Efficient Simulation of Light-Tailed Sums: an Old-Folk Song Sung to a Faster New Tune...

Jose H. Blanchet, Kevin Leder, and Peter W. Glynn

Abstract We revisit a classical problem in rare-event simulation, namely, efficient estimation of the probability that the sample mean of n independent identically distributed light tailed (i.e. with finite moment generating function in a neighborhood of the origin) random variables lies in a sufficiently regular closed convex set that does not contain their mean. It is well known that the optimal exponential tilting (OET), although logarithmically efficient, is not strongly efficient (typically, the squared coefficient of variation of the estimator grows at rate $n^{1/2}$). After discussing some important differences between the optimal change of measure and OET (for instance, in the one dimensional case the size of the overshoot is bounded for the optimal importance sampler and of order $O(n^{1/2})$ for OET) that indicate why OET is not strongly efficient, we provide a state-dependent importance sampling that can be proved to be strongly efficient. Our procedure is obtained based on computing the optimal tilting at each step, which corresponds to the solution of the Isaacs equation studied recently by Dupuis and Wang [8].

Jose H. Blanchet
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027
e-mail: jose.blanchet@columbia.edu

Kevin Leder
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027
e-mail: kl2457@columbia.edu

Peter W. Glynn
Department of Management Science and Engineering, Stanford University, Stanford, California 94305-4026
e-mail: glynn@stanford.edu

P. L'Ecuyer, A.B. Owen (eds.), *Monte Carlo and Quasi-Monte Carlo Methods 2008*,
DOI 10.1007/978-3-642-04107-5_13,

1 Introduction

Let $X, X_1, X_2, ...$ be a sequence of mean zero independent and identically distributed (iid) d-dimensional random variables (rv's). Assume that A is a sufficiently regular (see Section 5) convex set for which $0 \notin A$. We further assume that A satisfies a technical condition, which is detailed in Section 5. 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 $S_n = X_1 + X_2 + ... + X_n$. 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 it can be proven that 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) [12]. 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, see e.g., [12, 13]. In fact, it can be further shown that OET provides the only iid importance sampling algorithm that achieves logarithmic efficiency [2]. This might not be surprising, given that asymptotically, as $n \to \infty$, OET agrees with the conditional distribution of the S_k's $(k < n)$ given $\{S_n > n\beta\}$ (see Proposition 2 below). It is a simple calculation to check that the conditional distribution is in fact the ideal importance sampling change of measure since it creates an unbiased estimator with zero-variance. We will commonly refer to it as the zero-variance change of measure.

However, it turns out that the squared coefficient of variation [ratio of second moment of estimator to probability of interest squared, see Eq. (5)] 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 (zero-variance change of measure) at scales finer than that of the Law of Large Numbers (LLN). As one illustration of this phenomenon, Proposition 3 establishes that the "overshoot" $S_n - n\beta > 0$ is asymptotically exponentially distributed as $n \nearrow \infty$ under the conditional distribution. In particular the moments of the overshoot stay bounded as n grows. However, due to the Central Limit Theorem (CLT), 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 β, 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 (equivalently estimator) [2, 10]. To produce a strongly efficient estimator, we use "optimal state-dependent exponential tilting" (OSDET). 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 $(S_k : k \geq 0)$. We apply OSDET up until we basically reach the boundary (at which point we turn-off importance sampling) or the distance to the target is sufficiently large relative to the remaining time horizon (at which point we simply apply OET computed from such position until the end of the time horizon). It certainly seems intuitively clear that OSDET will likely reduce the growth of the coefficient of variation as $n \nearrow \infty$ relative to simply applying OET, but there is no reason to expect that one would obtain a bounded coefficient of variation. What plays a crucial role is the fact that (under the assumption that the target set A has a twice continuously differentiable boundary) the dynamic tilting induced by OSDET induces a twice continuously differentiable mapping given by the large deviations rate function. An additional polynomial decay rate of order $O\left(n^{-1/2}\right)$ necessary to control the behavior of the squared coefficient of variation arises thanks to the fact that the conditional expected value of a second order term in the Taylor expansion of this mapping has exactly the right behavior to control (after combining the contributions of all time steps) the previous polynomial decay rate. This result is stated in Lemma 2 and used in Proposition 4.

Let us briefly connect our work with the game-theoretic approach introduced by Dupuis and Wang in [8]. It turns out that the OSDET corresponds to applying importance sampling according to the solution to the associated Isaacs equation (Section 3.4 of [8]). They prove that if such a solution is continuously differentiable then one has a logarithmically efficient estimator. In our setting the solution to the Isaacs equation we work with is in fact twice continuously differentiable in the interior. By applying OSDET in a region where the large deviations scaling is applicable (i.e. before we basically reach the target level or the distance to the level is very large relative to the remaining time horizon), we obtain strong efficiency. It is important to note, however, that our sampler uses a small layer at the boundary to avoid sampling at the point where the solution to the Isaacs equation fails to be twice continuously differentiable. Our work then suggests a connection between the degree of smoothness of the solution to the associated Isaacs equation and the efficiency strength of the corresponding importance sampling estimator.

In previous works on importance sampling and large deviations settings for sample means, proofs of strong efficiency have been limited to systems with heavy tailed characteristics [5] or Gaussian increments [4]. The contribution of our paper is to construct an importance sampling algorithm that can be used in a general class of light tailed distributions that is also provably strongly efficient.

Our proof of strong efficiency relies on the analysis of several martingales that arise naturally from the description of the algorithm, and is of independent interest. Given that OSDET achieves bounded relative error, it may not be surprising then that OSDET also induces a bounded overshoot as $n \nearrow \infty$.

The rest of the paper is organized as follows. Sections 2 to 4 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 introduces the algorithm explicitly and provides a heuristic analysis behind its efficiency. The rigorous details are given in Section 4, where we also show that the overshoot under OSDET remains bounded as $n \nearrow \infty$. Section 5 treats the multidimensional case.

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 underlying our development in Sections 2 to 4.

i) $EX = 0$ and $\mathrm{Var}(X) = \sigma^2$

ii) The log-moment generating function $(\psi(\theta) : \theta \in \mathbb{R})$, defined as $\psi(\theta) = \log E \exp(\theta X)$, is assumed to be *steep* to the right in the sense that for each $w > 0$ there exists $\theta_w > 0$ such that $\psi'(\theta_w) = w$.

iii) We assume that $\inf_{\theta \geq 0} \psi''(\theta) > 0$.

iv) The random variable (rv) X is nonlattice (i.e. the characteristic function has modulus strictly less than one except at the origin).

Assumption i) is obviously introduced without loss of generality. The *steepness* assumption is standard in the large deviations literature and it is useful to rule out distributions with extremely light tails (in particular with compact support). Assumption iii) although satisfied by most models of practical interest beyond light-tailed random variables with finite support (in particular, the condition allows Gaussian, gamma random variables and mixtures thereof) is more technical and is applied only in Lemma 2. Nevertheless, such a condition certainly rules out tails that are lighter than Gaussian. The last assumption is again common in the development of exact asymptotics in large deviations, which are required in our setting because we are concerned with strong efficiency rather than logarithmic efficiency.

We are now ready to describe some results from large deviations that will be useful in our development. The so-called rate function plays a crucial role in the theory of large deviations. In our context, we work with a variant of the standard rate function, $J(\cdot)$, defined for $w \geq 0$ by

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

The standard rate function is defined by optimizing $\theta \in (-\infty, \infty)$. Both $J(\cdot)$ and the standard rate function agree on the positive real line. In particular, note that we have for $w \geq 0$

$$J(w) \triangleq w\theta_w - \psi(\theta_w) \text{ and } J'(w) = \psi'^{-1}(w) = \theta_w. \tag{2}$$

For $w < 0$ we have that $J(w) \equiv 0$ and that $J(\cdot)$ is continuously differentiable at zero. The algorithmic implication of defining $J(\cdot)$ in this way, as we shall see, is that no importance sampling is applied when one reaches the level above $n\beta$. Note, however, that $J(\cdot)$ is not twice continuously differentiable at zero.

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

$$dF_\theta = \exp(\theta x - \psi(\theta))\, dF. \tag{3}$$

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 often use the notation $E_{J'(w)}(\cdot)$, which just means $E_{\theta_w}(\cdot)$.

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 fixed* $\varepsilon > 0$*). In fact, for each* $w > 0$*, the function* $J(w + \cdot)$ *is infinitely differentiable at zero and its Taylor series converges in a neighborhood of the origin.*

Proof. First it is clear from the formula (2) that on the positive line $J(\cdot)$ inherits the smoothness properties of $\psi(\cdot)$, this gives the last two results of the above proposition. The first two results follow from a Taylor expansion of the function $J(\cdot)$ and therefore by necessity a Taylor expansion of θ_w which is obtained using the inverse function theorem. □

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 [3] that will be useful in our development.

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

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

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

The following proposition 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 $\{S_n > n\beta\}$ is $P_{\theta_\beta}(\cdot)$, for a proof see, e.g., [6].

Proposition 2. *Suppose that i), ii) and iv) are in force. Then, for any positive integers $k_1 < k_2 < ... < k_m < \infty$ and for each $x_{k_1}, ..., x_{k_m}$, continuity points of $F(\cdot)$,*

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

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 $\beta > 0$. Then, for all $x > 0$*

$$\lim_{n\to\infty} P\left(S_n - n\beta > x \middle| S_n > n\beta\right) = \exp\left(-\theta_\beta x\right)$$

Proof. Note that from Theorem 1 we have that

$$\lim_{n\to\infty} \frac{P\left(S_n > n\beta + x\right)}{P\left(S_n > n\beta\right)} = \lim_{n\to\infty} P\left(S_n > n\beta + x\right) e^{nJ(\beta)} \theta_\beta \sqrt{2\pi n \psi''(\theta_\beta)}.$$

Thus it remains to show that

$$\lim_{n\to\infty} P\left(S_n > n\beta + x\right) e^{nJ(\beta)} \theta_\beta \sqrt{2\pi n \psi''(\theta_\beta)} = e^{-\theta_\beta x}.$$

Following the notation of [7], Theorem 3.7.4, define the following

$$Y_i = \frac{X_i - \beta}{\sqrt{\psi''(\theta_\beta)}}, \ \psi_n = \theta_\beta \sqrt{n\psi''(\theta_\beta)}, \ W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \text{ and } F_n(y) = P(W_n \leq y).$$

Then a simple calculation (see [7], page 111, for details) gives

$$P\left(S_n > n\beta + x\right) e^{nJ(\beta)} \theta_\beta \sqrt{2\pi n \psi''(\theta_\beta)} = \psi_n \sqrt{2\pi} \int_{\theta_\beta x/\psi_n}^{\infty} e^{-\psi_n y} dF_n(y).$$

One can now follow nearly the same procedure as in the proof of Theorem 3.7.4 of [7], the only difference being the lower limit of integration. Due to the fact that our lower limit of integration is $\theta_\beta x/\psi_n$ the limit of the previous display is $e^{-\theta_\beta x}$, as desired. □

The previous result implies that $P_{\theta_\beta}(\cdot)$ does not accurately describe the behavior of the random walk, conditioned on $\{S_n > n\beta\}$, at time n. In particular under $P_{\theta_\beta}(\cdot)$, the CLT implies that $n^{-1/2}(S_n - n\beta) \overset{D}{\approx} N\left(0, \psi''\left(\theta_\beta\right)\right)$ and thus the overshoot is $(S_n - n\beta) = O\left(n^{1/2}\right)$ in distribution. 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

$$L = \exp\left(-\theta_\beta S_n + n\psi\left(\theta_\beta\right)\right) I\left(S_n > \beta n\right), \tag{4}$$

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 by definition

$$\left(cv_n(L)\right)^2 \triangleq \frac{\mathrm{Var}_{\theta_\beta}(L)}{\left(E_{\theta_\beta}(L)\right)^2} = \frac{E_{\theta_\beta}\left(L^2\right)}{P\left(S_n > n\beta\right)^2} - 1. \tag{5}$$

Under Assumptions i), ii) and iv) we have for a positive constant C_β

$$\begin{aligned} E_{\theta_\beta}\left(L^2\right) &= E_{\theta_\beta}\left[e^{2n\psi(\theta_\beta)-2\theta_\beta S_n} I\left(S_n > n\beta\right)\right] = E\left[e^{n\psi(\theta_\beta)-\theta_\beta S_n} I\left(S_n > n\beta\right)\right] \\ &= e^{-nJ(\beta)} P(S_n > n\beta) E\left\{\exp\left[-\theta_\beta(S_n - n\beta)\right] | S_n > n\beta\right\} \\ &\sim \frac{C_\beta}{\sqrt{n}} e^{-2nJ(\beta)} E\left\{\exp\left[-\theta_\beta(S_n - n\beta)\right] | S_n > n\beta\right\} = \frac{C_\beta}{2\sqrt{n}} e^{-2nJ(\beta)}. \end{aligned}$$

where we use Theorem 1 for the approximation, and Proposition 3 for the final equality. Using Theorem 1 once more we see that as $n \to \infty$, $\left(cv_n(L)\right)^2 = O\left(n^{1/2}\right)$.

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 Intuitive Analysis

The basic idea of our algorithm is that at each discrete time step (provided the random walk is inside a compact set to be described later) one recomputes the OET change of measure. There are two stopping criteria that must be introduced and that we shall discuss in more detail.

The algorithm is explicitly defined as follows. The constant λ below can be chosen arbitrarily as long as $\lambda > 2\beta$, (this is required in the proof of Proposition 5 below).

Algorithm 1

Set $w = \beta > 1/n^{1/2}$, $L = 1$, $s = 0$, $\bar{s} = 0$, $k = 0$, and $\lambda > 2\beta$.
Repeat STEP 1 until $n = k$ OR $w \leq 1/(n-k)^{1/2}$ OR $w > \lambda$.

STEP 1: Sample X from F_{θ_w} [defined by equations (2) and (3)] and set

$$L \longleftarrow \exp\left(-\theta_w X + \psi\left(\theta_w\right)\right) L,$$

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

STEP 2: If $k < n$ sample $X_{k+1}, ..., X_n$ iid rv's from F_{θ_w} and set

$$\begin{aligned} \bar{s} &\longleftarrow X_{k+1} + ... + X_n, \\ L &\longleftarrow \exp\left(-\theta_w \bar{s} + (n-k)\psi(\theta_w)\right) L. \end{aligned}$$

STEP 3: Output $Y_n = L \times I(s + \bar{s} > n\beta)$

The intuition behind the stopping conditions indicated in STEP 2, namely $w \leq 1/(n-k)^{1/2}$ or $w > \lambda$, is the following. First, when $w \leq 1/(n-k)^{1/2}$ there is no need for applying importance sampling sequentially until the end as the event of interest is not rare any more (we have reached the Central Limit Theorem region). It seems intuitive that if one replaces $w \leq 1/(n-k)^{1/2}$ simply by $w \leq 0$ (i.e. stop if we reach the boundary) then one still should obtain strong efficiency. Our analysis, however, requires a stopping criterion that is slightly removed from the origin, such as the one that we impose here. This criterion is used in the proof of Lemma 2 below and basically is imposed to deal with the fact that $J(\cdot)$ is not twice continuously differentiable at the origin. Now, whenever we have that $w > \lambda$, for some large constant λ, then we are approaching a scaling for which the large deviations asymptotics (which motivate the design of the algorithm) are no longer applicable (i.e. a situation where the distance to the target is no longer linearly related to the time to go). At that point, we simply apply the tilting once and for all up until the end of the time horizon.

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

$$W_j = (n\beta - S_j)/(n-j) \tag{6}$$

for $0 \leq j \leq n-1$, and $W_n \triangleq 0$. Next define the following stopping times $\tau_1^{(n)} = \inf\{0 \leq k < n : W_k > \lambda\}$, $\tau_0^{(n)} = \inf\{k \geq 0 : n\beta - S_k \leq (n-k)^{-1/2}\}$, and

$$\tau^{(n)} = \tau_0^{(n)} \wedge \tau_1^{(n)} \wedge n. \tag{7}$$

We define (allowing ourselves a slight abuse of notation) the change of measure used at step j to be

$$\theta_j \triangleq \theta_{W_j} = J'\left(W_j\right). \tag{8}$$

Let us write

$$Z_{1,n} = \exp\left(-\sum_{j=0}^{\tau^{(n)}-1} \left(\theta_j X_{j+1} - \psi\left(\theta_j\right)\right)\right),$$

$$Z_{2,n} = \exp\left(-\theta_{\tau^{(n)}}\left(S_n - S_{\tau^{(n)}}\right) + (n-\tau^{(n)})\psi\left(\theta_{\tau^{(n)}}\right)\right).$$

We can now define the OSDET (optimal state-dependent exponential tilting) estimator resulting from Algorithm 1 as

$$Y_n = Z_{1,n} Z_{2,n} I\left(S_n > n\beta\right) \quad (9)$$

where X_j follows the distribution F_{θ_j} for $1 \le j \le \tau^{(n)}$ and X_j is sampled according to the distribution $F_{\theta_{\tau(n)}}$ for $\tau^{(n)} + 1 \le j \le n$.

The next section is devoted to the rigorous efficiency analysis of Y_n. However, before we provide the full details behind such analysis we will spend the rest of this section explaining the main intuitive steps. It turns out that the most important contribution comes from term $Z_{1,n}$, so this object will be the focus of our discussion here. A substantial portion of the technical development in the next section is dedicated to showing that for any $p \ge 1$ $\sup_{n\ge 1} \widetilde{E}\left((n-\tau^{(n)})^p\right) < \infty$, where $\widetilde{E}(\cdot)$ is used throughout the rest of the paper to denote the expectation operator induced by the importance sampling strategy described in Algorithm 1 (see Proposition 5 below). This in turn is used to argue that the sum in the exponent in $Z_{1,n}$ has basically $n-m$ terms (where m is a constant).

The next results allows us to express the exponent in $Z_{1,n}$ in terms of a telescopic sum involving the function $J(\cdot)$.

Lemma 1. *For* $0 \le j \le n-2$ *let* $w_{j+1}(x) = (n\beta - s - x)/(n-j-1)$ *and* $w_j = (n\beta - s)/(n-j)$*, if* $(n-j)^{-1/2} < w_j < \lambda$ *then*

$$\begin{aligned}&(n-j-1)J\left(w_{j+1}(x)\right) - (n-j)J\left(w_j\right)\\ &= -J'\left(w_j\right)x + \psi\left(\theta_j\right)\\ &+ \frac{(x-w_j)^2}{(n-j-1)}\int_0^1\int_0^1 J''\left(w_j + vu(w_{j+1}(x) - w_j)\right)ududv.\end{aligned}$$

In addition,

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

Remark: A convenient representation that we will use in the future is

$$\begin{aligned}&\int_0^1\int_0^1 J''\left(w_j + vu(w_{j+1}(x) - w_j)\right)ududv \\ &= E\left(J''\left(w_j + VU(w_{j+1}(x) - w_j)\right)U\right)\end{aligned} \quad (10)$$

where U and V are independent uniformly distributed random variables over [0, 1].

Proof. The result is shown by looking at a Taylor expansion of $J\left(w_{j+1}(x)\right)$ about the point w_j. Recall that by definition $J(\cdot)$ is twice differentiable on $\mathbb{R}\backslash\{0\}$ and differentiable on $\mathbb{R}$. Note that

$$w_{j+1}(x) = w_j + \frac{1}{n-j-1}\left(w_j - x\right). \tag{11}$$

On the other hand,

$$J\left(w_{j+1}(x)\right) - J(w_j) = \int_0^1 J'(w_j + u(w_{j+1}(x) - w_j))(w_{j+1}(x) - w_j)du \tag{12}$$

and

$$\begin{aligned} &J'(w_j + u(w_{j+1}(x) - w_j)) - J'(w_j) \\ &= \int_0^1 J''\left(w_j + vu(w_{j+1}(x) - w_j)\right) u(w_{j+1}(x) - w_j)dv. \end{aligned} \tag{13}$$

Note that the integral representation in the previous display is valid for any values of w_j, $w_{j+1}(x)$ and u because $J''(\cdot)$ is continuous except at the origin. Combining (11), (12) and (13) we obtain that

$$\begin{aligned} J\left(w_{j+1}(x)\right) &= J(w_j) + \frac{1}{n-j-1}(w_j - x)J'(w_j) \\ &\quad + \frac{(w_j - x)^2}{(n-j-1)^2}\int_0^1\int_0^1 J''\left(w_j + vu(w_{j+1}(x) - w_j)\right) u\,du\,dv. \end{aligned}$$

The second statement follows from the relationship between ψ and J. □

We now provide an intuitive analysis of $Z_{n,1}$. Using the previous result, the definition of θ_j in (8), and assuming

$$\int_0^1\int_0^1 J''\left(w_j + vu(w_{j+1}(x) - w_j)\right) u\,du\,dv \approx J''(w_j)/2$$

we obtain (using formally $\tau^{(n)} \approx n - m$ for some positive integer m)

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

Under the sampler we have that $S_{n-m} \approx (n-m)\beta$ with high probability and therefore $mJ(W_{n-m}) \approx mJ(\beta)$. One then arrives at the following plausible upper bound (for some constant $c \in (0, \infty)$)

$$\widetilde{E}Z_{n,1}^2 \le c \exp(-2nJ(\beta))\, n^{-1} \times \widetilde{E} \exp\left(-\sum_{j=0}^{n-m-1}\left(\frac{J''(W_j)(X_{j+1}-W_j)^2}{(n-j-1)} - \frac{1}{n-j-1}\right)\right).$$

The main issue then becomes understanding the behavior of the expectation

$$\widetilde{E} \exp\left(-\sum_{j=0}^{n-m-1}\left(\frac{J''(W_j)(X_{j+1}-W_j)^2}{(n-j-1)} - \frac{1}{n-j-1}\right)\right). \tag{14}$$

The crucial observation is that $W_j \in ((n-j)^{-1/2}, \lambda)$ throughout the course of the algorithm and that the random variables

$$\frac{J''(W_j)(X_{j+1}-W_j)^2}{(n-j-1)} - \frac{1}{n-j-1}$$

are martingale differences with conditional variance of order $O\left(1/(n-j-1)^2\right)$. So, working backwards in time, we shall argue that (14) remains bounded as $n \nearrow \infty$, thereby concluding that $\widetilde{E}Z_{n,1}^2 \le cP(S_n > n\beta)^2$ for some constant $c \in (0, \infty)$.

It is important to note that the fact that $J(\cdot)$ is twice continuously differentiable on $(0, \infty)$ seems crucial for the development.

The next section is devoted to the proof of the following result.

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

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

Note that the result stated in Theorem 2 is in fact stronger than just bounded relative error for the estimator, since the result is stated for arbitrary $p > 1$. For a discussion on the benefits of establishing this stronger result see [11].

4 Rigorous Efficiency Analysis

In order to provide the proof of Theorem 2 we need the following result which is a companion to Lemma 1.

Lemma 2. *In the context of Lemma 1 and equation (10), assume that $0 \le j \le n-2$, $(n-j)^{-1/2} < w_j \le \lambda$. Let U and V be independent, uniformly distributed random variables also independent of X_{j+1} given w_j. Set*

$$\eta_{j+1}(X_{j+1}) = VU\left(w_{j+1}(X_{j+1}) - w_j\right) = VU\frac{w_j - X_{j+1}}{n-j-1}.$$

Then, there exists a constant $c(\lambda) \in (0, \infty)$ *such that*

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

Proof. Define $\tilde{\eta}\left(X_{j+1}, w_j\right) \doteq w_j + \eta_{j+1}(X_{j+1})$. Then note that

$$\begin{aligned}
&\left| E_{J'(w_j)} \left(J''\left(\tilde{\eta}\left(X_{j+1}, w_j\right)\right)(X_{j+1} - w_j)^2 U \right) - \frac{1}{2} \right| \\
&= \left| E_{J'(w_j)} \left(\left(J''\left(\tilde{\eta}\left(X_{j+1}, w_j\right)\right) - J''\left(w_j\right)\right) U \left(X_{j+1} - w_j\right)^2 \right) \right| \\
&\leq E_{J'(w_j)} \left(\left| J''\left(\tilde{\eta}\left(X_{j+1}, w_j\right)\right) - J''\left(w_j\right)\right| \left(X_{j+1} - w_j\right)^2 \right).
\end{aligned}$$

Let $\kappa > 0$ fixed (to be chosen later) and write

$$\begin{aligned}
&E_{J'(w_j)} \left(\left| J''\left(\tilde{\eta}\left(X_{j+1}, w_j\right)\right) - J''\left(w_j\right)\right| \left(X_{j+1} - w_j\right)^2 \right) \\
&= E_{J'(w_j)} \left(\left| J''\left(\tilde{\eta}\left(X_{j+1}, w_j\right)\right) - J''\left(w_j\right)\right| \left(X_{j+1} - w_j\right)^2 ; \tilde{\eta}\left(X_{j+1}, w_j\right) \leq 0 \right) \\
&+ E_{J'(w_j)} \left(\left| J''\left(\tilde{\eta}\left(X_{j+1}, w_j\right)\right) - J''\left(w_j\right)\right| \left(X_{j+1} - w_j\right)^2 ; \tilde{\eta}\left(X_{j+1}, w_j\right) \in (0, \kappa) \right) \\
&+ E_{J'(w_j)} \left(\left| J''\left(\tilde{\eta}\left(X_{j+1}, w_j\right)\right) - J''\left(w_j\right)\right| \left(X_{j+1} - w_j\right)^2 ; \tilde{\eta}\left(X_{j+1}, w_j\right) > \kappa \right).
\end{aligned}$$

Let us write I_1, I_2 and I_3 for the last three expectations in the previous display respectively. We have for any positive even integer m, that on $w_j \in ((n-j)^{-1/2}, \lambda)$

$$\begin{aligned}
I_1 &= J''\left(w_j\right) E_{J'(w_j)} \left(\left(X_{j+1} - w_j\right)^2 ; w_j + VU \frac{w_j - X_{j+1}}{n-j-1} \leq 0 \right) \\
&= J''\left(w_j\right) E_{J'(w_j)} \left(\left(X_{j+1} - w_j\right)^2 ; VU \frac{X_{j+1} - w_j}{n-j-1} \geq w_j \right) \\
&\leq J''\left(w_j\right) E_{J'(w_j)} \left(\left(X_{j+1} - w_j\right)^2 ; VU \frac{X_{j+1} - w_j}{n-j-1} \geq \frac{1}{(n-j)^{1/2}} \right) \\
&= J''\left(w_j\right) E_{J'(w_j)} \left(\left(X_{j+1} - w_j\right)^2 ; VU \left(X_{j+1} - w_j\right) \geq \frac{(n-j-1)}{(n-j)^{1/2}} \right) \\
&\leq \frac{J''\left(w_j\right)(n-j)^{m/2}}{(n-j-1)^m} E_{J'(w_j)} \left(\left(X_{j+1} - w_j\right)^{m+2} \right) \leq \frac{c(\lambda)}{(n-j)^{m/2}}.
\end{aligned}$$

In the penultimate inequality we have used $E\left(Z^2; Z/a > 1\right) \leq E\left(Z^{2+m}/a^m\right)$ for $a > 0$ any random variable Z, and positive, even integer m, and in the last line we use the fact that $w_j \in ((n-j)^{-1/2}, \lambda)$. Next, we have that

$$\begin{aligned}
I_2 &= E_{J'(w_j)}\left(\left|J''\left(\tilde{\eta}\left(X_{j+1}, w_j\right)\right) - J''\left(w_j\right)\right|\left(X_{j+1} - w_j\right)^2; \tilde{\eta}\left(X_{j+1}, w_j\right) \in (0, \kappa)\right) \\
&\leq \sup_{0<s<\kappa}\left|J'''(s)\right| E_{J'(w_j)}\left(\left|\eta_{j+1}(X_{j+1})\right|\left(X_{j+1} - w_j\right)^2; \tilde{\eta}\left(X_{j+1}, w_j\right) \in (0, \kappa)\right) \\
&\leq \sup_{0<s<\kappa}\left|J'''(s)\right| E_{J'(w_j)}\left(\frac{\left|X_{j+1} - w_j\right|^3}{n-j-1}; \tilde{\eta}\left(X_{j+1}, w_j\right) \in (0, \kappa)\right) \leq \frac{c(\lambda)}{n-j}.
\end{aligned}$$

for some constant $c(\lambda) \in (0, \infty)$ (this follows because $J(\cdot)$ is smooth on $(0, \kappa)$). Finally, we can use the relationship $J''(\theta) = 1/\psi''\left(J'(\theta)\right)$ to see that if $\kappa > \lambda$

$$\begin{aligned}
I_3 &= E_{J'(w_j)}\left(\left|J''\left(\tilde{\eta}\left(X_{j+1}, w_j\right)\right) - J''\left(w_j\right)\right|\left(X_{j+1} - w_j\right)^2; \tilde{\eta}\left(X_{j+1}, w_j\right) > \kappa\right) \\
&\leq \frac{1}{\inf_{\theta \geq 0} \psi''(\theta)} E_{J'(w_j)}\left(\left(w_j - X_{j+1}\right)^2; VU\frac{w_j - X_{j+1}}{n-j-1} > \kappa - w_j\right) \\
&\leq \frac{1}{\inf_{\theta \geq 0} \psi''(\theta)} E_{J'(w_j)}\left(\left(w_j - X_{j+1}\right)^2; VU\frac{w_j - X_{j+1}}{n-j-1} > \kappa - \lambda\right) \\
&\leq \frac{1}{\inf_{\theta \geq 0} \psi''(\theta)} \frac{E_{J'(w_j)}\left(\left(w_j - X_{j+1}\right)^{2+m}\right)}{(\kappa - \lambda)^m (n-j-1)^m} \leq \frac{c(\lambda)}{(n-j)^m},
\end{aligned}$$

where the previous inequality follows from Assumption 2 just as we did for the previous to last line in the analysis of I_1. Combining our estimates for I_1, I_2 and I_3 we obtain the result. □

We will now use the previous result to analyze the second moment of Y_n under the law induced by the importance sampling distribution $\widetilde{P}$ associated with Algorithm 1. First we need to define the following terms. For $0 \leq j \leq n-2$, define,

$$D_{j+1}\left(X_{j+1}, w_j\right) = E\left[J''\left(w_j + \eta_{j+1}\left(X_{j+1}\right)\right)(X_{j+1} - w_j)^2 U | X_{j+1}\right].$$

Next, for $0 \leq j \leq n-2$, $d_{j+1}\left(w_j\right) = E_{J'(w_j)}\left(D_{j+1}\left(X_{j+1}, w_j\right)\right)$. Finally, write

$$\overline{D}_{j+1}\left(X_{j+1}, w_j\right) = \left(D_{j+1}\left(X_{j+1}, w_j\right) - d_{j+1}\left(w_j\right)\right) I\left(\tau^{(n)} > j\right) \tag{15}$$

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

Proposition 4. *There exists a constant $m(\lambda) \in (0, \infty)$ such that*

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

Proof. First note that Lemma 1 guarantees that given $W_j = w_j$, for $j \leq n-2$

$$-\theta_j X_{j+1} + \psi\left(\theta_j\right) = (n-j-1) J\left(w_{j+1}\left(X_{j+1}\right)\right) - (n-j) J\left(w_j\right)$$

$$-\frac{D_{j+1}\left(X_{j+1},w_j\right)}{2(n-j-1)}.$$

Therefore, on $\tau^{(n)}=n$ we have that

$$-\sum_{j=0}^{\tau^{(n)}-1}\left(\theta_j X_{j+1}-\psi\left(\theta_j\right)\right)=-\sum_{j=0}^{n-2}\left(\theta_j X_{j+1}-\psi\left(\theta_j\right)\right)-\left(\theta_{n-1}X_n-\psi\left(\theta_{n-1}\right)\right)$$

$$=-nJ(\beta)-\sum_{j=0}^{n-2}\frac{D_{j+1}\left(X_{j+1},w_j\right)}{2(n-j-1)}+J\left(W_{n-1}\right)-\left(\theta_{n-1}X_n-\psi\left(\theta_{n-1}\right)\right).$$

On the other hand,

$$J\left(W_{n-1}\right)-\left(X_n\theta_{n-1}-\psi\left(\theta_{n-1}\right)\right)=\theta_{n-1}\left(W_{n-1}-X_n\right)=-\theta_{n-1}\left(S_n-n\beta\right).$$

Therefore,

$$Y_n I\left(\tau^{(n)}=n\right)\leq\exp\left(-nJ(\beta)-\sum_{j=0}^{\left(\tau^{(n)}-1\right)\wedge(n-2)}\frac{D_{j+1}\left(X_{j+1},w_j\right)}{2(n-j)}\right)I\left(S_n>n\beta\right).$$

Similarly, on $\tau^{(n)}<n$ we have

$$-\sum_{j=0}^{\tau^{(n)}-1}\left(\theta_j X_{j+1}-\psi\left(\theta_j\right)\right)=-\sum_{j=0}^{(\tau^{(n)}-1)\wedge(n-2)}\left(\theta_j X_{j+1}-\psi\left(\theta_j\right)\right)\qquad(16)$$

$$=-nJ(\beta)+\left(n-\tau^{(n)}\right)J\left(W_{\tau^{(n)}}\right)-\sum_{j=0}^{(\tau^{(n)}-1)\wedge(n-2)}\frac{D_{j+1}\left(X_{j+1},W_j\right)}{2(n-j)}.$$

On the other hand, we can use the definition of J and θ to get the following equality on $\tau^{(n)}<n$

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

Recalling the definition of our estimator, Y_n we obtain that

$$Y_n I(\tau^{(n)}<n)\leq\exp\left(-nJ(\beta)-\sum_{j=0}^{\left(\tau^{(n)}-1\right)\wedge(n-2)}\frac{D_{j+1}\left(X_{j+1},w_j\right)}{2(n-j)}\right)I\left(S_n>n\beta\right).$$

Therefore, we obtain that

$$Y_n\leq\exp\left(-nJ(\beta)-\sum_{j=0}^{\left(\tau^{(n)}-1\right)\wedge(n-2)}\frac{D_{j+1}\left(X_{j+1},w_j\right)}{2(n-j)}\right)I\left(S_n>n\beta\right).\qquad(17)$$

On the other hand, one can use the fact that $\overline{D}_j = 0$ for $j > \tau^{(n)}$ to see,

$$-nJ(\beta) - \sum_{j=0}^{(\tau^{(n)}-1)\wedge(n-2)} \frac{D_{j+1}\left(X_{j+1}, w_j\right)}{2(n-j)} = -nJ(\beta) - \sum_{j=0}^{n-2} \frac{\overline{D}_{j+1}\left(X_{j+1}, W_j\right)}{2(n-j)}$$
$$- \sum_{j=0}^{(\tau^{(n)}-1)\wedge(n-2)} \left(\frac{d_{j+1}\left(W_j\right)}{2(n-j)} - \frac{1}{2(n-j)}\right) - \sum_{j=0}^{(\tau^{(n)}-1)\wedge(n-2)} \frac{1}{2(n-j)}.$$

We now use Lemma 2 to bound the penultimate term in the previous display,

$$\left| \sum_{j=0}^{(\tau^{(n)}-1)\wedge(n-2)} \left(\frac{d_{j+1}\left(W_j\right)}{2(n-j)} - \frac{1}{2(n-j)}\right) \right| \leq c(\lambda) \sum_{j=1}^{\infty} j^{-2} < \infty. \tag{18}$$

Next using standard bounds on harmonic numbers we have the following

$$\sum_{j=0}^{(\tau^{(n)}-1)\wedge(n-2)} \frac{1}{n-j} \geq \sum_{j=1}^{n} \frac{1}{j} - \sum_{j=1}^{n-\tau^{(n)}+1} \frac{1}{j} \geq \log(n) - \log\left(n - \tau^{(n)} + 1\right) - \frac{1}{2}. \tag{19}$$

Putting the estimates from bounds (16), (18), (19) together into (17) we see that there exists a constant $m(\lambda) \in (0, \infty)$ such that

$$Y_n \leq m(\lambda)\, I\left(S_n \geq n\beta\right) \exp\left[-nJ(\beta) - \sum_{j=0}^{n-2} \frac{\overline{D}_{j+1}\left(X_{j+1}, W_j\right)}{2(n-j)}\right] \left(\frac{n - \tau^{(n)} + 1}{n}\right)^{1/2}.$$

The result then follows. □

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

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

Using the Cauchy-Schwarz inequality, we obtain

$$\left(\widetilde{E} Y_n^p\right)^2 \leq \frac{m(\lambda)\, e^{-2pnJ(\beta)}}{n^p} \widetilde{E} \left(n - \tau^{(n)} + 1\right)^p \widetilde{E} \exp\left(-\sum_{j=0}^{n-2} \frac{p\overline{D}_{j+1}\left(X_{j+1}, W_j\right)}{(n-j)} \right). \tag{20}$$

In order to verify strong efficiency of the algorithm it suffices to show that

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

$$\widetilde{E}\left(\left(n-\tau^{(n)}+1\right)^{p}\right)=O(1)$$

as $n \nearrow \infty$. 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*

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

Proof. By definition $\widetilde{E}[(n-\tau^{(n)})^{p}]=\sum_{k=1}^{n-1}(n-k)^{p}\widetilde{P}(\tau^{(n)}=k)$ and

$$\widetilde{P}\left(\tau^{(n)}=k\right)\leq\widetilde{P}\left(\tau^{(n)}>k-1,W_{k}\geq\lambda\right)+\widetilde{P}\left(\tau^{(n)}>k-1,W_{k}\leq 1/(n-k)^{1/2}\right).$$

Now, define the martingale difference $\tilde{D}_{j}=\left(W_{j}-W_{j-1}\right)\times I\left(\tau^{(n)}>j-1\right)$ (for $1\leq j\leq n-1$) and note that (recall that $W_{0}=\beta$)

$$\widetilde{P}\left(\tau^{(n)}>k-1,W_{k}\geq\lambda\right)=\widetilde{P}\left(\tau^{(n)}>k-1,\sum_{j=1}^{k}\tilde{D}_{j}\geq\lambda-\beta\right),$$

$$\widetilde{P}\left(\tau^{(n)}>k-1,W_{k}\leq(n-k)^{-1/2}\right)=\widetilde{P}\left(\tau^{(n)}>k-1,\sum_{j=1}^{k}\tilde{D}_{j}\leq(n-k)^{-1/2}-\beta\right).$$

We will show that there exists a constant $m_{1}\in(0,\infty)$ such that

$$\widetilde{P}\left(\sum_{j=1}^{k}\tilde{D}_{j}\geq\lambda-\beta\right)\leq m_{1}\exp\left(-(n-k)^{1/3}\right). \tag{21}$$

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

$$\tilde{D}_{k}=\left(\frac{w_{k-1}-X_{k}}{n-k}\right)I(\tau^{(n)}>k-1).$$

Thus, if $\eta\in(0,\infty)$ then $\widetilde{E}\left(\exp\left(\eta\tilde{D}_{k}\right)\Big|\,\tilde{D}_{1},...,\tilde{D}_{k-1}\right)=\exp\left(\chi_{k}\left(\frac{\eta}{n-k}\right)\right)$, where

$$\chi_{k}\left(\frac{\eta}{n-k}\right)=\frac{\eta w_{k-1}I\left(\tau^{(n)}>k-1\right)}{n-k} + \psi\left(\frac{-\eta I\left(\tau^{(n)}>k-1\right)}{n-k}+\theta_{k-1}\right)-\psi\left(\theta_{k-1}\right).$$

If $\eta=(n-k)^{1/3}$, then because $\tau^{(n)}>k-1$ (which implies $W_{k-1}\in(1/(n-k)^{1/2},\lambda)$), the smoothness of ψ, and $\psi'(\theta_{k-1})=w_{k-1}$, we can use a Taylor expansion with remainder term to see that 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

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

Chebyshev's inequality then yields inequality (21) as indicated. A completely analogous estimate can be obtained for $\widetilde{P}\left(\sum_{j=1}^{k}\tilde{D}_j \leq (n-k)^{-1/2} - \beta\right)$.
Therefore we conclude that

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

which is clearly bounded as $n \nearrow \infty$. □

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

Proposition 6. *For each* $\eta > 0$ *and* $p > 1$ *we have that*

$$\sup_{n\geq 1}\widetilde{E}\left(\exp\left(-p\sum_{j=0}^{n-2}\frac{\eta\overline{D}_{j+1}\left(X_{j+1}, W_j\right)}{(n-j)}\right)\right) < \infty.$$

Proof. We have that for any $\eta > 0$, given $W_j = w_j$, on $\tau^{(n)} > j$ and $0 \leq j \leq n-2$

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

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

From the definition of $\overline{D}_{j+1}$, and the convexity of $J(\cdot)$ 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 $(n-j)^{-1/2} \leq w_j \leq \lambda$) there exists $m_3(\lambda) \in (0, \infty)$ such that

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

Iterating the previous calculations for $j = n-2, n-3, ..., 0$ we obtain that

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

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

Theorem 2 follows easily from the previous two propositions by recalling the bound in display (20).

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

$$\begin{aligned}|S_{\tau^{(n)}} - n\beta| &\leq |S_{\tau^{(n)}-1} - (\tau^{(n)} - 1)\beta| + |X_{\tau^{(n)}} - (n - \tau^{(n)} + 1)\beta| \\ &\leq 2\lambda(n - \tau^{(n)} + 1) + |X_{\tau^{(n)}}|,\end{aligned}$$

where the second inequality follows from the definition of $\tau^{(n)}$ and that $\lambda > \beta$. Of course the expected value of $n - \tau^{(n)}$ stays bounded as $n \nearrow \infty$ thanks to Proposition 5. Therefore it remains to look at the expected value of $X_{\tau^{(n)}}$ based on the value of $\tau^{(n)}$. In particular, we have that

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

It follows from steepness and the fact that $\tau^{(n)} > j$ that there exists a constant $c_0(\lambda) \in (0,\infty)$ such that $\widetilde{E}(|X_{j+1}|^2 \;|\tau^{(n)} > j) \leq c_0(\lambda)$.

As we established in the proof of Proposition 5, it follows that $\widetilde{P}(\tau^{(n)} = j+1) \leq m_1 \exp\left(-(n-j)^{1/3}\right)$ for some constant $m_1 \in (0,\infty)$. Therefore, we obtain that

$$\sup_{n\geq 1} \widetilde{E}|X_{\tau^{(n)}}| \leq c_0(\lambda)^{1/2} m_1^{1/2} \sum_{j=1}^{\infty} \exp\left(-j^{1/3}/2\right) < \infty \tag{22}$$

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

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

$$\widetilde{E}\left(|S_n - S_{\tau^{(n)}}|; \tau_0^{(n)} \leq \tau_1^{(n)}\right) \leq (E|X_1|)\,\widetilde{E}(n - \tau_0^{(n)}).$$

Observe $E|X_1|$ appears because from time $\tau_0^{(n)} + 1$ up to n the sampling is done under the original / nominal distribution. Again we can use Proposition 5 to bound the expectation of $n - \tau^{(n)}$. Thus it suffices to consider

$$\widetilde{E}\left(|S_n - S_{\tau^{(n)}}|;\tau_0^{(n)} > \tau_1^{(n)}\right).$$

Let us define,

$$\mu(W_{\tau_1^{(n)}}) = \widetilde{E}\left(\left|X_{\tau_1^{(n)}+1}\right|\left|W_0,...,W_{\tau_1^{(n)}},\tau_1^{(n)}\right.\right) = \int_{-\infty}^{\infty}|x|\exp\left(\theta_{\tau_1^{(n)}}x - \psi\left(\theta_{\tau_1^{(n)}}\right)\right)dF(x).$$

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

$$\widetilde{E}\left(|S_n - S_{\tau^{(n)}}|;\tau_1^{(n)} < \tau_0^{(n)}\right) \le \widetilde{E}\left(\sum_{j=\tau_1^{(n)}+1}^{n}\left|X_j\right|;\tau_1^{(n)} < \tau_0^{(n)}\right)$$
$$\le \sum_{k=1}^{n-1}\sum_{j=k+1}^{n}\widetilde{E}\left(\left|X_j\right|;\tau_1^{(n)} = k\right) \le \sum_{k=1}^{n-1}(n-k)\widetilde{E}\left(\mu(W_{\tau_1^{(n)}});\tau_1^{(n)} = k\right)$$
$$\le \widetilde{E}\left(\mu(W_{\tau_1^{(n)}})^2\right)^{1/2}\sum_{k=1}^{n-1}(n-k)\left(\widetilde{P}\left(\tau_1^{(n)} = k\right)\right)^{1/2}.$$

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

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

It now remains to show that $\widetilde{E}(\mu(W_{\tau_1^{(n)}})^2)$ stays bounded as n goes to infinity. Notice that $0 \le W_{\tau_1^{(n)}} \le 2\lambda + |X_{\tau_1^{(n)}}|$. A similar analysis behind Eq. (22) then allows us to conclude

$$\sup_{n\ge 1}\widetilde{E}(W_{\tau_1^{(n)}})^2 < \infty. \tag{23}$$

Observe that $W_{\tau_1^{(n)}} = \int_{-\infty}^{\infty} x\exp[\theta_{\tau_1^{(n)}}x - \psi(\theta_{\tau_1^{(n)}})]dF(x)$ therefore

$$\mu\left(W_{\tau_1^{(n)}}\right) = W_{\tau_1^{(n)}} + 2\int_{-\infty}^{0}|x|\exp\left(\theta_{\tau_1^{(n)}}x - \psi\left(\theta_{\tau_1^{(n)}}\right)\right)dF(x)$$
$$\le W_{\tau_1^{(n)}} + 2\int_{-\infty}^{0}|x|\exp\left(-\psi\left(\theta_{\tau_1^{(n)}}\right)\right)dF(x).$$

Due to strict convexity, the fact that $\psi(0) = 0$, and $\psi'(0) = 0$ we have that $\psi(\theta) \ge 0$ for $\theta \ge 0$, thus

$$\mu(W_{\tau_1^{(n)}}) \le W_{\tau_1^{(n)}} + 2E|X_1|.$$

The proof is completed by combining the bound in the previous display with the result from (23). □

5 The Multidimensional Case

A vector $x \in R^d$ is always assumed to be a column vector, and we denote its transpose by x^T. Therefore the inner product of two vectors x, y is denoted by $x^T y$. The Hessian matrix of a function $f : \mathbb{R}^d \to \mathbb{R}$ is denoted by $D^2 f$. In this section we impose the following assumptions:

[i] $\left(\tilde{X}_j : j \geq 1\right)$ is a sequence of iid d-dimensional random vectors with mean zero and continuous distribution.

[ii] Let A be a closed convex set for which $0 \notin A$.

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

$$I(\xi_*) = \phi_*^T \xi_* - \varrho(\phi_*) = \inf_{z \in A} I(z). \tag{24}$$

[iv] Assume there is a local change of coordinates $T : \mathbb{R}^{d-1} \supset U \to \partial A$ (where U is an open set) so that the Hessian of $I \circ T$ is well defined and positive definite at $T^{-1}(\xi_*)$, see [1] for details.

[v] Define $X_j = \phi_*^T \tilde{X}_j$ and put $\psi(\theta) = \log E \exp\left(\theta X_j\right)$ for $\theta \in \mathbb{R}$. Suppose that $\psi(\cdot)$ satisfies Assumptions ii) and iii) from Section 2.

Assumption [iv] in particular requires ∂A to be twice continuously differentiable at ξ_*. The geometric interpretation, explained in [9] and [1], is that the boundary of ∂A must be more flat than the level curve of I corresponding to the value $I(\xi_*)$ at ξ_*. If assumption [iv] is violated then our algorithm is still logarithmically efficient. However, the relative error will grow at a polynomial rate which can be shown to be not larger than that of OET.

Analogous to the one-dimensional setting we define the exponential family $(\widetilde{F}_\phi : \phi \in \mathbb{R}^d)$ generated by the distribution $\widetilde{F}(\cdot) = P(\widetilde{X} \leq \cdot)$ (inequality is taken componentwise)

$$d\widetilde{F}_\phi = \exp\left(\phi^T x - \varrho(\phi)\right) d\widetilde{F}.$$

Note that by the definition of ϕ_* and ξ_*,

$$\xi_* = \frac{E\left[\tilde{X} \exp\left(\phi_*^T \tilde{X}\right)\right]}{E\left[\exp\left(\phi_*^T \tilde{X}\right)\right]} \text{ and thus } \phi_*^T \xi_* = \frac{E\left[\phi_*^T \tilde{X} \exp\left(\phi_*^T \tilde{X}\right)\right]}{E\left[\exp\left(\phi_*^T \tilde{X}\right)\right]}.$$

We define $\beta = \phi_*^T \xi_*$ and use exactly the same notation as in Section 2 in the context of equations (3), (1) and (2). So, we see that $\theta_\beta = 1$ and $J(\beta) = I(\xi_*)$. Moreover, since the analysis of the estimator will be reduced to the one dimensional setting taking advantage of the random variables X_j's defined in Assumption [iv], we also refer the reader to the definitions of W_j, the associated stopping time $\tau^{(n)}$, the change of measure θ_j and the likelihood ratio in equations (11), (7), (8) and (9).

Under Assumptions [i] to [v] we shall develop a strongly efficient estimator for computing $P\left(\tilde{S}_n / n \in A\right)$ as $n \nearrow \infty$ where $\tilde{S}_n = \tilde{X}_1 + \ldots + \tilde{X}_n$. First, let us recall the following result from [9].

Theorem 3. *Under Assumptions [i] to [iv] there exists a constant* $c(A)$ *such that*

$$P\left(\tilde{S}_n/n \in A\right) \sim \frac{c(A)}{n^{1/2}} \exp\left(-nJ(\beta)\right) \tag{25}$$

as $n \nearrow \infty$.

The previous result allows us to reduce, under assumptions [i] to [v], the multi-dimensional case problem to the one dimensional case studied in Sections 3 and 4. Note that Assumptions [ii] and [iv] are particularly important because they ensures that the premultiplying constant in (25) is $c(A)/n^{1/2}$ (i.e. the same order as in the one dimensional case). The premultiplying factor can in fact take the form $c(A)n^{\gamma}$ for $-\infty < \gamma \leq (d-2)/2$. Only with Assumptions [ii] and [iv] are we assured that $\gamma = -1/2$, [9] addresses the issue of identifying γ for smooth Borel subsets of R^d. If $\gamma = 1/2$ then modulo constants $P(\tilde{S}_n \in nA)$ behaves like $P(S_n \geq n\beta)$. In order for that to occur one must ensure that the boundary of A does not curve away too sharply from the level set of I at the dominating point ξ_*, Assumption [iv] ensures that the curvature of A with respect to I is sufficiently small.

We now provide an explicit description of the proposed algorithm.

Algorithm 2

Set $w = \beta = \phi_*^T \xi_* > 0$, $L = 1$, $s = 0$, $\bar{s} = 0$, $k = 0$, and λ a large positive constant. Repeat STEP 1 until $n = k$ OR $w \leq (n-k)^{-1/2}$ OR $w \geq \lambda$.

STEP 1: Sample $\tilde{X}$ from $\widetilde{F}_{\theta_w \phi_*}$ and set

$$\begin{aligned} L &\longleftarrow \exp\left(-\theta_w \phi_*^T \tilde{X} + \psi(\theta_w)\right) L, \\ s &\longleftarrow s + X, \\ k &\longleftarrow k + 1, \\ w &\longleftarrow \left(n\beta - \phi_*^T s\right)/(n-k). \end{aligned}$$

STEP 2: If $k < n$ sample $\tilde{X}_{k+1}, ..., \tilde{X}_n$ iid rv's from $\widetilde{F}_{\theta_w \phi_*}$ and set

$$\begin{aligned} \bar{s} &\longleftarrow \tilde{X}_{k+1} + ... + \tilde{X}_n, \\ L &\longleftarrow \exp\left(-\theta_w \phi_*^T \bar{s} + (n-k)\psi(\theta_w)\right) L. \end{aligned}$$

STEP 3: Output $Z_n = L \times I(s + \bar{s} \in nA)$.

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*

$$\sup_{n \geq 1} \frac{\widetilde{E}\left(Z_n^p\right)}{P\left(\tilde{S}_n/n \in A\right)^p} < \infty.$$

Proof. Since $I\left(\tilde{S}_n \in nA\right) \leq I(S_n \geq n\beta) = I\left(S_n \geq n\phi_*^T \xi_*\right)$ we obtain that the estimator obtained by running Algorithm 2 is bounded by

$$Y_n = \exp\left(- \sum_{j=0}^{\tau^{(n)}-1} \left(\theta_j X_{j+1} - \psi\left(\theta_j\right) \right) \right)$$
$$\times I\left(S_n \geq n\beta\right) \exp\left(-\theta_{\tau^{(n)}} \left(S_n - S_{\tau^{(n)}} \right) + (n - \tau^{(n)}) \psi\left(\theta_{\tau^{(n)}} \right) \right).$$

Therefore

$$\sup_{n\geq 1} \frac{\tilde{E} Z_n^p}{P(\tilde{S} \in nA)^p} \leq \sup_{n\geq 1} \frac{\tilde{E} Y_n^p}{P(S_n \geq n\beta)^p} \sup_{n\geq 1} \frac{P(S_n \geq n\beta)^p}{P(\tilde{S} \in nA)^p}.$$

The proof is completed by using Theorems 1 and 3 because $\sup_{n\geq 1} \frac{P(S_n \geq n\beta)}{P(\tilde{S}_n \in nA)} < \infty$. □

Acknowledgements The work of Jose Blanchet and Kevin Leder was supported by NSF grants NSF-0846816 and DMS-0806145.

References

1. Andriani, C., Baldi, P.: Sharp estimates of deviations of the sample mean in many dimensions. Ann. Inst. Henri Poincare **33**, 371–385 (1997)
2. Asmussen, S., Glynn, P.: Stochastic Simulation: Algorithms and Analysis. Springer-Verlag, New York, NY, USA (2008)
3. Bahadur, R., Rao, R.R.: On deviations of the sample mean. Ann. Math. Stat. (1960)
4. Blanchet, J., Glynn, P.: Strongly efficient estimators for light-tailed sums. In: valuetools '06: Proceedings of the 1st international conference on Performance evaluation methodolgies and tools, p. 18. ACM, New York, NY, USA (2006). DOI http://doi.acm.org/10.1145/1190095.1190118
5. Blanchet, J., Liu, J.C.: State-dependent importance sampling for regularly varying random walks. Adv. in Appl. Probab. **40**, 1104–1128 (2008)
6. Borovkov, A.A.: On the limit conditional distributions connected with large deviations. Siberian Mathematical Journal **37**, 635–646 (1996)
7. Dembo, A., Zeitouni, O.: Large deviations techniques and applications, second edn. Springer, New York (1998)
8. Dupuis, P., Wang, H.: Importance sampling, large deviations, and differential games. Stoch. and Stoch. Reports **76**, 481–508 (2004)
9. Iltis, M.: Sharp asymptotics of large deviations in $\mathbb{R}^d$. Journal of Theoretical Probability **8**, 501–522 (1995)
10. Juneja, S., Shahabuddin, P.: Rare event simulation techniques: An introduction and recent advances. In: S.G. Henderson, B.L. Nelson (eds.) Simulation, Handbooks in Operations Research and Management Science. Elsevier, Amsterdam, The Netherlands (2006)
11. L'Ecuyer, P., Blanchet, J., Tuffin, B., Glynn, P.: Asymptotic robustness of estimators in rare-event simulation. ACM TOMACS. To Appear
12. Sadowsky, J.S.: On Monte Carlo estimation of large deviations probabilities. Ann. Appl. Prob. **6**, 399–422 (1996)
13. Siegmund, D.: Importance sampling in the Monte Carlo study of sequential tests. Ann. Stat. **3**, 673–684 (1976)