Asymptotic Simulation Efficiency Based on Large Deviations

Peter Glynn

Department of Management Science and Engineering

Stanford University

Stanford, CA 94305, U.S.A.

Sandeep Juneja

School of Technology and Computer Science

Tata Institute of Fundamental Research

Mumbai, INDIA - 400005

February 18, 2013

Abstract

Consider a simulation estimator $\alpha(c)$ based on expending c units of computer time to estimate a quantity α. In comparing competing estimators for α, a natural figure of merit is to choose the estimator that minimizes the computation time needed to reduce the error probability $P(|\alpha(c) - \alpha| > \epsilon)$ to below some prescribed value δ. In this paper, we develop large deviations results that provide approximations to the computational budget necessary to reduce the error probability to below δ when δ is small. This approximation depends critically on both the distribution of the estimator itself and that of the random amount of computer time required to generate the estimator, and leads to different conclusions regarding the choice of preferred estimator than those obtained when one requires the error tolerance ϵ to be small. The "small ϵ" regime leads to variance-based selection criteria, and has a long history in the simulation literature going back to Hammersley and Handscomb.
[1].

Keywords: *Simulation, Efficiency, Large Deviations*

_____

[1]A preliminary version of this paper appeared in Glynn and Juneja (2008)

1

1 Introduction

This paper is concerned with the theoretical foundations underlying algorithmic efficiency in the Monte Carlo simulation setting. In particular, when one has two alternatives available for computing a given quantity α, a natural criterion for selecting one estimator over another is to choose, for a given error tolerance ϵ, that estimator which minimizes the computational budget c necessary to reduce the probability of the estimator deviating from α by more than ϵ to below some level δ. A natural asymptotic regime that simplifies the mathematical analysis is to consider this criterion in the setting where the error probability δ is small.

For a given estimator $\alpha(c)$, this paper proves, in significant generality, that

$$\frac{1}{c} \log P(|\alpha(c) - \alpha| > \epsilon) \to -I(\epsilon) \tag{1}$$

as $c \to \infty$, and computes $I(\epsilon)$ explicitly. This result relies heavily upon the existing theory of large deviations, and can be viewed as a large deviations result for the random process $\{\alpha(c)\}$. In the presence of (1), we find that the computational budget c needed to reduce the error probability to δ is asymptotic to $(1/I(\epsilon)) \log(1/\delta)$ as $\delta \to 0$ (ϵ fixed). Hence, the result (1) that is the central contribution of this paper establishes that when the error probability δ is small, one should select an estimator with the larger value of $I(\epsilon)$.

An alternative asymptotic regime that also simplifies the analysis of our estimator selection criterion is to consider the computational budget c in the setting where the error tolerance ϵ is small. In this context, establishing a central limit theorem (CLT) for $\alpha(c)$ of the form

$$c^{1/2}(\alpha(c) - \alpha) \Rightarrow \eta N(0, 1)$$

as $c \to \infty$ (where $N(0, 1)$ is a standard normal random variable, and $\Rightarrow$ denotes weak convergence) ensures that the computational budget c needed to reduce the error tolerance to ϵ is asymptotic to $z(\delta)^2 \eta^2 / \epsilon^2$ as $\epsilon \to 0$ (for a fixed δ), where $z(\delta)$ is the solution to the equation $(P(|N(0, 1)| > z) = \delta)$. Hence, in this asymptotic regime, one should choose an estimator with the smaller value of η^2. In general, minimizing η^2 is not equivalent to maximizing $I(\epsilon)$. A principal contribution of this paper is that consideration of these two asymptotic regimes leads to two different criteria for judging the efficiency of one estimator relative to another.

The second approach, in which ϵ is sent to zero, is the cornerstone of the predominant theoretical methodology used today for assessing computational efficiency. This idea goes back to the early days of the Monte Carlo method. When $\alpha(c)$ is computed by generating independent identically distributed (iid) copies of a common random variable (rv) X until the computational budget c is exausted, Hamersley and Handscomb (1964) argued that η^2 is the product of the

mean time to generate X and the variance of X; Glynn and Whitt (1992) later provided the full theoretical support for this conclusion. However, we note that sending ϵ to zero suggests that the simulation is intended to compute α to a large number of significant digits. Given the slow rate of convergence of Monte Carlo sampling-based methods, Monte Carlo is often used when only a small number of significant figures of accuracy is necessary. In such settings, one could argue that making δ small is more appropriate and realistic, and hence the selection criterion introduced in this paper would then be more appropriate.

This paper is organized as follows. Section 2 provides the mathematical framework for our problem. Here we also state our main large deviations result for the error probability for $\alpha(c)$ as $c \to \infty$ in a setting that permits the observations collected to be dependent (as in the steady-state simulation case). In Section 3, we state the exact asymptotic for $\alpha(c)$'s error probability when the underlying samples are independent. We conclude the paper in Section 4 by discussing some of the drawbacks of maximizing $I(\epsilon)$ as a measure of efficiency, and elaborate on some of the implications of our theory for "selection of best system" problems. As we discuss there, the CLT-based criterion η^2 turns out to be more tractable and useful as a guide in constructing efficient algorithms. Nevertheless, this paper points out that $I(\epsilon)$ is, at least conceptually, an equally defensible measure of algorithmic efficiency, and is to be preferred in settings where making the error probability small takes higher priority than making the error tolerance small. The task of making $I(\epsilon)$ a more useful and tractable measure of computational efficiency is left for future work.

2   The General Result

We develop here the most general version of our result, in a context which allows for the type of dependence that arises, for example, in the steady-state single replication simulation setting. Our framework consists of a sequence $((X_n, \tau_n) : n \geq 1)$ of random vectors in which τ_n is a positive random variable (rv) that represents the amount of computer time required to generate X_n. In the steady-state setting, X_n then corresponds to the simulation output generated at time n, so that the (X_n, τ_n)'s are then a dependent (auto-correlated) sequence. Of course, in the classic Monte Carlo setting in which one wishes to compute $\alpha = EX$, where X can be generated in finite time, the X_i's are iid copies of X (and, more generally, the (X_n, τ_n)'s are iid).

Put $X_n^+ = \max(X_n, 0)$ and $X_n^- = \max(-X_n, 0)$ so that $X_n = X_n^+ - X_n^-$. Let $S_n = X_1 + \cdots + X_n$, $S_n^+ = X_1^+ + \cdots + X_n^+$, $S_n^- = X_1^- + \cdots + X_n^-$ (with $S_0 = S_0^+ = S_0^- = 0$), $T_n = \tau_1 + \cdots + \tau_n$ (with

$T_0 = 0$), and $\xi_n = (S_n^+, S_n^-, T_n)^T$ (the superscript T denotes the transpose operator). Finally, for $0 = t_0 < t_1 < \ldots < t_m$, set $\vec{t} = (t_1, \ldots, t_m)$ and

$$\tilde{\psi}_{r,\vec{t}}(\beta_1, \ldots, \beta_m) \triangleq \frac{1}{r} \log E \exp(\sum_{i=1}^{m} \beta_i^T (\xi_{\lfloor rt_i \rfloor} - \xi_{\lfloor rt_{i-1} \rfloor}))$$

for $\beta_i \triangleq (u_i, v_i, \eta_i)^T \in \Re^3$, $1 \le i \le m$, where $\lfloor s \rfloor$ is the greatest integer less than or equal to s. Because log moment generating functions are known to be convex, $\tilde{\psi}_{r,\vec{t}}$ is convex. Furthermore, the non-negativity of $\xi_{\lfloor rt_i \rfloor} - \xi_{\lfloor rt_{i-1} \rfloor}$ ensures that $\tilde{\psi}_{r,\vec{t}}$ is non-decreasing in $\beta_1, \ldots, \beta_m$.

The following (Gartner-Ellis type) assumption is commonly used within the large deviations literature, as it covers wide variety of stochastic processes that exhibit dependence (e.g., Gaussian sequences, Markov-dependent sequences, etc.).

A1. Suppose that there exists an open convex set $\mathcal{N} \subset \Re^3$ and a function $\tilde{\psi} : \Re^3 \to \Re$ satisfying:

i.) For each $m \ge 1$ and $0 = t_0 < t_1 < \ldots < t_m$ and $\beta_1, \ldots, \beta_m \in \mathcal{N}$,

$$\tilde{\psi}_{r,\vec{t}}(\beta_1, \ldots, \beta_m) \to \sum_{i=1}^{m} (t_i - t_{i-1}) \tilde{\psi}(\beta_i)$$

as $r \to \infty$;

ii.) $\tilde{\psi}$ is strictly convex and twice continuously differentiable on $\mathcal{N}$;

iii.) $\Re_+^2 \times \{0\} \subseteq \mathcal{N}$.

To enhance our understanding of the above assumption, consider the Markov-dependent setting in which $X_n = f(Z_n)$, $\tau_n = h(Z_n)$, and $(Z_n : n \ge 0)$ is a finite-state S-valued irreducible Markov chain with transition matrix $P = (P(x, y) : x, y \in S)$, and with $f : S \to \Re$ and $h : S \to \Re_+$ suitably non-constant functions. Then **A1** is satisfied with $\mathcal{N} = \Re^3$ and

$$\tilde{\psi}(u, v, \eta) = \log \lambda(u, v, \eta),$$

where $\lambda(u, v, \eta)$ is the Perron-Frobenius eigenvalue of the non-negative irreducible matrix $K = (K(x, y) : x, y \in S)$ and

$$K(x, y) = \exp(u \max(f(x), 0) + v \max(-f(x), 0) + \eta h(x)) P(x, y).$$

(see, e.g., Iscoe, Ney and Nummelin 1985). In addition, note that because $\tilde{\psi}$ inherits the monotonicity of $\tilde{\psi}_{r,\vec{t}}$, the strict convexity of $\tilde{\psi}$ implies that $\frac{\partial}{\partial \eta} \tilde{\psi}(u, v, \eta) > 0$ for $(u, v, \eta) \in \mathcal{N}$. Set

$$\psi(\theta, \eta) = \tilde{\psi}(\theta, -\theta, \eta),$$

so that

$$\psi(\theta, \eta) = \lim_{n \to \infty} \psi_n(\theta, \eta)$$

where $\psi_n(\theta, \eta) = \frac{1}{n} \log E \exp(\theta S_n + \eta T_n)$.

Under **A1**,

$$\bar{X}_n \triangleq n^{-1} \sum_{i=1}^{n} X_i \Rightarrow \alpha \tag{2}$$

as $n \to \infty$, where $\alpha = \frac{\partial}{\partial \theta} \psi(0, 0)$; see, for example, Dembo and Zeitouni (1998), pg. 54. Thus, $\bar{X}_n$ is an estimator for the parameter α. Of course, when the (X_i, τ_i)'s are independent and identically distributed, $\alpha = EX_1$. On the other hand, when $(X_n : n \geq 1)$ is a Markov-dependent sequence, $\alpha = EX_\infty$, where EX_∞ is the equilibrium expectation of $(X_n : n \geq 1)$. In fact, **A1** actually implies almost sure convergence of $\bar{X}_n$ to α.

Proposition 1 *Under* **A1**, *$\bar{X}_n \to \alpha$ almost surely as $n \to \infty$.*

Proof. We will show that for each $\epsilon > 0$,

$$P(\bar{X}_n > \alpha + \epsilon \text{ infinitely often}) = 0. \tag{3}$$

It follows that there exists a set having probability one on which (3) holds simultaneously for all rational values of ϵ, proving that

$$\overline{\lim}_{n \to \infty} \bar{X}_n \leq \alpha \ \ a.s.$$

A similar argument establishes that $\underline{\lim}_{n \to \infty} \bar{X}_n \geq \alpha \ \ a.s.$ proving the result. To obtain (3), we verify that

$$\sum_{n=1}^{\infty} P(\bar{X}_n > \alpha + \epsilon) < \infty, \tag{4}$$

so that the Borel-Cantelli lemma yields (3).

But (4) is a consequence of the Gartner-Ellis theorem, since **A1** and Markov inequality then guarantees that for each $\theta > 0$ and $\epsilon > 0$,

$$
\begin{aligned}
P(\bar{X}_n > \alpha + \epsilon) \ &\leq \ \exp(-n\theta(\alpha + \epsilon) + n\psi_n(\theta, 0)) \\
&= \ \exp(-n\theta(\alpha + \epsilon) + n(\psi(\theta, 0) + o(1))) \\
&= \ \exp(-n\theta(\alpha + \epsilon) + n(\psi(\theta, 0) - \psi(0, 0)) + o(n)) \ \ (\text{since } \psi(0,0) = 0) \\
&= \ \exp\left(-n\theta(\alpha + \epsilon) + n\left(\theta\alpha + \frac{\theta^2}{2} \frac{\partial^2}{\partial \theta^2} \psi(\tilde{\theta}, 0)\right) + o(n)\right) \\
&= \ \exp\left(-n\theta\epsilon + n\frac{\theta^2}{2} \frac{\partial^2}{\partial \theta^2} \psi(\tilde{\theta}, 0) + o(n)\right),
\end{aligned}
$$

where $\tilde{\theta}$ lies between 0 and θ and $o(b_n)$ represents a sequence $(a_n : n \geq 1)$ for which $a_n/b_n \to 0$ as $n \to \infty$. By choosing $\theta > 0$ sufficiently small, we can then ensure that $\theta \epsilon > \frac{\theta^2}{2} \frac{\partial^2}{\partial \theta^2} \psi(\tilde{\theta}, 0)$, thereby yielding (3). ∎

For $c \geq 0$, let
$$N(c) = \max\{n \geq 0 : T_n \leq c\},$$

so that $N(c)$ is the number of observations generated within a computer time budget of magnitude c. Furthermore,

$$
\begin{aligned}
\alpha(c) &= \bar{X}_{N(c)} \text{ if } N(c) \geq 1, \\
&= 0 \text{ if } N(c) = 0
\end{aligned}
$$

is the estimator available after c units of computer time have been expended. In the presence of Proposition 1's almost sure convergence, it is easily verified that $\alpha(c) \Rightarrow \alpha$ as $c \to \infty$ under **A1** (and, in fact, $\alpha(c) \to \alpha$ a.s. as $c \to \infty$ under **A1**).

Our concern, in this section, rests on identifying the constant η^* for which

$$\frac{1}{c} \log P(|\alpha(c) - \alpha| > \epsilon) \to -\eta^* \tag{5}$$

as $c \to \infty$; we refer to (5) as a "rough asymptotic", since the asymptotic involves the logarithm of the probability rather than the probability itself.

A2. Given $a \in \Re$, assume that there exists $(\theta^*, \eta^*) = (\theta^*(a), \eta^*(a))$ for which

i.) $(\theta^*, -\theta^*, \eta^*) \in \mathcal{N}$;

ii.) $\psi(\theta^*, \eta^*) = \theta^* a$;

iii.) $\frac{\partial}{\partial \theta} \psi(\theta^*, \eta^*) = a$;

iv.) $(0, 0, \eta^*) \in \mathcal{N}$.

We will now offer a heuristic argument to justify the approximations

$$P(\alpha(c) \geq a) \approx \exp(-\eta^*(a)c) \tag{6}$$

for $a > \alpha$ and

$$P(\alpha(c) \leq a) \approx \exp(-\eta^*(a)c) \tag{7}$$

for $a < \alpha$. In view of (6) and (7), this then suggests the desired rough asymptotic

$$\frac{1}{c} \log P(|\alpha(c) - \alpha| > \epsilon) \to -\min(\eta^*(\alpha + \epsilon), \eta^*(\alpha - \epsilon)) \triangleq -I(\epsilon) \tag{8}$$

as $c \to \infty$. The heuristic argument proceeds as follows: For $a > \alpha$,

$$\begin{aligned}
P(\alpha(c) \geq a) &= \sum_{n=0}^{\infty} P(\bar{X}_n \geq a, N(c) = n) \\
&\approx \max_{n \geq 0} P(S_n \geq na, T_n \approx c) \\
&\approx \max_{n \geq 0} \exp(-\theta(n)na - \eta(n)c + n\psi(\theta(n), \eta(n))) \\
&\approx \max_{t \geq 0} \exp(-\theta(t)ta - \eta(t)c + t\psi(\theta(t), \eta(t)))
\end{aligned}$$

where $\theta(t)$ and $\eta(t)$ are defined as the roots of $\frac{\partial}{\partial \theta} \psi(\theta(t), \eta(t)) = a$ and $\frac{\partial}{\partial \eta} \psi(\theta(t), \eta(t)) = c/t$, and the second last line applies the standard Gartner-Ellis large deviations approximation (see, for example, Dembo and Zeitouni 1998, pg 44). At the maximizing t^*, the derivative of the exponent with respect to t should vanish:

$$\begin{aligned}
&- \quad \theta'(t^*)t^*a - \theta(t^*)a - \eta'(t^*)a + \psi(\theta(t^*), \eta(t^*)) + t^* \frac{\partial}{\partial \theta} \psi(\theta(t^*), \eta(t^*))\theta'(t^*) \\
&+ \quad t^* \frac{\partial}{\partial \eta} \psi(\theta(t^*), \eta(t^*))\eta'(t^*) = 0.
\end{aligned} \tag{9}$$

Because $\theta(t^*)$ and $\eta(t^*)$ are defined as roots of $\frac{\partial}{\partial \theta} \psi(\theta(t), \eta(t)) = a$ and $\frac{\partial}{\partial \eta} \psi(\theta(t), \eta(t)) = c/t$, (9) simplifies to

$$-\theta(t^*)a + \psi(\theta(t^*), \eta(t^*)) = 0. \tag{10}$$

Hence, if we choose $\theta(t^*) = \theta^*$ and $\eta(t^*) = \eta^*$, where θ^* and η^* satisfy **A2** $ii.$) and $iii.$), (10) follows and we obtain (6).

We note that in **A2**, condition $i.$) is a technical condition that ensures that $\tilde{\psi}$ (and hence ψ) is finite-valued and twice continuously differentiable at $(\theta^*, -\theta^*, \eta^*)$ (so that conditions $ii.$) and $iii.$) are meaningful). To explain the presence of condition $iv.$), observe that

$$\begin{aligned}
P(\alpha(c) \geq a) &\geq P(\alpha(c) \geq a, N(c) = 1) \\
&= P(X_1 \geq a, T_1 < c < T_2).
\end{aligned}$$

Note that $iv.$) will be violated if

$$\underline{\lim}_{c \to \infty} \frac{1}{c} \log P(X_1 \geq a, T_1 \leq c < T_2) > -\eta^*(a). \tag{11}$$

7

Since typically $\{X_1 > a\}$ has positive probability and T_2 has a heavier tail than T_1, (11) will usually hold whenever

$$\underline{\lim}_{c \to \infty} \frac{1}{c} \log P(T_2 \geq c) > -\eta^*(a). \tag{12}$$

A natural condition to preclude (12) is to require $iv.$). Not surprisingly, this condition also plays a role in the exact asymptotic that is developed in Section 3.

Given the key role that $\theta^*(a)$ and $\eta^*(a)$ play in our results, it is worth noting that $\theta^*(a) > 0$ for $a > \alpha$ and $\theta^*(a) < 0$ for $a < \alpha$, while $\eta^*(a) > 0$ for $a \neq \alpha$. To see this, note that $(\lambda\theta^*(a), \lambda\eta^*(a)) \in \mathcal{N}$ for $0 \leq \lambda \leq 1$. It follows that

$$H(\lambda) \triangleq \psi(\lambda\theta^*(a), \lambda\eta^*(a)) - \theta^*(a)a\lambda$$

is a continuously differentiable strictly convex function for $\lambda \in [0, 1]$. Furthermore, $H(0) = H(1) = 0$. The strict convexity then implies that $H'(0) < 0 < H'(1)$ and hence

$$H'(1) = \theta^*(a)\frac{\partial}{\partial\theta}\psi(\theta^*(a), \eta^*(a)) + \eta^*(a)\frac{\partial}{\partial\eta}\psi(\theta^*(a), \eta^*(a)) - \theta^*(a)a > 0,$$

so that $\eta^*(a)\frac{\partial}{\partial\eta}\psi(\theta^*(a), \eta^*(a)) > 0$. Since $\frac{\partial}{\partial\eta}\psi(\theta, \eta) > 0$ for $(\theta, \eta) \in \mathcal{N}$, we conclude that $\eta^*(a) > 0$. On the other hand,

$$H'(0) = \theta^*(a)(\alpha - a) + \eta^*(a)\frac{\partial}{\partial\eta}\psi(0, 0) < 0,$$

so that the positivity of $\eta^*(a)\frac{\partial}{\partial\eta}\psi(0, 0)$ and negativity of $\alpha - a$ requires that $\theta^*(a) > 0$ for $a > \alpha$.

Theorem 1 *Assume* $\mathbf{A1} - \mathbf{A2}$. *Then,*

$$\frac{1}{c}\log P(\alpha(c) \geq a) \to -\eta^*(a)$$

as $c \to \infty$ *when* $a > \alpha$ *and*

$$\frac{1}{c}\log P(\alpha(c) \leq a) \to -\eta^*(a)$$

as $c \to \infty$ *when* $a < \alpha$.

   Proof: We prove the result for $a > \alpha$; the proof for $a < \alpha$ is (very) similar. We start by establishing an upper bound. For $\epsilon > 0$, $\rho > 0$ and m sufficiently large,

$$
\begin{aligned}
P(\alpha(c) \geq a) &= \sum_{n=1}^{\infty} P(S_n \geq na, N(c) = n) \\
&= \sum_{n=1}^{\infty} P(S_n \geq na, T_n \leq c < T_{n+1}) \\
&\leq \sum_{n=1}^{\infty} P(S_n \geq na, T_{n+1} > c) \\
&\leq mP(T_m > c) + \sum_{n=m}^{\infty} P(S_n \geq na, T_{n+1} > c) \\
&= mP(T_m > c) + \sum_{n=m}^{\infty} P(S_n \geq na, S_{n+1} \geq (n+1)(a-\epsilon), T_{n+1} > c) \\
&\quad + \sum_{n=m}^{\infty} P(S_n \geq na, S_{n+1} < (n+1)(a-\epsilon), T_{n+1} > c) \\
&\leq mP(T_m > c) + \sum_{n=m}^{\infty} P(S_{n+1} \geq (n+1)(a-\epsilon), T_{n+1} > c) \\
&\quad + \sum_{n=m}^{\infty} P(X_{n+1} < -(n+1)\epsilon + a) \\
&\leq mP(T_m > c) + \sum_{n=m}^{\infty} P(S_{n+1} \geq (n+1)(a-\epsilon), T_{n+1} > c) \\
&\quad + \sum_{n=m}^{\infty} P(X_{n+1}^- > (n+1)\epsilon - a) \\
&\leq mP(T_m > c) + \sum_{n=m}^{\infty} P(S_{n+1} \geq (n+1)(a-\epsilon), T_{n+1} > c) \\
&\quad + \sum_{n=m}^{\infty} P(S_{\lfloor n+\rho n \rfloor}^- - S_n^- > (n+1)\epsilon - a).
\end{aligned}
\tag{13}
$$

By **A2** *iv.*), there exists $\delta > 0$ for which $(0, 0, \eta^* + \delta) \in \mathcal{N}$. Markov inequality then implies that

$$
\begin{aligned}
mP(T_m > c) &\leq m\exp(-(\eta^* + \delta)c)E\exp((\eta^* + \delta)T_m) \\
&= \exp(-(\eta^* + \delta)c + m\tilde{\psi}(0, 0, \eta^* + \delta) + o(m)).
\end{aligned}
$$

If we choose $m = \lfloor rc \rfloor$ with $r = \frac{\delta}{2\tilde{\psi}(0,0,\eta^*+\delta)}$, then

$$
\begin{aligned}
mP(T_m > c) &= \exp(-(\eta^* + \delta/2)c + o(c)) \\
&= o(\exp(-\eta^* c))
\end{aligned}
\tag{14}
$$

as $c \to \infty$.

For the second term in (13), note that for $n \geq m$, and $\tilde{\epsilon} > 0$, $P(S_{n+1} \geq (n+1)(a-\epsilon), T_{n+1} > c)$

$$
\begin{aligned}
&\leq \quad \exp(-\theta^*(n+1)(a-\epsilon) - (\eta^* - \tilde{\epsilon})c) E \exp(\theta^* S_{n+1} + (\eta^* - \tilde{\epsilon}) T_{n+1}) \\
&= \quad \exp(-\theta^* n(a-\epsilon) - (\eta^* - \tilde{\epsilon})c + n\psi(\theta^*, \eta^* - \tilde{\epsilon}) + o(n)).
\end{aligned}
$$

Set $\gamma = \frac{\partial}{\partial \eta} \psi(\theta^*, \eta^*)$ and observe that $\gamma > 0$. A Taylor expansion of ψ about (θ^*, η^*) implies that

$$
\psi(\theta^*, \eta^* - \tilde{\epsilon}) = \psi(\theta^*, \eta^*) - \tilde{\epsilon}\gamma + \frac{\tilde{\epsilon}^2}{2} \frac{\partial^2}{\partial \eta^2} \psi(\theta^*, \eta_0)
$$

for some $\eta_0 \in (\eta^* - \tilde{\epsilon}, \eta^*)$. By virtue of **A2** $ii.$), $\sum_{n=m}^{\infty} P(S_{n+1} \geq (n+1)(a-\epsilon), T_{n+1} > c)$

$$
\begin{aligned}
&\leq \quad \exp(-(\eta^* - \tilde{\epsilon})c) \sum_{n=m}^{\infty} \exp\left(-n(\tilde{\epsilon}\gamma - \theta^*\epsilon) + n\frac{\tilde{\epsilon}^2}{2} \frac{\partial^2}{\partial \eta^2} \psi(\theta^*, \eta_0) + o(n)\right) \\
&= \quad \exp(-(\eta^* - \tilde{\epsilon})c) o(1), \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (15)
\end{aligned}
$$

provided that we choose $\tilde{\epsilon} > 0$ and $\epsilon > 0$ sufficiently small so that $\tilde{\epsilon}\gamma - \theta^*\epsilon > \tilde{\epsilon}^2 \frac{\partial^2}{\partial \eta^2} \psi(\theta^*, \eta_0)$.

On the other hand, for $\theta > 0$, $P(S_{\lfloor n+\rho n \rfloor}^- - S_n^- > (n+1)\epsilon - a)$

$$
\begin{aligned}
&\leq \quad \exp(-\theta((n+1)\epsilon - a)) E \exp(\theta(S_{\lfloor n+\rho n \rfloor}^- - S_n^-)) \\
&= \quad \exp(-\theta n\epsilon + \rho n \tilde{\psi}(0, \theta, 0) + o(n)).
\end{aligned}
$$

Choose $\theta = 2\eta^*/(r\epsilon)$ and $\rho = \epsilon/(r\tilde{\psi}(0, 2\eta^*/(r\epsilon), 0))$ (recall that **A1** $iii.$) ensures that $\tilde{\psi}(0, 2\eta^*/(r\epsilon), 0)$ is finite for any $\epsilon > 0$). Then, $\sum_{n=m}^{\infty} P(S_{\lfloor n+\rho n \rfloor}^- - S_n^- > (n+1)\epsilon - a)$

$$
\begin{aligned}
&\leq \quad \sum_{n=m}^{\infty} \exp(-2\eta^* n/r + \epsilon n/r + o(n)) \\
&= \quad o(\exp(-\eta^* c)) \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (16)
\end{aligned}
$$

as $c \to \infty$. Combining (13), (14), (15) and (16), we conclude that

$$
\overline{\lim_{c \to \infty}} \frac{1}{c} \log P(\alpha(c) \geq a) \leq -\eta^* + \tilde{\epsilon}.
$$

Since $\tilde{\epsilon}$ may be made arbitrarily small, this proves the required upper bound.

For the lower bound, observe that for $r_1 \le r_2$ and $\epsilon > 0$,

$$
\begin{aligned}
& P(\alpha(c) \ge a) \\
\ge\ & P(\bar{X}_{N(c)} \ge a, r_1 \le N(c) < r_2) \\
=\ & P(S_{N(c)} - aN(c) \ge 0, r_1 \le N(c) < r_2) \\
\ge\ & P(\min_{r_1 \le n < r_2} (S_n - na) \ge 0, r_1 \le N(c) < r_2) \\
\ge\ & P(S_{r_1} \ge r_1(a+\epsilon), \min_{0 \le n < r_2 - r_1} (S_{r_1+n} - (r_1+n)a) \ge 0, r_1 \le N(c) < r_2) \\
\ge\ & P(S_{r_1} \ge r_1(a+\epsilon), \min_{0 \le n < r_2 - r_1} (S_{r_1+n} - S_{r_1}) \ge -\epsilon r_1 + a(r_2 - r_1), r_1 \le N(c) < r_2) \\
\ge\ & P(S_{r_1} \ge r_1(a+\epsilon), r_1 \le N(c) < r_2) \\
& -P(\min_{0 \le n < r_2 - r_1} (S_{r_1+n} - S_{r_1}) < -\epsilon r_1 + a(r_2 - r_1)) \\
\ge\ & P(S_{r_1} \ge r_1(a+\epsilon), r_1 \le N(c) < r_2) - P(S_{r_2}^- - S_{r_1}^- > \epsilon r_1 - a(r_2 - r_1)) \\
=\ & P(S_{r_1} \ge r_1(a+\epsilon), T_{r_2} > c) - P(S_{r_1} \ge r_1(a+\epsilon), T_{r_1} > c) \\
& -P(S_{r_2}^- - S_{r_1}^- > \epsilon r_1 - a(r_2 - r_1)). \qquad (17)
\end{aligned}
$$

For $\delta > 0$, choose $r_1 = \lfloor c(1 - \delta\epsilon)/\gamma \rfloor$ and $r_2 = \lfloor c(1 + \delta\epsilon)/\gamma \rfloor$. The Gärtner-Ellis theorem and **A1, A2** imply that

$$
\lim_{r_1 \to \infty} \frac{1}{r_1} \log P(S_{r_1} \ge r_1(a+\epsilon), T_{r_1} \ge c) = -\sup_{\theta, \eta} \left(\theta(a+\epsilon) + \eta(\frac{\gamma}{1 - \delta\epsilon}) - \psi(\theta, \eta) \right) \qquad (18)
$$

and

$$
\lim_{r_1 \to \infty} \frac{1}{r_1} \log P(S_{r_1} \ge r_1(a+\epsilon), T_{r_2} \ge c) = -\sup_{\theta, \eta} \left(\theta(a+\epsilon) + \eta(\frac{\gamma}{1 - \delta\epsilon}) - \psi(\theta, \eta) - \frac{2\delta\epsilon}{1 - \delta\epsilon} \psi(0, \eta) \right). \qquad (19)
$$

Because the η's that achieve the supremum in the above two cases are positive and $\psi(0, \eta) > 0$, it follows that the quantity being maximized in (19) is strictly smaller than that in (18) for each $\eta > 0$. Hence,

$$
P(S_{r_1} \ge r_1(a+\epsilon), T_{r_1} > c) = o(P(S_{r_1} \ge r_1(a+\epsilon), T_{r_2} > c)) \qquad (20)
$$

as $c \to \infty$. From (19), it follows that

$$
\underline{\lim}_{c \to \infty} \frac{1}{c} \log P(S_{r_1} \ge r_1(a+\epsilon), T_{r_2} > c) = -\tilde{\theta}a/\gamma - \tilde{\eta} + \psi(\tilde{\theta}, \tilde{\eta})/\gamma + O(\epsilon) \qquad (21)
$$

where $\tilde{\theta}$ and $\tilde{\eta}$ are such that $\frac{\partial}{\partial \theta} \psi(\tilde{\theta}, \tilde{\eta}) = a + \epsilon$ and

$$
\frac{\partial}{\partial \eta} \psi(\tilde{\theta}, \tilde{\eta}) + \frac{2\delta\epsilon}{1 - \delta\epsilon} \frac{\partial}{\partial \eta} \psi(0, \tilde{\eta}) = \gamma
$$

(and $O(\epsilon)$ represents a function which when divided by ϵ remains bounded). Finally, for $\theta > 0$,

$$
P(S_{r_2}^- - S_{r_1}^- > \epsilon r_1 - a(r_2 - r_1)) \le \exp(-\theta(\epsilon r_1 - a(r_2 - r_1)) + (r_2 - r_1)\tilde{\psi}(0, \theta, 0) + o(r_2 - r_1)).
$$

11

Choose θ to equal $4\eta^*\gamma/\epsilon$ and $\delta = \min(\epsilon, 1/(4\epsilon), \frac{1}{8|a|}, \frac{\eta^*\gamma}{4\epsilon\tilde{\psi}(0, 4\eta^*\gamma/\epsilon, 0)})$ so that

$$P(S^-_{r_2} - S^-_{r_1} > \epsilon r_1 - a(r_2 - r_1)) = O(\exp(-3/2\eta^* c)) \tag{22}$$

as $c \to \infty$. Combining (17), (20), (21) and (22), we conclude that

$$\underline{\lim} \frac{1}{c} \log P(\alpha(c) \geq a) \geq -\frac{\tilde{\theta}}{\gamma}(a + \epsilon) - \tilde{\eta} + \frac{\psi(\tilde{\theta}, \tilde{\eta})}{\gamma} + O(\epsilon).$$

Letting $\epsilon \downarrow 0$ yields the required lower bound needed to complete the proof for $a > \alpha$. $\square$

Note that **A1** *iii.*) implies that when the (X_i, τ_i)'s are iid, both X_1^+ and X_1^- must have super-exponentially decaying tails, in the sense that for each $\theta > 0$, both

$$P(X_1^+ > x) = o(\exp(-\theta x))$$

and

$$P(X_1^- > x) = o(\exp(-\theta x))$$

as $x \to \infty$. A close examination of the proof of Theorem 1 shows that only the super-exponentiality of the tail of X_1^- is used to prove the logarithmic asymptotic for $P(\alpha(c) \geq a)$ for $a > \alpha$; the super-exponentiality of the tail of X_1^+ is then used to establish the limit for $P(\alpha(c) \leq a)$ for $a < \alpha$.

3   An Exact Asymptotic for Terminating Simulations

We consider here the special case in which the (X_i, τ_i)'s are iid, as occurs (for example) in the setting of terminating simulations. We assume that:

A3. (X_1, τ_1) has a (joint) bounded density.

In this context, $\psi(\theta, \eta) = \log E \exp(\theta X_1 + \eta \tau_1)$. Let $\mathcal{M} = \{(\theta, \eta) : \psi(\theta, \eta) < \infty\}$. In view of the iid structure, we can simplify **A1** and **A2** to the following hypothesis:

A4. There exists (θ^*, η^*) lying in the interior $\mathcal{M}^o$ of $\mathcal{M}$ for which $\theta^* > 0$ and $\eta^* > 0$ and:

  i.) $\psi(\theta^*, \eta^*) = \theta^* a$;

  ii.) $\frac{\partial}{\partial \theta} \psi(\theta^*, \eta^*) = a$;

  iii.) $(0, \eta^*) \in \mathcal{M}^o$.

Given **A4**, let P^* be the probability measure under which

$$P^*((X_i, \tau_i) \in A_i, 1 \le i \le n) = \prod_{i=1}^{n} E\left(I((X_i, \tau_i) \in A_i) \exp(\theta^* X_i + \eta^* \tau_i - \psi(\theta^*, \eta^*))\right)$$

for Borel sets $A_i \in \Re^2$, and let E^* denote the associated expectation operator. Then, the $(X_i, \tau_i)'s$ are iid under P^*, and $E^* X_1 = \frac{\partial}{\partial \theta} \psi(\theta^*, \eta^*) = a$ with $E^* \tau_1 = \frac{\partial}{\partial \eta} \psi(\theta^*, \eta^*) \triangleq \gamma$. Since, $N(c) + 1$ is an a.s. finite stopping time under both P and P^*, it follows that

$$P(\bar{X}_{N(c)} > a) = E^*\left(I(\bar{X}_{N(c)} > a) \exp(-\theta^* S_{N(c)+1} - \eta^* T_{N(c)+1} + (N(c)+1)\psi(\theta^*, \eta^*))\right).$$

Set $\tilde{X}_n = X_n - a$ and $\tilde{S}_n = S_n - na$. Observe that **A4** implies that $P(\bar{X}_{N(c)} > a, N(c) = n)$

$$
\begin{aligned}
&= P(\bar{X}_{N(c)} > a, T_n \le c < T_{n+1}) \\
&= E^*\left(I(\bar{X}_n > a, T_n \le c < T_{n+1}) \exp(-\theta^* S_{n+1} - \eta^* T_{n+1} + (n+1)\psi(\theta^*, \eta^*))\right) \\
&= E^*\left(I(\tilde{S}_n > 0, T_n \le c < T_{n+1}) \exp(-\theta^* \tilde{S}_{n+1} - \eta^* T_{n+1})\right) \\
&= \int_{[0,c]} \int_{(0,\infty)} P^*(\tilde{S}_n \in dx, T_n \in dt) \exp(-\theta^* x - \eta^* t) \\
&\quad \times E^*\left(I(\tau_{n+1} > c - t) \exp(-\theta^* \tilde{X}_{n+1} - \eta^* \tau_{n+1})\right) \\
&= e^{-\eta^* c} \int_{[0,c]} \phi(c-t) \int_{(0,\infty)} \exp(-\theta^* x) P^*(\tilde{S}_n \in dx, T_n \in dt) \qquad (23)
\end{aligned}
$$

where,

$$
\begin{aligned}
\phi(t) &= E^*\left(I(\tau_1 > t) e^{\eta^* t} \exp(-\theta^* \tilde{X}_1 - \eta^* \tau_1)\right) \\
&= E\left(I(\tau_1 > t) e^{\eta^* t} \exp(-\theta^*(X_1 - a) - \eta^* \tau_1 + \theta^* X_1 + \eta^* \tau_1 - \psi(\theta^*, \eta^*))\right) \\
&= e^{\eta^* t} E I(\tau_1 > t).
\end{aligned}
$$

Hence, we have the following identity:

$$P(\bar{X}_{N(c)} > a) = e^{-\eta^* c} \int_{[0,c]} \int_{(0,\infty)} f(x,t) R(dx, dt), \qquad (24)$$

where R is the (multivariate) renewal measure defined by

$$R(dx, dt) = \sum_{n=0}^{\infty} P^*(\tilde{S}_n \in dx, T_n \in dt)$$

and

$$f(x,t) = \exp(-\theta^* x)\phi(c-t).$$

To proceed further, we decompose $\int_{[0,c]} \int_{(0,\infty)} f(x,t) R(dx, dt)$ into

$$\int_{[0,c-c^\xi]} \int_{(0,\infty)} f(x,t) R(dx, dt) + \int_{[c-c^\xi, c]} \int_{(0,\infty)} f(x,t) R(dx, dt),$$

13

where $0 < \xi < 1/2$. First observe that

$$\lim_{c \to \infty} c^{1/2} \int_{[0,c-c^\xi]} \int_{(0,\infty)} f(x,t) R(dx, dt) = 0. \tag{25}$$

To see this, note that ($\mathbf{A4}$) iii.) implies the existence of $\delta > 0$ for which

$$P(\tau_1 > t) = O(e^{-(\eta^* + \delta)t})$$

as $t \to \infty$, so that $\phi(t) = O(e^{-\delta t})$ as $t \to \infty$. In particular, $f(x,t) \le O(e^{-\delta c^\xi})$ for $x \ge 0$ and $t \in [0, c - c^\xi]$. Hence, $\int_{[0,c-c^\xi]} \int_{(0,\infty)} f(x,t) R(dx, dt)$ is bounded from above by

$$O(e^{-\delta c^\xi}) E^*[N(c)]$$

and (25) follows from the elementary renewal theorem.

To asymptotically evaluate the key term $\int_{[c-c^\xi,c]} \int_{(0,\infty)} f(x,t) R(dx, dt)$, we will apply Theorem 1 of Keener (1990). This theorem describes the precise sense in which the following approximation of the renewal measure R holds:

$$R(dx, dt) \approx \frac{1}{\gamma} \tilde{r}(x|t) dx dt \tag{26}$$

for $x, t \ge 0$, where $\tilde{r}(\cdot|t)$ is the density of a normal rv with mean 0 and variance $t\sigma^2$, where $\sigma^2 = E^*(X_i - (E^*X/E^*\tau)\tau_1)^2 / E^*\tau_1$. In the presence of this approximation,

$$
\begin{aligned}
& c^{1/2} \int_{[c-c^\xi,c]} \int_{(0,\infty)} f(x,t) R(dx, dt) \\
\approx\ & c^{1/2} \int_{[c-c^\xi,c]} \int_{(0,\infty)} e^{-\theta^* x} \phi(c-t) \frac{1}{\gamma} \tilde{r}(x|t)\, dx\, dt \\
=\ & c^{1/2} \int_{[c-c^\xi,c]} \phi(c-t) \int_{(0,\infty)} \int_{(\theta^* x,\infty)} e^{-u} \frac{1}{\gamma} \tilde{r}(x|t)\, du\, dx\, dt \\
=\ & c^{1/2} \int_{[c-c^\xi,c]} \phi(c-t) \int_{(0,\infty)} e^{-u} \int_{(0,\infty)} I(0 \le x \le u/\theta^*) \frac{1}{\gamma} \tilde{r}(x|t)\, dx\, du\, dt \\
=\ & c^{1/2} \int_{[c-c^\xi,c]} \phi(c-t) \int_{(0,\infty)} e^{-u} P(0 \le N(0,\sigma^2)t \le u/\theta^*) \frac{du}{\gamma}\, dt \\
=\ & c^{1/2} \int_{[c-c^\xi,c]} \phi(c-t) \int_{(0,\infty)} e^{-u} P\left(0 \le N(0,1) \le \frac{u}{\sqrt{t}\sigma\theta^*}\right) \frac{du}{\gamma}\, dt. \\
=\ & \int_{[0,c^\xi]} \phi(t) \sqrt{\frac{c}{c-t}} \int_{(0,\infty)} e^{-u} \sqrt{c-t} P\left(0 \le N(0,1) \le \frac{u}{\sqrt{c-t}\sigma\theta^*}\right) \frac{du}{\gamma}\, dt \tag{27}
\end{aligned}
$$

Note that

$$e^{-u} \sqrt{t} P\left(0 \le N(0,1) \le \frac{u}{\sqrt{t}\sigma\theta^*}\right) \to \frac{u e^{-u}}{\sigma\theta^* \sqrt{2\pi}} \tag{28}$$

14

as $t \to \infty$, and the left-hand side of (28) is dominated (uniformly in t) by the limit appearing in the right-hand side of (28) (since the density of a $N(0,1)$ is bounded by $1/\sqrt{2\pi}$). Consequently, the Dominated Convergence Theorem implies that

$$\int_{(0,\infty)} e^{-u}\sqrt{t}P(0 \le N(0,1) \le \frac{u}{\sqrt{t}\sigma\theta^*})\frac{du}{\gamma} \to \frac{1}{\gamma\sigma\theta^*\sqrt{2\pi}}$$

as $t \to \infty$. Hence, the right-hand side of (27) converges to

$$\begin{aligned}
&\int_{(0,\infty)} \phi(t)dt \cdot \frac{1}{\gamma\sigma\theta^*}\frac{1}{\sqrt{2\pi}} \\
=\ & E\int_{(0,\infty)} e^{\eta^*t}I(\tau_1 > t)dt \cdot \frac{1}{\gamma\sigma\theta^*}\frac{1}{\sqrt{2\pi}} \\
=\ & E\int_{(0,\tau_1)} e^{\eta^*t}dt \cdot \frac{1}{\gamma\sigma\theta^*}\frac{1}{\sqrt{2\pi}} = \frac{1}{\eta^*\gamma\sigma\theta^*\sqrt{2\pi}}E(e^{\eta^*\tau} - 1).
\end{aligned} \tag{29}$$

The above calculation supports the following theorem.

Theorem 2 *Under* **A3** *and* **A4**,

$$P(\alpha(c) > a) \sim \frac{e^{-\eta^*c}}{\sqrt{c}} \cdot \frac{E(e^{\eta^*\tau} - 1)}{\eta^*\gamma\sigma\theta^*\sqrt{2\pi}}$$

as $c \to \infty$.

All that remains to prove Theorem 2 is verification of the validity of the approximation (26).

Proof: For $\epsilon > 0$, let

$$\omega_f(x,t,\epsilon) = \sup\{|f(x,t) - f(x',t')| : |x - x'| + |t - t'| < \epsilon\}$$

be the *oscillation function* associated with f. Straightforward (but tedious) calculations establish that

$$\omega_f(x,t,\epsilon) \le \epsilon e^{-\theta^*x - \delta(c-t)}O(1), \tag{30}$$

where $O(1)$ is bounded and δ is as in (27). Applying Theorem 1 of Keener (1990) guarantees the existence of $\nu > 0$ and $1/2 + \xi < p < 1$, such that,

$$\begin{aligned}
&\int_{[k,k+1]}\int_{(0,\infty)} f(x,t)R(dx,dt) \\
=\ & \int_{[k,k+1]}\int_{(0,\infty)} f(x,t)\tilde{r}(x|t)(1 + \frac{e(x)}{\sqrt{t}})\frac{dx}{\gamma}\,dt + o(k^{-p}) + o(k^{-p})\int_{(0,\infty)}[f(x,k) + \omega_f(x,k,e^{-\nu k})]\,dx \\
+\ & O(1)\int_{[k,k+1]}\int_{(0,\infty)} \omega_f(x,t,e^{-\nu k})\tilde{r}(x|t)(1 + \frac{e(x)}{\sqrt{t}})\,dx\,dt,
\end{aligned}$$

where $e(\cdot)$ is a second-degree (Hermite) polynomial. Below, we use this approximation for $c - c^\xi \le k \le c$. Because $e^{-\theta^*x}e(x)\tilde{r}(x|t)$ is integrable in x and $e(\cdot)$ is independent of t, steps

15

(26) through (29) continue to apply with $\tilde{r}(x|t)$ modified to $\tilde{r}(x|t)(1 + \frac{e(x)}{\sqrt{t}})$, for $t \in [c - c^\xi, c]$, without changing the final result of the calculation. Consequently,

$$
\int_{[c-c^\xi,c]} \int_{(0,\infty)} f(x,t) R(dx, dt)
$$
$$
= \int_{[c-c^\xi,c]} \int_{(0,\infty)} f(x,t) \tilde{r}(x|t) \frac{dx}{\gamma} \, dt
$$
$$
+ \sum_{c-c^\xi \le k \le c} O(k^{-p}) + \sum_{c-c^\xi \le k \le c} O(e^{-\delta(c-k)-\nu k})
$$
$$
= \int_{[c^\xi,c]} \int_{(0,\infty)} f(x,t) \tilde{r}(x|t) \frac{dx}{\gamma} \, dt + o(c^{-1/2})
$$

as $c \to \infty$ proving the required approximation.

The above proof explicitly identifies constants that are only implicitly developed in related work of Chi (2007) and Kuczek and Crank (1991), and offers a more direct proof of this result.

4 Concluding Remarks and Discussion

In this paper, we have argued that when one measures the quality of an estimator through the error probability rather than through the error tolerance, the natural figure of merit for comparing the efficiency of two competing estimators is the rate constant $I(\epsilon)$. In particular, when the error probability δ is sent to zero, the computational effort required to achieve that error probability is asymptotic to $(1/I(\epsilon)) \log(1/\delta)$. We now discuss some of the implications of the theory we have developed in this paper:

1. When competing estimators are available, the choice of "best estimator" depends upon whether one is interested in small error probabilities (in which case $I(\epsilon)$ is the relevant figure of merit) or small error tolerances (in which case η^2 is the relevant figure of merit). In general, the two figures of merit do not lead to identical choices about how to select the best estimator. For example, suppose that X is a rv for which we wish to compute $\alpha = EX$ via simulation. One means of computing α is by generating iid copies of X. But we might also decide to use another zero mean rv Z as a control variate, leading to an estimator based on iid replication of the rv $X - \lambda Z$, where λ is the control coefficient to be optimized. Assuming (for simplicity) that the computer time required to compute both estimators is deterministic and equal to 1, the conventional variance-based figures of merit for the two estimators are, respectively, $\eta^2 = Var(X)$ and $\eta^2 = Var(X) - 2\lambda \, Cov(X,Z) + \lambda^2 Var(Z)$.

To compute our large deviations-based figure of merit, suppose that $\psi(\theta, \eta) = \log E \exp(\theta X + \eta Z)$ is the joint log moment generating function of X and Z. For each fixed value of λ, let $\alpha(c, \lambda)$ be the corresponding control variate estimator associated with computational budget c. Note that under appropriate regularity conditions,

$$\frac{1}{c} \log P(\alpha(c, \lambda) > \alpha + \epsilon) \to -\theta(\lambda)(\alpha + \epsilon) + \psi(\theta(\lambda), -\theta(\lambda)\lambda)$$

as $c \to \infty$, where $\theta(\lambda)$ is the root of the equation

$$\frac{\partial}{\partial \theta} \psi(\theta, -\theta\lambda) - \lambda \frac{\partial}{\partial \eta} \psi(\theta, -\theta\lambda) = \alpha + \epsilon. \tag{31}$$

Suppose that we wish to select λ to minimize the one-sided error probability $P(\alpha(c, \lambda) > \alpha + \epsilon)$. When this error probability needs to be small, this requires that we minimize $\theta(\lambda)(\alpha + \epsilon) + \psi(\theta(\lambda), -\theta(\lambda)\lambda)$ over λ. (Note that when the associated one-sided error tolerance needs to be small, we continue to minimize $Var(X) - 2\lambda \, Cov(X, Z) + \lambda^2 Var(Z)$). The derivative of this function should vanish at minimizer λ^*, and hence

$$\theta'(\lambda^*)(\alpha + \epsilon) + \frac{\partial}{\partial \theta} \psi(\theta(\lambda^*), -\theta(\lambda^*)\lambda^*)\theta'(\lambda^*) + \frac{\partial}{\partial \eta} \psi(\theta(\lambda^*), -\theta(\lambda^*)\lambda^*)[-\theta'(\lambda^*)\lambda^* - \theta(\lambda^*)] = 0.$$

Taking advantage of (31) shows that the pair $(\theta(\lambda^*), \lambda^*)$ should satisfy the two equations

$$\frac{\partial}{\partial \theta} \psi(\theta(\lambda^*), -\theta(\lambda^*)\lambda^*) = \alpha + \epsilon$$

and

$$\frac{\partial}{\partial \eta} \psi(\theta(\lambda^*), -\theta(\lambda^*)\lambda^*) = 0.$$

(A curious aspect of this calculation is that it implies that, conditional on the rare event $\{\alpha(c, \lambda) > \alpha + \epsilon\}$, the optimizing control variates estimator has the property that the control continues to have zero mean.) It is easily seen that it is almost never the case that the λ^* minimizing the one-sided error probability coincides with the minimizing λ for the one-sided error tolerance. (A notable exception is the case where Z is independent of X, in which case both minimizers coincide at $\lambda^* = 0$.) A full analytical study of the two-sided error probability is more complicated, because it involves minimizing the maximum of the two one-sided error probability functions. Nevertheless, we can easily argue, by example, that choosing λ to minimize $-I(\epsilon)$ leads to a different choice of "best λ" than when minimizing $Var(X) - 2\lambda Cov(X, Z) + \lambda^2 Var(Z)$. Suppose, for example, that X is exponentially distributed with mean 1 and $Z = X - Y$, where Y is independent and has a Gaussian distribution with mean 1 and variance 2. In this case, the variance-minimizing value of λ is $\lambda^* = 1/3$. However, it can be seen, via calculation, that $-I(\epsilon)$ takes on a smaller value at $\lambda = 0.4$ than at $\lambda = 1/3$ when $\epsilon = 0.2$.

2. The above example also illustrates the fact that the large deviations-based choice of best estimator depends on the value of the error tolerance ϵ. An estimator that is preferable for one choice of ϵ may be suboptimal for another choice of ϵ. This is a complication that does not arise in the setting of the conventional variance-based figure of merit. In particular, observe that when the error probability δ is fixed, the computational effort required to achieve error tolerance ϵ, namely $c = z(\delta)2\eta^2/\epsilon^2$, has the characteristic that if one estimator is to be preferred for one value of δ, it is to be preferred for all values of δ. One implication is therefore that constructing estimators with a small error probability leads to subtleties that do not arise in the setting of small error tolerances.

3. In addition, the formulae for the optimizing choice of control coefficient are more complex in the setting of small error probabilities than in the small error tolerance context, and the formulae depend on "exponential moments" of the underlying random variables rather than second moments (as for η^2). Typically, the sample sizes necessary to accurately estimate exponential moments will be much larger than those needed to estimate second moments, so that estimating the optimizing control coefficient from a set of trial runs will require much larger sample sizes for our large deviations-based figure of merit than for the conventional figure of merit.

4. An important simulation area for which the associated theory frequently rests on asymptotic analysis of small error probabilities is the "selection of best system" literature. In the "best system" setting, one wants to build algorithms for which the probability that one has incorrectly eliminated the actual best system from the set of remaining alternatives is (very) small. This leads naturally to an algorithmic development in which the large deviations-based figure of merit discussed in this paper plays a key role. As a consequence, there are a number of papers on the selection of best system problem that have appeared in recent years in which large deviations plays a key role (see, for example, Dai and Chen 1997, Chen et. al. 2000, Glynn and Juneja 2004, Szechtman and Yucesan 2008, Hunter and Pasupathy 2010). However, these papers have all implicitly assumed that the time τ_i to generate X_i is a.s. equal to 1. Thus, these papers have used the large deviations rate function associated with the simulation outputs from each system, but ignored the contribution to $I(\epsilon)$ due to the randomness induced by the time needed to generate such samples. Our results in this paper correct this oversight and provide the appropriate figure of merit $I(\epsilon)$ that reflects the random behavior of the τ_i's that arises in actual "selection of best system" problems.

As the above discussion makes clear, a principle conclusion of this paper is that the design of good simulation algorithms intended to provide small error probabilities is inherently more challenging than design that is intended to produce small error tolerances. The development of algorithmic tools appropriate to the design of such "small error probability" computations is a fertile area for future research.

References

CHEN, C.H., J. LIN, E. YUCESAN AND S. E. CHICK. 2000. Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Journal of Discrete Event Dynamic Systems: Theory and Applications*, **10**, 251-270.

CHI, Z. 2007. Uniform convergence of exact large deviations for renewal reward process, *Annals of Applied Probability* **17**, 1019–1048.

DAI, L., and CHEN, C.H. 1997. Rate of convergence for ordinal comparison of dependent simulations in discrete event dynamic systems. *Journal of Optimization Theory and Applications* **94**, 29-54.

DEMBO, A. and ZEITOUNI, O. 1998. *Large Deviations Techniques and Applications*, Jones and Bartlett Publishers. Second edition.

GLYNN, P. W. AND S. JUNEJA. 2004. A large deviations perspective on ordinal optimization. Proceedings of the 2004 Winter Simulation Conference, IEEE Press. 577-585.

GLYNN, P. W. AND S. JUNEJA. 2008. A large deviations view of asymptotic efficiency for simulation estimators. Proceedings of the 2008 Winter Simulation Conference, IEEE Press. 396-406.

HUNTER, R. S., AND R. PASUPATHY. 2010. Large-deviation sampling laws for constrained simulation optimization on finite sets. Proceedings of the 2010 Winter Simulation Conference, IEEE Press. 995-1002.

KUCZEK, T., AND K. N. CRANK. 1991. A large deviations result for regenerative processes. *J. Theoretical Prob.*, **4**, 551-561.

GLYNN, P.W., AND W. WHITT. 1992. The asymptotic efficiency of simulation estimators. *Operations Research* **40**, 3, 505-520.

HAMMERSLEY, J. M., AND D. C. HANDSCOMB. 1964. *Monte Carlo Methods.* Methuen, London.

ISCOE, I., P. NEY, AND E NUMMELIN. 1985. Large deviations of uniformly recurrent Markov additive processes. *Adv. Appl. Math.* 6, 4, 373-412.

KEENER, R. 1990. Asymptotic expansions in multivariate renewal theory. *Stochastic Pro-*

cesses and Their Applications **34**, 137-153.

SZECHTMAN, R., AND E. YUCESAN. 2008. A new perspective on feasibility determination. Proceedings of the 2008 Winter Simulation Conference, IEEE Press. 273-280.