDET (optimal state-dependent exponential tilting) estimator as follows

$$\begin{aligned} Y_n = \exp & \left(- \sum_{j=0}^{\tau^{(n)}-1} \left(\theta_j X_{j+1} - \psi\left(\theta_j\right)\right) \right) \\ & \times \exp\left(-\theta_{\tau^{(n)}}\left(S_n - S_{\tau^{(n)}}\right) + (n - \tau^{(n)})\psi\left(\theta_{\tau^{(n)}}\right)\right) \\ & \times I\left(S_n > n\beta\right), \end{aligned} \tag{7}$$

where X_j follows the distribution F_{θ_j} for $1 \leq j \leq \tau^{(n)}$ and X_j is sampled according to the distribution $F_{\theta_{\tau^{(n)}}}$ for $\tau + 1 \leq j \leq n$ (note that if $\tau_0^{(n)} < \tau_1^{(n)} < n$ then the remaining $n - \tau_0^{(n)}$ increments are sampled according to the original/nominal distribution).

In simple words, the algorithm recomputes the OET at every step up until the boundary is reached or when the random walk is far away from the boundary (i.e. the ratio between the distance to the boundary and the remaining time is sufficiently large). If we have not reached the time horizon, then we just simulate the remaining increments using a standard (state-independent) OET corresponding to the current position.

The rest of the section is devoted to the proof of the following result. Let us write $\widetilde{E}(\cdot)$ to denote the expectation operator induced by the importance sampling algorithm described in Algorithm 1.

Theorem 2 *For each $p > 1$,*

$$\sup_{n \geq 1} \frac{\widetilde{E} Y_n^p}{P\left(S_n > n\beta\right)^p} < \infty$$

as $n \nearrow \infty$.

Remark: In [4] an estimator is proposed for Gaussian random walks based on a direct approximation of the zero-variance change-of-measure by means of sharp large deviations results. In order to simplify their analysis the authors proposed a change-of-measure which is applied all the time up to n (even if the level $n\beta$ is reached before time n, i.e. $\tau_0^{(n)} < n$). The proof of Theorem 6 in [4] requires verifying Proposition 5 for $k = n-1$ as well (and not only $k \leq n-2$ as it is done there). A similar efficiency analysis as the one that follows below can be done by turning off importance sampling at time $\tau_0^{(n)}$. Time $\tau_1^{(n)}$ is not really necessary for Gaussian random walks. One then obtains strong efficiency for the proposed scheme.

To prove the Theorem 2, we first recognize that $\tau^{(n)}$is likely to occur within $O(1)$ time of n under the importance sampling distribution $\widetilde{P}$ associated with Algorithm 1, so that the key element in the exponent of the likelihood ratio is then the term $-\sum_{j=0}^{\tau^{(n)}-1}\left(\theta_j X_{j+1} - \psi\left(\theta_j\right)\right)$. This term in the exponent can then be simplified by first noting that

$$\theta_j X_{j+1} - \psi\left(\theta_j\right) = J'\left(W_j\right) X_{j+1} - \psi\left(W_j\right)$$

where $W_j = (n\beta - S_j)/(n-j)$. Furthermore,

$$\begin{aligned} J'\left(W_j\right) X_{j+1} - \psi\left(W_j\right) &= (n-j-1)J(W_{j+1}) - (n-j)J\left(W_j\right) \\ &\quad - \frac{1}{2}\frac{J''\left(W_j + \eta_{j+1}\left(X_{j+1}\right)\right)}{(n-j)}\left(X_{j+1} - W_j\right)^2, \end{aligned}$$

where $\left|\eta_{j+1}\left(X_{j+1}\right)\right| \leq \left|W_{j+1} - W_j\right|$. Summing the above expression over j, the "collapsing sum" on the right-hand side then yields

$$\begin{aligned} -\sum_{j=0}^{\tau^{(n)}-1}\left(\theta_j X_{j+1} - \psi\left(\theta_j\right)\right) &= -nJ\left(\beta\right) + (n - \tau^{(n)})J\left(W_{\tau^{(n)}}\right) \\ &\quad + \sum_{j=0}^{\tau^{(n)}-1}\frac{J''\left(W_j + \eta_{j+1}\left(X_{j+1}\right)\right)}{2\left(n-j\right)}\left(X_{j+1} - W_j\right)^2. \end{aligned}$$

To proceed further, we now require the following result.

Lemma 1 *Let $w_j = (n\beta - s)/(n-j)$.*

$$J''(w_j)^{-1} = E_{J'(w_j)}(X_{j+1} - w_j)^2 = Var_{J'(w_j)}(X_{j+1}).$$

Consequently, assuming that $|w_j| \le \lambda$,then there exists a constant $c(\lambda) \in (0,\infty)$ such that

$$\left| E_{J'(w_j)} \frac{J''(w_j + \eta_{j+1}(X_{j+1}))}{(n-j)} (X_{j+1} - w_j)^2 - \frac{1}{n-j} \right| \le \frac{c(\lambda)}{(n-j)^2}.$$

Proof. The expression for $J''(w_j)$ follows from 1. For the second part we note that

$$W_{j+1} = w_j + \frac{w_j}{n-j-1} - \frac{X_{j+1}}{n-j-1},$$

therefore,

$$|\eta_{j+1}(X_{j+1})| \le |W_{j+1} - w_j| = \left| \frac{w_j}{n-j-1} - \frac{X_{j+1}}{n-j-1} \right|.$$

On the other hand,

$$J''(w_j + \eta_{j+1}(X_{j+1})) - J''(w_j) = \int_0^{\eta_{j+1}(X_{j+1})} J'(w_j + s)\,ds,$$

which yields,

$$\begin{aligned} &|J''(w_j + \eta_{j+1}(X_{j+1})) - J''(w_j)| \\ &\le |\eta_{j+1}(X_{j+1})| (|w_j| + |\eta_{j+1}(X_{j+1})|) \le \frac{(2\lambda + |X_{j+1}|)(|X_{j+1}| + \lambda)}{n-j}. \end{aligned}$$

We conclude that

$$\begin{aligned} &\left| E_{J'(w_j)} \frac{J''(w_j + \eta_{j+1}(X_{j+1}))}{(n-j)} (X_{j+1} - w_j)^2 - \frac{1}{n-j} \right| \\ &= \left| E_{J'(w_j)} \frac{J''(w_j + \eta_{j+1}(X_{j+1})) - J''(w_j)}{(n-j)} (X_{j+1} - w_j)^2 \right| \\ &\le \left| E_{J'(w_j)} \frac{(2\lambda + |X_{j+1}|)(|X_{j+1}| + \lambda)}{(n-j)^2} (X_{j+1} - w_j)^2 \right| \le \frac{c(\lambda)}{(n-j)^2}, \end{aligned}$$

where the previous inequality follows by steepness and because $w_j \le \lambda$. ■

With this result in hand, define

$$D_{j+1}(X_{j+1}, w_j) = J''(w_j + \eta_{j+1}(X_{j+1}))(x - w_j)^2$$

and put

$$d_{j+1}(w_j) = E_{J'(w_j)}(D_{j+1}(X_{j+1}, w_j)).$$

Finally, write

$$\overline{D}_{j+1}\left(X_{j+1}, w_j\right) = D_{j+1}\left(X_{j+1}, w_j\right) - d_{j+1}\left(w_j\right)$$

and note that the $\overline{D}_{j+1}\left(X_{j+1}, W_j\right)$'s form a sequence of martingale differences. Therefore,

$$\begin{aligned}
& - \sum_{j=0}^{\tau^{(n)}-1} \left(\theta_j X_{j+1} - \psi\left(\theta_j\right)\right) \\
& = -nJ\left(\beta\right) + (n - \tau^{(n)}) J\left(W_{\tau \wedge n}\right) - \sum_{j=0}^{\tau^{(n)}-1} \frac{\overline{D}_{j+1}\left(X_{j+1}, W_j\right)}{2(n-j)} \\
& - \sum_{j=0}^{\tau^{(n)}-1} \left(\frac{d_{j+1}\left(W_j\right)}{2(n-j)} - \frac{1}{2(n-j)} \right) \\
& - \sum_{j=0}^{\tau^{(n)}-1} \frac{1}{2(n-j)}.
\end{aligned}$$

Once again, using Lemma 1, since $|W_j| \leq \lambda$ on $j < \tau$ we obtain that

$$\left| \sum_{j=0}^{\tau^{(n)}-1} \left(\frac{d_{j+1}\left(W_j\right)}{2(n-j)} - \frac{1}{2(n-j)} \right) \right| \leq c\left(\lambda\right) \sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{c\left(\lambda\right)\pi^2}{6} < \infty.$$

On the other hand, we have that

$$\begin{aligned}
& - \theta_{\tau^{(n)}}\left(S_n - S_{\tau^{(n)}}\right) + (n - \tau^{(n)}) \psi\left(\theta_{\tau^{(n)}}\right) \\
& = -(n - \tau^{(n)}) J\left(W_{\tau^{(n)}}\right) - \theta_{\tau^{(n)}}\left(S_n - n\beta\right)
\end{aligned}$$

and therefore, combining all of the previous calculations, we arrive at the following proposition.

Proposition 4 *There exists a constant* $m\left(\lambda\right) \in (0, \infty)$

$$\begin{aligned}
Y_n^p & \leq m\left(\lambda\right) \exp\left(-pnJ\left(0\right) - \sum_{j=0}^{\tau^{(n)}-1} \frac{p\overline{D}_{j+1}\left(X_{j+1}, W_j\right)}{2(n-j)} \right) \\
& \times \frac{\left(n - \tau^{(n)} + 1\right)^{p/2}}{n^{p/2}} \exp\left(-p\theta_{\tau^{(n)}}\left(S_n - n\beta\right)\right) I\left(S_n > n\beta\right) \\
& \leq m\left(\lambda\right) \frac{\exp\left(-pnJ\left(0\right)\right)}{n^{p/2}} \left(n - \tau^{(n)} + 1\right)^{p/2} \\
& \times \exp\left(- \sum_{j=0}^{\tau^{(n)}-1} \frac{p\overline{D}_{j+1}\left(X_{j+1}, W_j\right)}{2(n-j)} \right).
\end{aligned}$$

Let us use $\widetilde{E}(\cdot)$ to denote the change-of-measure induced by Algorithm 1. The previous proposition indicates that

$$\widetilde{E}Y_n^2 \le \frac{m(\lambda)\exp(-pnJ(0))}{n^{p/2}} \times \widetilde{E}\left(\exp\left(-\sum_{j=0}^{\tau^{(n)}-1}\frac{p\overline{D}_{j+1}(X_{j+1},W_j)}{2(n-j)}\right)\left(n-\tau^{(n)}+1\right)\right).$$

Using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}&\widetilde{E}\left(\exp\left(-\sum_{j=0}^{\tau^{(n)}-1}\frac{p\overline{D}_{j+1}(X_{j+1},W_j)}{2(n-j)}\right)\left(n-\tau^{(n)}+1\right)\right)\\ &\le \widetilde{E}\left(\exp\left(-\sum_{j=0}^{\tau^{(n)}-1}\frac{p\overline{D}_{j+1}(X_{j+1},W_j)}{(n-j)}\right)\right)^{1/2}\\ &\times \widetilde{E}\left(\left(n-\tau^{(n)}+1\right)^p\right)^{1/2}.\end{aligned}$$

In order to verify strong efficiency of the algorithm we must show that

$$\begin{aligned}\widetilde{E}\left(\exp\left(-\sum_{j=0}^{\tau^{(n)}-1}\frac{p\overline{D}_{j+1}(X_{j+1},W_j)}{2(n-j)}\right)\right) &= O(1),\\ \widetilde{E}\left(\left(n-\tau^{(n)}+1\right)^p\right) &= O(1)\end{aligned}$$

as $n\nearrow\infty$. Observe that the latter $O(1)$ statement asserts, as suggested earlier, that $\tau^{(n)}$ does indeed occur within $O(1)$ time of n.

We first establish the required property for $\widetilde{E}\left(\left(n-\tau^{(n)}+1\right)^p\right)$.

Proposition 5 *For any $p\in(1,\infty)$ we have that*

$$\sup_{n\ge1}\widetilde{E}\left(\left(n-\tau^{(n)}\right)^p\right)<\infty.$$

Proof. Clearly,

$$\widetilde{E}\left(\left(n-\tau^{(n)}\right)^p\right)=\sum_{k=1}^{n-1}(n-k)^p\widetilde{P}\left(\tau^{(n)}=k\right)$$

and

$$\widetilde{P}\left(\tau^{(n)}=k\right)\le\widetilde{P}\left(\tau^{(n)}>k-1,W_k\ge\lambda\right)+\widetilde{P}\left(\tau^{(n)}>k-1,W_k\le 0\right).$$

Now, define the martingale difference $D_j' = (W_j - W_{j-1}) \times I\left(\tau^{(n)} > j-1\right)$ (for $1 \leq j \leq n-1$) and note that

$$\widetilde{P}\left(\tau^{(n)} > k-1, W_k \geq \lambda\right) = \widetilde{P}\left(\tau^{(n)} > k-1, \sum_{j=1}^{k} D_j' \geq \lambda - \beta\right),$$

$$\widetilde{P}\left(\tau^{(n)} > k-1, W_k \leq 0\right) = \widetilde{P}\left(\tau^{(n)} > k-1, \sum_{j=1}^{k} D_j' \leq -\beta\right).$$

We will show that there exist constant $m_1 \in (0, \infty)$ such that

$$\widetilde{P}\left(\sum_{j=1}^{k} D_j' \geq \lambda - \beta\right) \leq m_1 \exp\left(-(n-k)^{1/3}\right). \tag{8}$$

To see this, note that given $W_{k-1} = w_{k-1}$ we can write

$$D_k' = \left(\frac{w_{k-1} - X_k}{n-k}\right) \times I\left(\tau^{(n)} > k-1\right).$$

Thus, if $\eta \in (0, \infty)$ then

$$\widetilde{E}\left(\exp\left(\eta D_k'\right) \middle| D_1', ..., D_{k-1}'\right) = \exp\left(\chi_k\left(\frac{\eta}{n-k}\right)\right),$$

where

$$\begin{aligned}\chi_k\left(\frac{\eta}{n-k}\right) &= \frac{\eta w_{k-1} I\left(\tau^{(n)} > k-1\right)}{n-k} \\ &\quad + \psi\left(\frac{-\eta I\left(\tau^{(n)} > k-1\right)}{n-k} + \theta_{k-1}\right) \\ &\quad - \psi\left(\theta_{k-1}\right).\end{aligned}$$

If $\eta = (n-k)^{1/3}$, then because $\tau > k-1$ and $\psi'\left(\theta_{k-1}\right) = w_{k-1}$, there exists a constant $m_2(\lambda) \in (0, \infty)$ such that

$$\chi_k\left(\frac{\eta}{n-k}\right) \leq \frac{m_2(\lambda)}{(n-k)^{4/3}}.$$

Applying the previous considerations subsequently for $j = k-1, k-2, ..., 1$ we obtain that

$$\begin{aligned}\widetilde{E}\left(\exp\left((n-k)^{1/3} \sum_{j=1}^{k} D_j'\right)\right) &\leq \exp\left(\sum_{j=1}^{k} \frac{m_2(\lambda)}{(n-j)^{4/3}}\right) \\ &\leq \exp\left(\sum_{j=1}^{\infty} \frac{m_2(\lambda)}{j^{4/3}}\right) = m_1.\end{aligned}$$

Chebyshev's inequality then yields inequality (8) as indicated. A completely analogous estimate can be obtained for

$$\widetilde{P}\left(\sum_{j=1}^{k} D_j' \leq -\beta\right)$$

and therefore we conclude that

$$\begin{aligned}\widetilde{E}\left(\left(n-\tau^{(n)}\right)^p\right) &= \sum_{k=1}^{n-1}(n-k)^p \widetilde{P}\left(\tau^{(n)}=k\right) \\ &\leq 2\sum_{k=1}^{n-1}(n-k)^p m_1 \exp\left(-(n-k)^{1/3}\right) = O\left(1\right)\end{aligned}$$

as $n \nearrow \infty$. ■

Finally, we turn to the remaining result required to establish strong efficiency, namely,

Proposition 6 *For each $\eta > 0$ we have that*

$$\sup_{n\geq 1} \widetilde{E}\left(\exp\left(-\sum_{j=0}^{\tau^{(n)}-1} \frac{\eta \overline{D}_{j+1}\left(X_{j+1}, W_j\right)}{(n-j)}\right)\right) < \infty$$

Proof. Define $\overline{D}_{j+1}' = \overline{D}_{j+1}\left(X_{j+1}, W_j\right) I\left(\tau^{(n)} > j\right)$and note that for any $\eta > 0$ we have

$$-\sum_{j=0}^{\tau^{(n)}-1} \frac{\eta \overline{D}_{j+1}\left(X_{j+1}, W_j\right)}{(n-j)} = -\sum_{j=0}^{n-1} \frac{\eta \overline{D}_{j+1}'}{(n-j)}.$$

Now, we have that $W_j = w_j$ and $\tau^{(n)} > j$

$$\widetilde{E}\left(\exp\left(-\frac{\eta \overline{D}_{j+1}'}{(n-j)}\right)\middle| X_1, ..., X_j\right) = \exp\left(\xi_j\left(\frac{-\eta I\left(\tau^{(n)} > j\right)}{n-j}, w_j\right)\right),$$

where

$$\xi_j\left(\theta, w_j\right) = \log \widetilde{E} \exp\left(\theta \overline{D}_{j+1}\left(X_{j+1}, w_j\right)\right).$$

It is important to observe that $\xi_j\left(\theta, w_j\right) < \infty$ for $\theta < 0$. Moreover, we have that $\xi_j'\left(0, w_j\right) = 0$ and therefore on $\tau^{(n)} > j$ (which implies that $0 \leq w_j \leq \lambda$) there exists $m_3\left(\lambda\right) \in (0, \infty)$ such that

$$\xi_j\left(\frac{-\eta}{n-j}\right) \leq \frac{\eta^2}{2\left(n-j\right)^2} \sup_{-\eta/(n-j)\leq\theta\leq 0} \xi_j''\left(\theta\right) \leq \frac{\eta^2 m_3\left(\lambda\right)}{2\left(n-j\right)^2}.$$

Iterating the previous calculations for $j = n-1, n-2, ..., 0$ we obtain that

$$\widetilde{E}\left(\exp\left(-\sum_{j=0}^{\tau^{(n)}-1} \frac{\eta \overline{D}_{j+1}\left(X_{j+1}, W_j\right)}{(n-j)}\right)\right) \leq \exp\left(\sum_{j=1}^{n-1} \frac{\eta^2 m_3\left(\lambda\right)}{2\left(n-j\right)^2}\right) = O\left(1\right)$$

as $n \nearrow \infty$, which yields the result. ■

In the Introduction we emphasized the distinction between OET and the zero-variance change-of-measure in the sense that the overshoot is controlled as $n \nearrow \infty$ under the zero-variance change-of-measure but grows under OET. As the next result shows, under our sampler the overshoot stays bounded in expectation.

Proposition 7

$$\sup_{n\geq 1} \widetilde{E}\left(|S_n - n\beta|\right) < \infty.$$

Proof. We first write

$$\widetilde{E}\,|S_n - n\beta| \leq \widetilde{E}|S_n - S_{\tau^{(n)}}| + \widetilde{E}|S_{\tau^{(n)}} - n\beta|.$$

We analyze the latter term first, therefore note that

$$|S_n - S_{\tau^{(n)}}| \leq \max\left[\lambda(n-\tau^{(n)}), |X_{\tau^{(n)}}| + \lambda\right].$$

Then look at the expected value of $X_{\tau^{(n)}}$ based on the value of $\tau^{(n)}$. In particular, we have that for each $p \geq 1$

$$\begin{aligned}\widetilde{E}|X_{\tau^{(n)}}|^p &= \sum_{j=0}^{n-1} \widetilde{E}\left(|X_{j+1}|^p; \tau^{(n)} = j+1, \tau^{(n)} > j\right)\\ &\leq \sum_{j=0}^{n-1} \widetilde{E}\left(|X_{j+1}|^{2p}; \tau^{(n)} > j\right)^{1/2} \widetilde{P}\left(\tau^{(n)} = j+1\right)^{1/2}.\end{aligned}$$

It follows from steepness and the fact that $\tau^{(n)} > j$ that there exists a constant $c_0(\lambda) \in (0,\infty)$ such that

$$\widetilde{E}\left(|X_{j+1}|^{2p} \middle|\, \tau^{(n)} > j\right) \leq c_0(\lambda).$$

As we established in the proof of Proposition 5, it follows that

$$\widetilde{P}\left(\tau^{(n)} = j+1\right) \leq \widetilde{P}\left(\tau^{(n)} \leq j+1\right) \leq m_1 \exp\left(-(n-j)^{1/3}\right)$$

for some constant $m_1 \in (0,\infty)$. Therefore, we obtain that

$$\sup_{n\geq 1} \widetilde{E}|X_{\tau^{(n)}}|^p \leq c_0(\lambda)^{1/2} m_1^{1/2} \sum_{j=1}^{\infty} \exp\left(-j^{1/3}/2\right) < \infty \tag{9}$$

and therefore,

$$\sup_{n\geq 1} E\left[|S_{\tau^{(n)}} - n\beta|\right] < \infty.$$

The proof will be completed once we show that $\widetilde{E}\left[|S_n - S_{\tau^{(n)}}|\right]$ stays bounded with n. First, note that

$$\widetilde{E}\left(|S_n - S_{\tau^{(n)}}|; \tau_2^{(n)} \leq \tau_1^{(n)}\right) \leq (E|X_1|)\,\widetilde{E}(n - \tau_2^{(n)}),$$

(observe $E|X_1|$ appears because from time $\tau_2^{(n)}+1$ up to n the sampling is done under the original / nominal distribution). So, it suffices to consider

$$\widetilde{E}\left(|S_n - S_{\tau^{(n)}}|; \tau_2^{(n)} > \tau_1^{(n)}\right).$$

Let us define

$$\mu(W_{\tau_1^{(n)}}) = \widetilde{E}\left(|X_j| \,\middle|\mathcal{F}_{\tau_1^{(n)}}\right).$$

Using the triangle inequality, conditioning and Cauchy-Schwarz we get the following,

$$\begin{aligned}
&\widetilde{E}\left(|S_n - S_{\tau^{(n)}}|; \tau_1^{(n)} < \tau_2^{(n)}\right) \\
&\leq \widetilde{E}\left(\sum_{j=\tau_1^{(n)}+1}^{n} |X_j|\,; \tau_1^{(n)} < \tau_2^{(n)}, \tau_1^{(n)} = k\right) \\
&\leq \sum_{k=1}^{n-1}\sum_{j=k+1}^{n} \widetilde{E}\left(|X_j|\,; \tau_1^{(n)} = k\right) \\
&\leq \sum_{k=1}^{n-1}(n-k)\widetilde{E}\left(\mu(W_{\tau_1^{(n)}}); \tau_1^{(n)} = k\right) \\
&\leq \widetilde{E}\left(\mu(W_{\tau_1^{(n)}})^2\right)^{1/2}\sum_{k=1}^{n-1}(n-k)\left(\widetilde{P}\left(\tau_1^{(n)} = k\right)\right)^{1/2}.
\end{aligned}$$

As we noted before, from the proof of Proposition 5, it follows that

$$\sup_{n\geq 1}\sum_{k=1}^{n-1}(n-k)\left(\widetilde{P}\left(\tau_1^{(n)} = k\right)\right)^{1/2} < \infty.$$

Therefore, it remains to show that $\widetilde{E}\left(\mu(W_{\tau_1^{(n)}})^2\right)$ stays bounded as n goes to infinity. Notice that

$$0 \leq W_{\tau_1^{(n)}} \leq \lambda + 1 + \left|X_{\tau_1^{(n)}}\right|.$$

A similar analysis behind equation (9) then allows us to conclude

$$\sup_{n\geq 1}\widetilde{E}\left(\left(W_{\tau_1^{(n)}}\right)^p\right) < \infty.$$

Finally, observe that

$$\begin{aligned}
\mu(W_{\tau_1^{(n)}})^2 &\leq \widetilde{E}\left(X_j^2|\mathcal{F}_{\tau_1^{(n)}}\right) \\
&= \psi''\left(\theta_{\tau_1^{(n)}}\right) + \left(\psi'\left(\theta_{\tau_1^{(n)}}\right)\right)^2.
\end{aligned}$$

However,

$$\psi''\left(\theta_{\tau_1^{(n)}}\right) \leq \theta_{\tau_1^{(n)}}\psi'\left(\theta_{\tau_1^{(n)}}\right) = \theta_{\tau_1^{(n)}}W_{\tau_1^{(n)}}.$$

Because $\psi(\cdot)$ is steep, $\psi'(\cdot)$ must exhibit at least linear growth and therefore there exists a constant $c_1 \in (0,\infty)$ such that

$$\psi''\left(\theta_{\tau_1^{(n)}}\right) \leq c_1 W_{\tau_1^{(n)}}^2$$

and therefore

$$\sup_{n\geq 1} \widetilde{E}\left(\mu(W_{\tau_1^{(n)}})^2\right) \leq (1+c_1) \sup_{n\geq 1} \widetilde{E}\left(\left(W_{\tau_1^{(n)}}\right)^2\right) < \infty.$$

This estimate allows us to conclude the proof of the proposition. ■

4 The Multidimensional Case

In this section we impose the following assumptions:

[i] $X, X_1, X_{2,}...$ is a sequence of iid d-dimensional random vectors with mean zero and continuous distribution

[ii] Let $A \subseteq R^d$ be a convex set with twice continuously differentiable boundary.

[iii] Given $\phi \in R^d$ define $\varrho(\phi) = \log E \exp\left(\phi^T X\right)$, put $I(z) = \max_{\theta \in R^d}(\phi^T z - \varrho(\phi))$ and suppose that there exists $\xi_* \in A$ and $\phi_* \in R^d$ such that

$$I(\xi_*) = \phi_*^T \xi_* - \varrho(\phi_*) = \inf_{z\in A} I(z).$$

[iv] Define $Y_j = \phi_*^T X_j$, $\psi(\theta) = \log E \exp(\theta Y_j)$ and assume that for each $a > 0$ there exists $\theta_a > 0$ such that $\psi'(\theta_a) = a$. Moreover, put

$$J(w) = \theta_w w - \psi(\theta_w)$$

for $w > 0$ and $J(w) = 0$ for $w \leq 0$.

Under assumptions [i] to [iv], we shall develop a strongly efficient estimator for computing $P(S_n/n \in A)$ as $n \nearrow \infty$ where $S_n = X_1 + ... + X_n$. The key reduction step involves noting that because we are assuming that A has a smooth boundary, the most likely path associated with the event $S_n \in nA$ essentially involves a one dimensional large deviation on the part of the random walk associated with the Y_j's. This one dimensional reduction is suggested by the following result by [6]. Such a one dimensional reduction is not, in general, possible. (Consider, for example, the case where $A = [1,\infty) \times [1,\infty)$ and the two components of the random vector X evolve independently, in which case the asymptotic below involves a prefactor of order $1/n$ rather than the "one-dimensional" $n^{-1/2}$seen below.)

Theorem 3 *Under assumptions [i] to [iii], there exists a constant $c(A)$ such that*

$$P(S_n/n \in A) \sim \frac{c(A)}{n^{1/2}} \exp(-nJ(0)) \tag{10}$$

as $n \nearrow \infty$.

We now provide an explicit description of the proposed algorithm.

Algorithm 2

Set $w = \phi_*^T \xi_* > 0$, $L = 1$, $s = 0$, $s' = 0$, $k = 0$, and λ a large positive constant.

WHILE $w > 0$ AND $w \leq \lambda$

Sample X from $F_{J'(w)\phi_*}$ and set

$$\begin{aligned} L &\longleftarrow \exp\left[-J'(w)\,\phi_*^T X + \psi\left(J'(w)\right)\right] L, \\ s &\longleftarrow s + X, \\ k &\longleftarrow k + 1, \\ w &\longleftarrow \phi_*^T \left(n\xi_* - s\right)/(n-k). \end{aligned}$$

LOOP

Sample $X_1, ..., X_{n-k}$ iid rv's from $F_{J'(w)\phi_*^T}$ and set

$$\begin{aligned} s' &\longleftarrow X_1 + ... + X_{n-k}, \\ L &\longleftarrow \exp\left[-J'(w)\,\phi_*^T s' + (n-k)\psi\left(J'(w)\right)\right] L. \end{aligned}$$

OUTPUT $Z_n = L \times I\left(s + s' \in nA\right)$

END

Theorem 4 *Let $\widetilde{E}(\cdot)$ be the expectation operator associated with the change-of-measure described by Algorithm 2. Then, for each $p > 1$ we have*

$$\frac{\widetilde{E}\left(Z_n^p\right)}{P\left(S_n/n \in A\right)^p} = O(1)$$

as $n \nearrow \infty$.

Proof. First define the half-space

$$H \doteq \left\{x \in \mathbb{R}^d : x^T \phi_* \geq \phi_*^T \xi_*\right\} \tag{11}$$

Note that we have the following reduction

$$P\left(S_n \in nH\right) = P\left(Y_1 + ...Y_n \geq n\phi_*^T \xi_*\right),$$

and therefore it is possible to use Algorithm 1 to form an estimator R_n for estimating this probability. Then we have to following three results.

1. $Z_n \leq R_n$ since by definition $A \subset H$, and the likelihood ratios generated for these two estimators are identical.

2. From Theorem 2 we know that for each $p > 1$ that

$$\sup_{n \geq 1} \frac{\tilde{E}\tilde{Y}_n^p}{P\left(S_n' \geq n\phi_*^T \xi_*\right)^p} < \infty.$$

3. Due to the definition of ϕ_*, ξ_* it follows that

$$0 < \inf_{n \geq 1} \frac{P\left(S_n \in nA\right)}{P\left(S_n \in nH\right)} \leq \sup_{n \geq 1} \frac{P\left(S_n \in nA\right)}{P\left(S_n \in nH\right)} < \infty.$$

Combining these three items gives the desired result. ∎

5 Numerical Results

For illustration of the OSDET algorithm, which is defined in equation (7), we consider the problem of estimating $P(S_n \in nA)$. The increments of the sum X_i are iid with multivariate Laplacian distribution. What is meant by this is that $X_i = \sqrt{Z}Y$ where Z_i and Y_i are mutually iid sequences with the following distributions $Z_i \sim \exp(\alpha)$ and $Y_i \sim N(0, \Gamma)$. For the example under consideration we set

$$\Gamma = \begin{pmatrix} 1.684 & .3459 & .6776 \\ .3459 & 4.09 & -1.5864 \\ .6776 & -1.5864 & 1.5411 \end{pmatrix}, \text{ and } \lambda = 1.7.$$

In addition we define the convex set

$$A \doteq \left\{x \in \mathbb{R}^3 : x_3 \geq (x_1 + a_1)^2 + (x_2 + a_2)^2 + b\right\},$$

with $a_1 = 0.8$, $a_2 = -0.7$, and $b = 0.5$.

For comparison we also run the OET algorithm, defined in equation (3), to estimate this problem. In the table and figure below results corresponding to this algorithm are denoted by Traditional IS. Results derived using Algorithm 2 are denoted by Dynamic IS.

In the table and figure below note that the coefficient of variation, see equation (4), of our algorithm does not grow with the large deviations parameter. However the coefficient of variation of the OET algorithm does grow.

Acknowledgement 1 *We gratefully acknowledge various illuminating conversations with Xiao-Li Meng about this problem, as well as his suggestion (which we happily adopted) for the title of this paper.*

n	OSDET	OET	
10	5.09×10^{-9} .0484 $[4.61 \times 10^{-9}, 5.57 \times 10^{-9}]$	4.24×10^{-9} .0622 $[3.72 \times 10^{-9}, 4.75 \times 10^{-9}]$	Estimate Coefficient of Variation Confid. Interval
30	4.35×10^{-21} .048 $[3.94 \times 10^{-21}, 4.76 \times 10^{-21}]$	4.63×10^{-21} .076 $[3.95 \times 10^{-21}, 5.32 \times 10^{-21}]$	
50	5.42×10^{-33} .046 $[4.92 \times 10^{-33}, 5.91 \times 10^{-33}]$	5.27×10^{-33} .09 $[4.34 \times 10^{-33}, 6.20 \times 10^{-33}]$	

Table 1: Comparison of estimates, coefficient of variation, and confidence intervals for both algorithms.

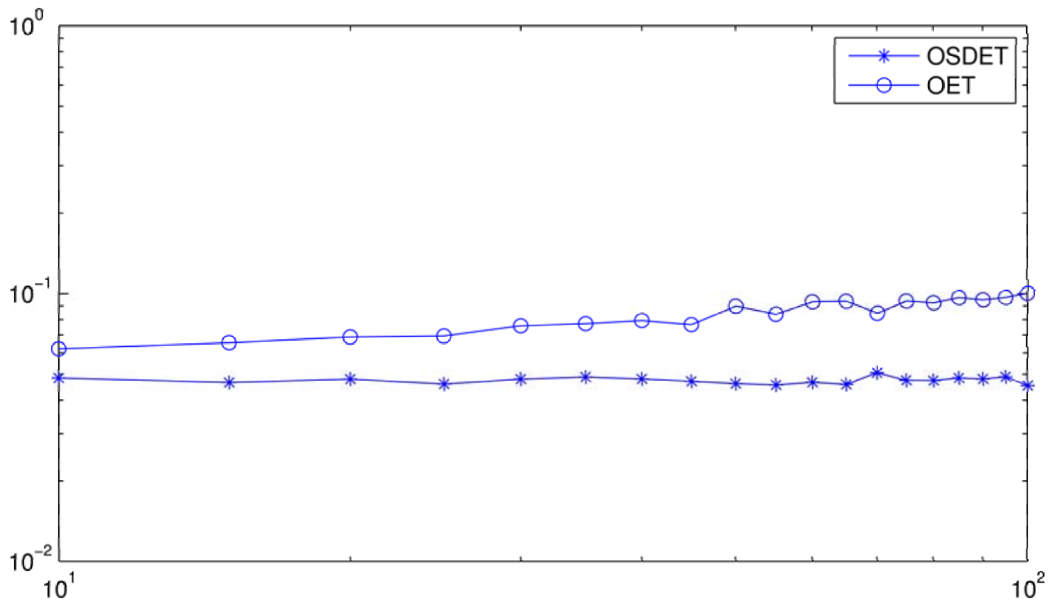

Figure 1: Log-Log plot of Coefficient of variation vs. Large deviations parameter n, for both algorithms.

References

[1] S. Asmussen. *Applied Probability and Queues*. Springer-Verlag, New York, 2003.

[2] S. Asmussen and P. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer-Verlag, New York, NY, USA, 2008.

[3] R. Bahadur and R. R. Rao. On deviations of the sample mean. *Ann. Math. Stat.*, 1960.

[4] J. Blanchet and P. Glynn. Strongly efficient estimators for light-tailed sums. In *valuetools '06: Proceedings of the 1st international conference on Performance evaluation methodolgies and tools*, page 18, New York, NY, USA, 2006. ACM.

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

[6] M. Iltis. Sharp asymptotics of large deviations in $\mathbb{R}^d$. *Journal of Theoretical Probability*, 8:501–522, 1995.

[7] S. Juneja and P. Shahabuddin. Rare event simulation techniques: An introduction and recent advances. In S. G. Henderson and B. L. Nelson, editors, *Simulation*, Handbooks in Operations Research and Management Science. Elsevier, Amsterdam, The Netherlands, 2006.

[8] J. S. Sadowsky. On Monte Carlo estimation of large deviations probabilities. *Ann. Appl. Prob.*, 6:399–422, 1996.