The Annals of Applied Probability
1997, Vol. 7, No. 3, 731–746

COUNTEREXAMPLES IN IMPORTANCE SAMPLING FOR LARGE DEVIATIONS PROBABILITIES[1]

BY PAUL GLASSERMAN AND YASHAN WANG

Columbia University

A guiding principle in the efficient estimation of rare-event probabilities by Monte Carlo is that importance sampling based on the change of measure suggested by a large deviations analysis can reduce variance by many orders of magnitude. In a variety of settings, this approach has led to estimators that are optimal in an asymptotic sense. We give examples, however, in which importance sampling estimators based on a large deviations change of measure have provably poor performance. The estimators can have variance that decreases at a slower rate than a naive estimator, variance that increases with the rarity of the event, and even infinite variance. For each example, we provide an alternative estimator with provably efficient performance. A common feature of our examples is that they allow more than one way for a rare event to occur; our alternative estimators give explicit weight to lower probability paths neglected by leading-term asymptotics.

1. Introduction. Among the most dramatic successes in variance reduction for efficient Monte Carlo simulation are applications of importance sampling for estimating rare-event probabilities. It has been shown in a variety of settings that a large deviations analysis of a rare event often suggests a particularly effective change of measure—one that is often optimal in an asymptotic sense. Siegmund [30] gave perhaps the first result of this type. He showed that the uniquely optimal exponential change of measure for estimating a gambler's ruin probability is determined by the exponential rate of decay of the probability as one of the boundaries recedes. The subsequent literature can be roughly divided in two: results showing that specific estimators have provably good performance, and the development of estimators—often evaluated experimentally—suggested by, but not strictly supported by, rare-event asymptotics. While the first type of result may be more mathematically complete, the second type is also essential, because in applications it is necessary to have good heuristics for developing estimators even if a rigorous analysis of their variance may be inaccessible.

The purpose of this article is to show, however, that there is some risk in assuming that an importance sampling estimator based on a large deviations analysis of a rare event is sure to perform well. Large deviations seeks to

Received January 1996; revised March 1997.

[1]Research partially supported by NSF Grants DMI-94-57189 and DMS-95-08709 and IBM Corporation.

AMS 1991 *subject classifications*. 60F10, 60J15, 65C05.

Key words and phrases. Monte Carlo methods, rare events, random walks, large deviations, importance sampling, simulation.

identify the most likely path and this may have adverse effects on the variance. Simply put, an analysis of a first moment cannot be expected to carry a guarantee about the behavior of a second moment. General necessary and sufficient conditions in [28], [12] and [27] point out the requirements beyond a large deviations limit that must be verified to ensure effective variance reduction. Our objective is to show just how poorly seemingly optimal estimators can perform when these types of conditions do not hold. Glasserman and Kou [16] make a related point in a queueing-network application, but the observation there is qualified by the fact that a complete large deviations analysis of the process is unavailable. The examples we consider here can be more thoroughly analyzed and lead to much stronger negative conclusions.

2. Problem setting. A reasonably generic formulation of the problem we consider starts with an indexed family of events $\{A_x,\ x > 0\}$ satisfying either a logarithmic limit

$$\lim_{x\to\infty} \frac{1}{x} \log P(A_x) = -\gamma, \tag{1}$$

for some $\gamma > 0$, or the stronger exponential asymptotic

$$P(A_x) \sim Ce^{-\gamma x}, \tag{2}$$

for some constant $C > 0$. An even more general formulation along the lines of the usual statement of a large deviations principle would consider an indexed sequence of probability measures, but the simpler case suffices for our purpose. [Asmussen and Binswanger [5] consider rare-event simulation in a heavy-tailed setting where not even (1) holds and a different formulation is therefore necessary.]

To estimate $\alpha(x) \triangleq P(A_x)$, straightforward simulation generates independent replications of the indicator $\mathbf{1}_{A_x}$ or something that approximates it. The variance of this estimator is $\alpha - \alpha^2$. If $P(A_x) \to 0$ then this variance approaches 0, but the *relative error* of the estimator (the ratio of its standard deviation to its mean) satisfies

$$\text{relative error} = \frac{\sqrt{\alpha(x) - \alpha^2(x)}}{\alpha(x)} \geq \frac{1}{\sqrt{\alpha(x)}} \to \infty.$$

If (1) holds, the increase to infinity is exponential. This further implies that the number of independent replications of $\mathbf{1}_{A_x}$ required to achieve a fixed relative error grows exponentially in x.

Importance sampling generates samples under a different measure $\overline{P}$ with expectation operator $\overline{E}$ and uses the representation

$$P(A_x) = \overline{E}[L_x; A_x] \triangleq \overline{E}[L_x \mathbf{1}_{A_x}], \tag{3}$$

in which L_x is a likelihood ratio. More precisely, L_x is the likelihood ratio of a restriction of $\overline{P}$ to a restriction of P, the restriction being to a sub-σ-algebra containing A_x on which $\bar{P} \gg P$. Often, absolute continuity of the unrestricted measures fails; this is the case in our examples and indeed in all examples where switching from P to $\bar{P}$ changes the common distribution of an infinite i.i.d. sequence.

Based on (3), we obtain an unbiased estimator of $P(A_x)$ by averaging independent replications of

$$\hat{\alpha}(x) = L_x \mathbf{1}_{A_x}$$

generated under $\bar{P}$. Assuming the computational requirements of the estimators are roughly comparable, the new one is better than the straightforward one if its standard deviation is smaller, at least for large x. If (1) holds, then nonnegativity of variance implies

$$\limsup_{x \to \infty} \left\{ -\frac{1}{x} \log \bar{E}\left[\hat{\alpha}^2(x) \right] \right\} \le 2\gamma,$$

meaning that the exponential rate of decrease of the second moment can be at most twice that of the probability itself. An estimator is called *asymptotically efficient* (or *asymptotically optimal*) if this maximum rate is achieved—that is, if

$$\lim_{x \to \infty} \left\{ -\frac{1}{x} \log \bar{E}\left[\hat{\alpha}^2(x) \right] \right\} = 2\gamma. \tag{4}$$

An estimator is said to have *bounded relative error* if it satisfies the stronger requirement that

$$\limsup_{x \to \infty} \frac{\sqrt{\bar{E}\left[\hat{\alpha}^2(x) \right]}}{P(A_x)} < \infty. \tag{5}$$

The number of independent replications required to achieve a fixed relative error grows at a subexponential rate under (4) and remains bounded under (5), no matter how small $P(A_x)$. (See [27, 26] for efficiency criteria sensitive to all moments of an estimator.) Specific estimators suggested by rare-event asymptotics have been shown to satisfy these types of properties in, for example, [3], [6], [8] (Section 8.B), [9], [10] (Section 2), [11], [16], [17], [19], [21], [23], [25] and [29]. In, for example, [8] (Section 8.C), [10] (Section 3), [1], [14], [20], [22] and [24], estimators are proposed based on large deviations or related asymptotics for probabilities but without a corresponding analysis for second moments. See [18] for further background on (4) and (5) and a survey of rare-event simulation.

Since our objective is to show that estimators suggested by asymptotics of $P(A_x)$ may fail to have desirable properties (and may even be inferior to straightforward estimators), we need to make precise the mechanism by which an estimator is suggested. One can imagine quite a few senses in which a change of measure is consistent with a large deviations result, and it seems unlikely that any one sense can meaningfully cover all settings.

Perhaps the most attractive notion is the existence of a conditioned limit theorem showing roughly that P conditioned on A_x converges to $\bar{P}$ as $x \to \infty$: simulating under the conditional law produces a zero-variance estimator. Asmussen [2] has established a conditioned limit theorem for random walks and queues, and this forms an important underpinning of the associated theory for simulation. A related notion is the existence of a large deviations result for an empirical measure on paths in A_x, generalizing Sanov's theorem (see [13]). We work with two apparently weaker notions, intended to accompany (1) and (2), respectively. We say that $\bar{P}$ is consistent with the large deviations result for $P(A_x)$ if $\liminf_{x\to\infty} \bar{P}(A_x) > 0$ and, under $\bar{P}$, either

$$-\frac{1}{x}\log L_x \to \gamma \quad \text{in probability} \tag{6}$$

or

$$e^{\gamma x} L_x \text{ has a proper limiting distribution,} \tag{7}$$

where *proper* excludes distributions putting mass on $\pm\infty$. These are not intended to cover all cases, but rather to provide *some* concrete link between a large deviations result and a proposed change of measure. Each of these suggests that P conditioned on A_x is approximated by $\bar{P}$ because the likelihood ratio is not too spread out: the conditional law itself has constant likelihood ratio with respect to P on A_x. In some settings, it would be preferable to multiply by the indicator $\mathbf{1}_{A_x}$ in (7); but in the examples of Section 4 we have $\bar{P}(A_x) = 1$ for all x, and in Section 3 $\bar{P}$ will be at the boundary of a parametric family of measures under which the probability of A_x converges to 1. One could argue that stronger notions of "consistent with large deviations" would be preferable. But in applications stronger notions may be impossible to verify, and in practice one often finds that no precise link is made between rare-event asymptotics and a suggested importance sampling measure.

Section 3 presents an example based on logarithmic asymptotics in which an estimator consistent with (6) turns out to have infinite variance. Section 4 gives examples based on exponential asymptotics in which estimators consistent with (7) have variances that increase to infinity. In each case, we present an alternative estimator that has bounded relative error or is at least asymptotically efficient. Proofs are deferred to Section 5.

3. Tail probabilities for sums of random variables. Let $X_1, X_2, \ldots$ be i.i.d. with cumulant generating function (c.g.f.)

$$\psi(\theta) = \log E[e^{\theta X_1}]$$

whose domain has nonempty interior. Suppose there is an $a > 0$ with $|EX_1| < a$ for which there are solutions θ_a and θ_{-a} in the interior of the domain of ψ to the equations $\psi'(\theta_a) = a$ and $\psi'(\theta_{-a}) = -a$. Let $S_n = \sum_{i=1}^n X_i$, $n = 1, 2, \ldots$. It is well known from Cramér's theorem (see [13], Corollary 2.2.19) that

$$-\lim_{n\to\infty}\frac{1}{n}\log P(S_n \geq an) = \theta_a a - \psi(\theta_a) \triangleq \gamma_a \tag{8}$$

and

$$-\lim_{n\to\infty}\frac{1}{n}\log P(S_n \leq -an) = -\theta_{-a} a - \psi(\theta_{-a}) \triangleq \gamma_{-a}. \tag{9}$$

Efficient estimation of $P(S_n \geq an)$ and its generalizations are among the most extensively studied applications of importance sampling to large deviations probabilities; see, in particular, [9] and [26]. The unique asymptotically efficient change of measure twists the common distribution of the X_i by θ_a: by definition

$$P_{\theta_a}(X_1 \leq x) = E\big[\exp(\theta_a X_1 - \psi(\theta_a)); X_1 \leq x\big],$$

and under P_{θ_a} the X_i are still i.i.d. (The measures P_{θ_a} and P are mutually absolutely continuous when restricted to the σ-algebra generated by $\{X_1, \dots, X_n\}$, though not on that generated by the infinite sequence $\{X_1, X_2, \dots\}$.) Thus,

$$P(S_n \geq an) = E_{\theta_a}\big[\exp(-\theta_a S_n + n\psi(\theta_a)); S_n \geq an\big].$$

We verify below that if $\gamma_a < \gamma_{-a}$ then P_{θ_a} is consistent with the large deviations behavior of $P(|S_n| \geq an)$, so the obvious extension of the method above generates independent replications of

$$\hat{\alpha}(n) = \exp(-\theta_a S_n + n\psi(\theta_a))\mathbf{1}_{\{|S_n| \geq an\}} \triangleq L_n \mathbf{1}_{\{|S_n| \geq an\}} \tag{10}$$

under P_{θ_a}. An alternative estimator generates, at each replication,

$$\hat{\beta}_+(n) = \exp(-\theta_a S_n + n\psi(\theta_a))\mathbf{1}_{\{S_n \geq an\}}$$

under P_{θ_a} and

$$\hat{\beta}_-(n) = \exp(-\theta_{-a} S_n + n\psi(\theta_{-a}))\mathbf{1}_{\{S_n \leq -an\}}$$

independently under $P_{\theta_{-a}}$ and sums them to get $\hat{\beta}(n)$. We analyze the estimators in the following two results. Their proofs are deferred to Section 5.

PROPOSITION 1. *Suppose $\gamma_a < \gamma_{-a}$. Then we have the following*:

(i) $-n^{-1}\log P(|S_n| \geq an) \to \gamma_a$;
(ii) L_n *in* (10) *satisfies* (6) *with* $\gamma = \gamma_a$;
(iii) $\hat{\beta}(n)$ *is asymptotically efficient.*

The next theorem analyzes the performance of $\hat{\alpha}(n)$, the estimator suggested by the leading-term asymptotics of $P(|S_n| \geq an)$. The performance of this estimator can be very poor because even though P_{θ_a} gives $\{S_n \leq -an\}$ little probability, L_n can be very large on this set.

THEOREM 1. *Suppose* $\gamma_a < \gamma_{-a}$.

(i) *If* $\psi(-\theta_a) = \infty$, *then* $\hat{\alpha}(n)$ *has infinite variance.*

(ii) *If* $\theta_a + \theta_{-a} \geq 0$ *and* $\psi(-\theta_a) < \infty$, *then* $E_{\theta_a}[\hat{\alpha}^2(n)] \to \infty$.

(iii) *If* $\theta_a + \theta_{-a} < 0$ *then* $-(1/n)\log E_{\theta_a}[\hat{\alpha}^2(n)] \to \min\{2\gamma_a, -\psi(\theta_a) - \psi(\theta_{-a}) - a(\theta_a + \theta_{-a})\}$.

To show that these results have content, we give two specific examples. For the first, let $X_1 = A - B$, with $A \sim N(1.5, 1)$, $B \sim \mathrm{Exp}(1)$ and A, B independent, so that $\psi(\theta) = 1.5\theta + 0.5\theta^2 - \log(1+\theta)$ for $\theta \in (-1, \infty)$. Let $a = 2.5$. Then $\theta_a = \sqrt{2}$ and $\gamma_a \approx 1.2956$; $\theta_{-a} = (-5 + \sqrt{13})/2 \approx -0.6972$ and $\gamma_{-a} \approx 1.3511$; but $-\theta_a < -1$ is outside the domain of ψ so we get infinite variance, as in Theorem 1(i). If we let $a = 1.5$, then $\theta_a = (\sqrt{5} - 1)/2 \approx 0.6180$ and $\gamma_a \approx 0.2902$; $\theta_{-a} = (-2 + \sqrt{2}) \approx -0.5858$ and $\gamma_{-a} \approx 0.7044$. Since $\theta_a + \theta_{-a} > 0$, the variance increases with n, as in Theorem 1(ii). This case is illustrated in Figure 1.

Our next example illustrates case (iii) in Theorem 1. Let $X_1 \sim N(1,1)$. For $a > 1$, we have $\theta_a = a - 1$ and $\gamma_a = (a-1)^2/2$; $\theta_{-a} = -(a+1)$ and $\gamma_{-a} = (a+1)^2/2$. The conditions $\gamma_a < \gamma_{-a}$ and $\theta_a + \theta_{-a} < 0$ always hold, and $-\psi(\theta_a) - \psi(\theta_{-a}) - a(\theta_a + \theta_{-a}) = -a^2 + 2a + 1$. For values of a in the ranges $(1, 2], (2, 1 + 2/\sqrt{3}], (1 + 2/\sqrt{3}, 1 + \sqrt{2}]$ and $(1 + \sqrt{2}, +\infty)$, the rate in (iii) falls in the ranges $2\gamma_a$, $[\gamma_a, 2\gamma_a)$, $[0, \gamma_a)$ and $(-\infty, 0)$, respectively.

REMARKS. (a) Examination of the proof of Proposition 1 [see especially (15)] reveals that P_{θ_a} is the *only* exponential change of measure for which the resulting likelihood ratio is consistent with the large deviations behavior of $P(|S_n| \geq an)$, even in the weaker sense (6). Notice also that for $\theta > \theta_a$,

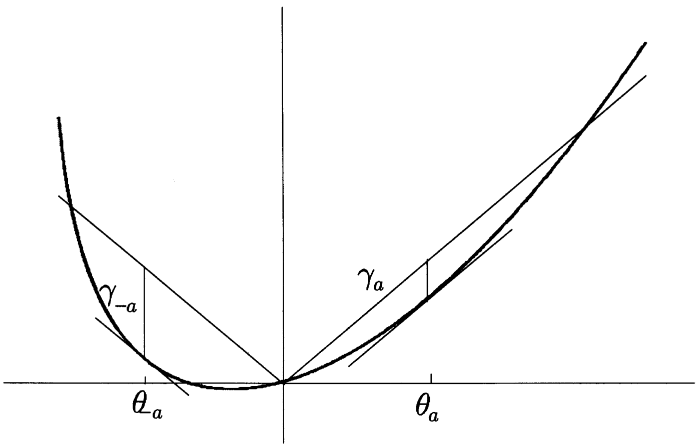

FIG. 1. *Graph of* ψ *for the first example. The tangent to* ψ *has slope* a *and* $-a$ *at points* θ_a *and* θ_{-a}, *respectively. The vertical distance from* $\psi(\theta_{\pm a})$ *to the line through the origin with slope* $\pm a$ *is* $\gamma_{\pm a}$.

$P_\theta(|S_n| \geq an) \to 1$ as suggested in the discussion following (7). For the second example after Theorem 1, it can be shown that the exponential change of measure that maximizes the exponential rate of decrease of the variance twists by $\min(1, a-1) = \min(1, \theta_a)$ for all $a > 1$, so the large deviations change of measure is the best exponential change of measure only for $1 < a \leq 2$. Twisting by $\theta = \min(1, a-1)$ maximizes the minimum of the rates of decrease of $E_\theta[L_n^2;\ S_n \geq an]$ and $E_\theta[L_n^2;\ S_n \leq -an]$. The resulting estimator fails to be asymptotically efficient for $a > 2$.

(b) The variance of the crude estimator $\mathbf{1}_{\{|S_n| \geq an\}}$ generated without importance sampling decreases exponentially at rate γ_a. In particular, then, the large deviations estimator is dramatically outperformed by the crude one when Theorem 1(i)–(ii) apply. This may also be the case in (iii), as illustrated by the second example above.

(c) We could alternatively have defined $\hat{\beta}$ as a mixture of $\hat{\beta}_+$ and $\hat{\beta}_-$ and still obtained asymptotic efficiency. In practice, one would allocate a fraction of runs to evaluation of $\hat{\beta}_+$ and the remainder to evaluation of $\hat{\beta}_-$. It is well known (and easily verified) that allocating runs in proportion to standard deviations minimizes variance. The optimal fraction allocated to $\hat{\beta}_-$ is therefore

$$\frac{\exp(-\gamma_{-a}n + o(n))}{\exp(-\gamma_{-a}n + o(n)) + \exp(-\gamma_a n + o(n))}.$$

The resulting estimator thus remains asymptotically efficient by allocating an asymptotically negligible fraction of runs to the secondary event $\{S_n \leq -an\}$, whereas the usual large deviations estimator produces potentially infinite variance in ignoring this event altogether.

(d) The example treated in Theorem 1 satisfies the necessary condition for asymptotic efficiency in [28], but not the sufficient condition given there. Violation of this sufficient condition might be taken to suggest poor performance, but Sadowsky and Bucklew [28] do not discuss what happens when their sufficient condition fails to hold. Interestingly, in the second example after Theorem 1, for values of a between 1 and 2, the sufficient condition is violated but $\hat{\alpha}$ is asymptotically efficient. For $a > 2$, the same example violates the necessary and sufficient condition of Proposition 2 of [12] for Gaussian probabilities. Sadowsky [27] gives a very general formulation of this necessary and sufficient condition that applies to arbitrary moments of an importance sampling estimator, not just the second moment.

(e) Fresnedo [15] and Sadowsky and Bucklew [28] also propose mixtures of importance sampling measures. Nakayama [23] notes the importance of secondary paths to failure in simulating highly reliable systems modeled as Markov chains, but contrasts this with the large deviations setting.

(f) Some might argue that the setting treated in Theorem 1 is rigged because $P(|S_n| \geq an)$ is *obviously* the sum of two more primitive probabilities. We counter that virtually all interesting rare events arising in applications are unions of meaningful events describing alternative ways in which the rare event can occur. One of these may dominate the probability of the rare event though the others carry sufficient weight to affect the variance. The setting of Theorem 1 merely makes transparent a possibility present in more general cases.

4. Level crossing probabilities. Just as the example of the previous section builds on classical results for sum tails, the examples of this section build on classical results for level crossing probabilities. We begin with some background, continuing to use notation from the previous section.

Suppose that $\psi'(0) \equiv E[X_1] < 0$ so that $\{S_n,\ n \geq 0\}$ is a negative-drift random walk. Suppose that in the interior of the domain of ψ there is strictly positive γ (necessarily unique, by the convexity of ψ) at which $\psi(\gamma) = 0$. Then $\psi'(\gamma)$ exists and is necessarily positive because $\psi(0) = 0$; $\psi'(0) < 0$ and ψ is convex. For $x > 0$ define

$$T_1(x) = \inf\{n \geq 0 \colon S_n > x\}.$$

It is well known that, if the distribution of X_1 is nonlattice, then

$$P(T_1(x) < \infty) \sim Ce^{-\gamma x} \quad \text{as } x \to \infty, \tag{11}$$

for some constant $C \in (0,1)$. [In the lattice case, (11) holds for x increasing through multiples of the span of X_1.] Importance sampling for the estimation of this probability starts from the representation

$$P(T_1(x) < \infty) = E_\gamma\left[\exp\left(-\gamma S_{T_1(r)}\right); T_1(x) < \infty\right] = E_\gamma\left[\exp\left(-\gamma S_{T_1(x)}\right)\right],$$

where E_γ is expectation with respect to the γ-twisted measure, and $P_\gamma(T_1(x) < \infty) = 1$ because $E_\gamma[X_1] = \psi'(\gamma) > 0$. The estimator $\exp(-\gamma S_{T_1(x)})$ generated under P_γ satisfies

$$\frac{\sqrt{E_\gamma\left[\exp\left(-2\gamma S_{T_1(x)}\right)\right]}}{P(T_1(x) < \infty)} \leq \frac{\sqrt{\exp(-2\gamma x)}}{P(T_1(x) < \infty)} = \frac{\exp(-\gamma x)}{P(T_1(x) < \infty)} \to \frac{1}{C}$$

and thus has bounded relative error. The examples in this section build on this background but in a two-dimensional setting.

Let $\{(X_n^1, X_n^2),\ n \geq 1\}$ be an i.i.d. sequence of two-dimensional vectors with $EX_1^i < 0$, $i = 1,2$. Set $S_0^i = 0$, $S_n^i = \sum_{j=1}^n X_j^i$ for $n \geq 0$, and $T_i = T_i(x) = \inf\{n \geq 0 \colon S_n^i > x\}$, $i = 1,2$. For simplicity, we suppose X_1^1 and X_1^2 have nonlattice distributions. Then

$$P(T_i < \infty) \sim C_i \exp(-\gamma_i x), \tag{12}$$

if $\gamma_i > 0$ solves $\psi_{X^i}(\gamma_i) = 0$ in the interior of the domain of ψ_{X^i}. Our objective is to estimate $\alpha(x) \triangleq P(T < \infty)$ with $T = \min\{T_1, T_2\}$.

We set

$$\psi(\theta_1, \theta_2) = \log E\left[\exp\left(\theta_1 X_1^1 + \theta_2 X_1^2\right)\right] \tag{13}$$

and work with bivariate twisted distributions of the form

$$\begin{aligned} &P_{\theta_1,\theta_2}\left(X^1 \leq x_1, X^2 \leq x_2\right) \\ &\qquad = E\left[\exp\left\{\theta_1 X^1 + \theta_2 X^2 - \psi(\theta_1, \theta_2)\right\}; X^1 \leq x_1, X^2 \leq x_2\right], \end{aligned}$$

repeatedly using the standard property $E_{\theta_1,\theta_2}[X^i] = \partial_{\theta_i}\psi(\theta_1, \theta_2)$, $i = 1,2$.

Since $P(T < \infty) = E_{\gamma_1, 0}[\exp(-\gamma_1 S_T^1)]$, the estimator

$$\hat{\alpha}(x) = \exp\left(-\gamma_1 S_T^1\right) \tag{14}$$

generated under $P_{\gamma_1, 0}$ is unbiased and even consistent with the large deviations behavior of $\{T(x) < \infty\}$ if $\gamma_1 < \gamma_2$. An alternative is $\hat{\beta}(x) = \hat{\beta}_1(x) + \hat{\beta}_2(x)$ with $\hat{\beta}_1(x) = \exp(-\gamma_1 S_T^1)\mathbf{1}_{\{T_1 \le T_2\}}$ generated under $P_{\gamma_1, 0}$ and $\hat{\beta}_2(x) = \exp(-\gamma_2 S_T^2)\mathbf{1}_{\{T_2 < T_1\}}$ generated under P_{0, γ_2}. We analyze these estimators in the following two results. Their proofs are deferred to Section 5.

PROPOSITION 2. *Suppose $\gamma_1 < \gamma_2$. Then we have the following*:

(i) $P(T(x) < \infty) \sim C_1 \exp(-\gamma_1 x)$.
(ii) *The likelihood ratio in* (14) *satisfies* (7).
(iii) *The estimator $\hat{\beta}(x)$ has bounded relative error.*

It remains to analyze the performance of the estimator $\hat{\alpha}(x)$, the one suggested by the leading-term asymptotics of $P(T(x) < \infty)$. For this analysis, we impose the mild regularity condition that for some $\varepsilon > 0$, the set $\{(\theta_1, \theta_2)\colon \psi(\theta_1, \theta_2) \le \varepsilon\}$ is compact (it is automatically convex). This then implies that the level curve $\{(\theta_1, \theta_2)\colon \psi(\theta_1, \theta_2) = 0\}$ is the boundary of the compact, convex set $\{(\theta_1, \theta_2)\colon \psi(\theta_1, \theta_2) \le 0\}$. The points $(0, 0)$, $(\gamma_1, 0)$ and $(0, \gamma_2)$ all lie on this level curve. Convexity of ψ implies that at any point θ on this level curve, the gradient $(\partial_{\theta_1}\psi(\theta), \partial_{\theta_2}\psi(\theta))$ is normal to the level curve. Define

$$\theta_1^- = \min\{\theta_1\colon \exists\, \theta_2,\ \psi(\theta_1, \theta_2) = 0\}$$

and θ_2^- as the solution to $\psi(\theta_1^-, \theta_2^-) = 0$. Then $\theta^- = (\theta_1^-, \theta_2^-)$ is the leftmost point on the level curve and thus $\partial_{\theta_1}\psi(\theta^-) < 0$ and $\partial_{\theta_2}\psi(\theta^-) = 0$. Define $\theta^+ = (\theta_1^+, \theta_2^+)$ by the requirement that $\psi(\theta^+) = 0$ and $\partial_{\theta_1}\psi(\theta^+) = \partial_{\theta_2}\psi(\theta^+) > 0$. This is the unique point at which the normal to the level curve points at a 45° angle. We always have $\theta_1^- < \theta_1^+$. Finally, set

$$\theta_2' = \sup\{\theta_2\colon \psi(-\gamma_1, \theta_2) \le 0\},$$

the highest point at which the line $\theta_1 = -\gamma_1$ crosses the level curve, if it crosses it at all. These points are illustrated in Figure 2.

THEOREM 2. *Suppose $\gamma_1 < \gamma_2$.*

(i) *If $-\gamma_1 < \theta_1^-$, then $E_{\gamma_1, 0}[\hat{\alpha}^2(x)] \to \infty$.*
(ii) *If $\theta_1^- \le -\gamma_1 < \theta_1^+$, then $E_{\gamma_1, 0}[\hat{\alpha}^2(x)] \sim C \exp(-\min\{2\gamma_1, \theta_2'\}x)$ for some finite, positive constant C.*
(iii) *If $\theta_1^+ \le -\gamma_1$, then $E_{\gamma_1, 0}[\hat{\alpha}^2(x)] \sim C \exp(-2\gamma_1 x)$ for some finite, positive constant C.*

Because we know from Proposition 2(i) that $\alpha(x)$ decreases exponentially at rate γ_1, the estimator $\hat{\alpha}(x)$ has bounded relative error only in case (iii) of

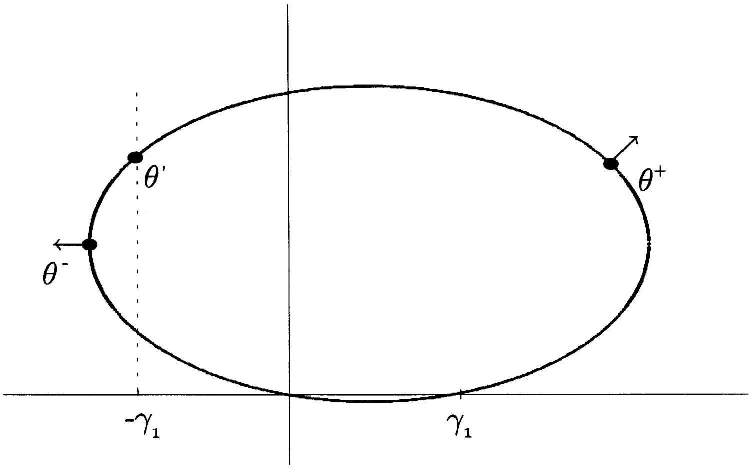

FIG. 2. *Level curve of ψ for a bivariate normal distribution with independent marginals. The point θ' has coordinates $(-\gamma_1, \theta_2')$. In this example, $-\theta_1^- < -\gamma_1 < \theta_1^+$, so case* (ii) *of Theorem* 2 *applies. Case* (i) *would apply if the vertical dashed line at $-\gamma_1$ were to the left of θ^-. Case* (iii) *would apply if the line were to the right of θ^+, which would require nonzero correlation.*

Theorem 2. In case (ii), the exponential rate of decrease of the variance of $\hat{\alpha}(x)$ may be less than $2\gamma_1$ [in which case $\hat{\alpha}(x)$ fails to be asymptotically efficient], and it may even be less than γ_1, the rate of decrease of the variance of $\mathbf{1}_{\{T(x)<\infty\}}$ under the original measure. The proof of Theorem 2 shows that in case (i), the rate of *increase* of the variance is faster than any exponential.

Each of these cases may prevail, even for well-behaved distributions, as the following examples show. Suppose X^1, X^2 are independent and $X^i \sim N(-\mu_i, \sigma_i^2)$, $i = 1, 2$. The requirement $\gamma_1 < \gamma_2$ entails $\mu_2/\mu_1 > \sigma_2^2/\sigma_1^2$. When $\Delta \equiv \mu_2^2 - 8\mu_1^2\sigma_2^2/\sigma_1^2 < 0$, $E_{\gamma_1, 0}[\hat{\alpha}^2(x)] \to \infty$; when $\Delta \geq 0$, $\theta_2' = (\mu_2 + \sqrt{\Delta})/\sigma_2^2$, and θ_2' may take values in $(0, \gamma_1)$ and $(\gamma_1, 2\gamma_1)$. See Figure 2.

For case (iii), we give a specific numerical example. Suppose X_1, X_2 are jointly normal with means -1 and -4.5, variances 4 and 2.25 and correlation 0.6. Then we have $\gamma_1 = 0.5 < \gamma_2 = 4$; $\theta_1^+ \approx -0.62265 < -\gamma_1$.

The analysis above applies to a problem in queueing by exploiting a standard connection between waiting time distributions and level crossing probabilities (see, e.g., Chapter VIII of [4]). The probability that at least one coordinate of a two-dimensional random walk crosses a threshold gives, with minor modification, the tail distribution of the response time in a fork-join queue—a system in which each job splits into two subjobs which join parallel queues, receive independent service, and are ultimately rejoined. (The details of this correspondence are in an earlier version of this paper, available from the authors.)

As an example, suppose arrivals are Poisson with rate 1 and service times at the two queues are exponentially distributed with rates 2 and 2.6 at queues 1 and 2, respectively. Then, in the notation of Theorem 2, $\gamma_1 = 1 <$

$\gamma_2 = 1.6$. Numerically we find that at $\theta_1 \approx -0.94026$ and $\theta_2 \approx 1.27013$ we are in the setting of Theorem 2(i). If the second service rate is changed to 2.9, then $\theta_1^- \approx -1.21278$, $\theta_1^+ \approx 0.65639$ and $\theta_2' \approx 1.86130$. Here, case (ii) of Theorem 2 applies, so $E_{\gamma_1, 0}[\hat{\alpha}^2]$ goes to zero exponentially at rate $\theta_2' \approx 1.8613$, which is slower than $2\gamma_1 = 2$. In particular, the estimator fails to be asymptotically efficient.

This example could hardly be considered pathological as it involves standard distributions and reasonable parameters. Based on observations in [1], [24] and [16], one might expect the effectiveness of importance sampling to deteriorate when the service rates at the two queues are close. It is perhaps surprising just how poorly the standard approach to importance sampling performs in this case, in spite of a substantial difference in the rates.

5. Proofs. We now prove the results stated in Section 3 and Section 4.

PROOF OF PROPOSITION 1. Part (i) follows from (8), (9) and the general property $\log(x+y) = \log(x) + O(y/x)$ as $(y/x) \to 0$. For (ii),

$$-\frac{1}{n}\log L_n = \theta_a \frac{S_n}{n} - \psi(\theta_a) \to \theta_a a - \psi(\theta_a) \equiv \gamma_a, \quad P_{\theta_a}\text{-a.s.,} \tag{15}$$

by the strong law of large numbers and the c.g.f. property $E_\theta[X_1] = \psi'(\theta)$.

For (iii) we know from Theorem 2 of [9] that $-n^{-1}\log E_{\theta_{\pm a}}[\hat{\beta}_\pm^2] \to 2\gamma_{\pm a}$. Because $\hat{\beta}(n)$ is the sum of $\hat{\beta}_+(n)$ and $\hat{\beta}_-(n)$, the logarithm of its second moment decreases asymptotically at rate no less than the smaller of the rates for the two estimators separately, that is, no less than $\min\{2\gamma_a, 2\gamma_{-a}\} = 2\gamma_a$. Nonnegativity of the variance of $\hat{\beta}(n)$ implies that the rate can be no greater than this, so the rate must get exactly $2\gamma_a$, as required for asymptotic efficiency. □

PROOF OF THEOREM 1. Write

$$\begin{aligned} E_{\theta_a}\big[\hat{\alpha}^2(n)\big] &= E_{\theta_a}\big[\exp(-2(\theta_a S_n - n\psi(\theta_a))); |S_n| \ge an\big] \\ &= E\big[\exp(-(\theta_a S_n - n\psi(\theta_a))); |S_n| \ge an\big] \\ &= E\big[\exp(-(\theta_a S_n - n\psi(\theta_a))); S_n \ge an\big] \\ &\quad + E\big[\exp(-(\theta_a S_n - n\psi(\theta_a))); S_n \le -an\big]. \end{aligned} \tag{16}$$

For (i), we have

$$\begin{aligned} E_{\theta_a}\big[\hat{\alpha}^2(n)\big] &\ge E\big[\exp(-(\theta_a S_n - n\psi(\theta_a))); S_n \le -an\big] \\ &= E\big[\exp(-(\theta_a S_n - n\psi(\theta_a)))\big] \\ &\quad - E\big[\exp(-(\theta_a S_n - n\psi(\theta_a))); S_n > -an\big] \\ &\ge \exp(n\psi(\theta_a))\big(E\big[\exp(-\theta_a S_n)\big] - \exp(n\theta_a a)\big) \\ &= \exp(n\psi(\theta_a))\big(\exp(n\psi(-\theta_a)) - \exp(n\theta_a a)\big) = \infty \end{aligned}$$

because $\psi(-\theta_a) = \infty$.

For (ii) and (iii), write the second term of (16) as

$$\begin{aligned}
&E\big[\exp(-(\theta_a S_n - n\psi(\theta_a))); S_n \le -an\big]\\
&\quad = E_{-\theta_a}\big[\exp(-\theta_a S_n + n\psi(\theta_a) - (-\theta_a)S_n + n\psi(-\theta_a)); \\
&\qquad\qquad S_n \le -an\big]\\
&\quad = \exp(n\psi(\theta_a) + n\psi(-\theta_a))P_{-\theta_a}(S_n \le -an).
\end{aligned} \tag{17}$$

If $\theta_a + \theta_{-a} \ge 0$ then $E_{-\theta_a}[X_1] = \psi'(-\theta_a) \le \psi'(\theta_{-a}) = -a$ and $\liminf_{n\to\infty} P_{-\theta_a}(S_n \le -an) > 0$. Convexity of ψ implies that $\psi(\theta_a) + \psi(-\theta_a) > 2\psi(0) = 0$ so the result in (ii) follows from (16) and (17).

If $\theta_a + \theta_{-a} < 0$ then $E_{-\theta_a}[X_1] = \psi'(-\theta_a) > -a$. Define

$$\begin{aligned}
\psi_{-\theta_a}(\theta) &\triangleq \log E_{-\theta_a}\big[\exp(\theta X_1)\big] = \log E\big[\exp(-\theta_a X_1 - \psi(-\theta_a) + \theta X_1)\big]\\
&= \psi(\theta - \theta_a) - \psi(-\theta_a).
\end{aligned}$$

We have $\psi'_{-\theta_a}(\theta) = \psi'(\theta - \theta_a)$, so $\bar\theta_{-a} \equiv \theta_a + \theta_{-a}$ solves

$$\psi'_{-\theta_a}(\bar\theta_{-a}) = -a.$$

Cramér's theorem gives

$$\begin{aligned}
-\frac{1}{n}\log P_{-\theta_a}(S_n \le an) &\to -\bar\theta_{-a}a - \psi_{-\theta_a}(\bar\theta_{-a})\\
&= -a(\theta_a + \theta_{-a}) - \psi(\theta_{-a}) + \psi(-\theta_a).
\end{aligned}$$

Recalling (17), we get

$$\begin{aligned}
&-\frac{1}{n}\log E\big[\exp(-(\theta_a S_n - n\psi(\theta_a))); S_n \le -an\big]\\
&\quad \to -\psi(\theta_a) - \psi(-\theta_a) - a(\theta_a + \theta_{-a}) - \psi(\theta_{-a}) + \psi(-\theta_a)\\
&\quad = -\psi(\theta_a) - \psi(\theta_{-a}) - a(\theta_a + \theta_{-a}).
\end{aligned}$$

Also, we know from Theorem 2 of [9] that

$$-\frac{1}{n}\log E\big[\exp(-(\theta_a S_n - n\psi(\theta_a))); S_n \ge an\big] \to 2\gamma_a.$$

The result in (iii) now follows from (16). □

Proof of Proposition 2. Case (i) follows from writing

$$P(T_1(x) < \infty) \le P(T(x) < \infty) \le P(T_1(x) < \infty) + P(T_2(x) < \infty)$$

and invoking (12). For (ii), we write

$$\exp\big(-\gamma_1 S^1_{T(x)}\big) = \exp\big(-\gamma_1 S^1_{T_1(x)}\big)\mathbf{1}_{\{T_1(x) \le T_2(x)\}} + \exp\big(-\gamma_1 S^1_{T_2(x)}\big)\mathbf{1}_{\{T_2(x) < T_1(x)\}}$$

and argue that the second term is zero for large x. If $\partial_{\theta_2}\psi(\gamma_1, 0) < 0$, then $T_2(x) = \infty$ for all sufficiently large x, $P_{\gamma_1,0}$-a.s., whereas $T_1(x) < \infty$ for all x. If $\partial_{\theta_2}\psi(\gamma_1, 0) = 0$ then

$$\frac{T_2(x)}{T_1(x)} \to \infty, \quad P_{\gamma_1,0}\text{-a.s.} \tag{18}$$

If $\partial_{\theta_2}\psi(\gamma_1, 0) > 0$ then

$$\frac{T_2(x)}{T_1(x)} \to \frac{E_{\gamma_1,0}[X^1]}{E_{\gamma_1,0}[X^2]} = \frac{\partial_{\theta_1}\psi(\gamma_1,0)}{\partial_{\theta_2}\psi(\gamma_1,0)}, \quad P_{\gamma_1,0}\text{-a.s.} \tag{19}$$

Convexity of ψ implies convexity of the set $\{(u,v): \psi(u,v) \le 0\}$ which further implies (essentially by the supporting hyperplane theorem; see, e.g., the argument used for Theorem 2.3.7 of [7])

$$(u - \gamma_1)\partial_{\theta_1}\psi(\gamma_1,0) + (v - 0)\partial_{\theta_2}\psi(\gamma_1,0) \le 0. \tag{20}$$

Applying this inequality at $(u,v) = (0, \gamma_2)$ and rearranging terms we get

$$\frac{\partial_{\theta_1}\psi(\gamma_1,0)}{\partial_{\theta_2}\psi(\gamma_1,0)} \ge \frac{\gamma_2}{\gamma_1} > 1.$$

Thus, (18) and (19) imply that for all sufficiently large x, $T(x) = T_1(x)$ and $S^1_{T(x)} - x$ has the same limit in distribution as $S^1_{T_1(x)} - x$. That the overshoot $S^1_{T_1(x)} - x$ converges in distribution follows from the renewal theorem through the analysis of ladder heights; for details see [4], Theorem XII.5.3 or [31], Corollary 8.33.

For (iii), observe that the variance of $\hat{\beta}(x)$ is bounded above by $\exp(-2\gamma_1 x) + \exp(-2\gamma_2 x)$, recall that $\gamma_1 < \gamma_2$ and invoke (i) to conclude that $\hat{\beta}(x)$ has bounded relative error. $\square$

Proof of Theorem 2. For (i), choose $m > 0$ arbitrarily large. Abbreviating $E_{\theta_1^-,\theta_2^-}$ to E_{θ^-}, we have

$$\begin{aligned}
&E_{\gamma_1,0}\Big[\exp\big(-2\gamma_1 S^1_{T(x)}\big)\Big]\\
&\quad= E\Big[\exp\big(-\gamma_1 S^1_{T(x)}\big); T(x) < \infty\Big]\\
&\quad\ge E\Big[\exp\big(-\gamma_1 S^1_{T(x)}\big); T_2(x) < \infty\Big]\\
&\quad= E_{\theta^-}\Big[\exp\big(-\gamma_1 S^1_{T(x)} - \theta_1^- S^1_{T_2(x)} - \theta_2^- S^2_{T_2(x)}\big)\Big]\\
&\quad\ge E_{\theta^-}\Big[\exp\big(-(\gamma_1 + \theta_1^-)S^1_{T_2(x)} - \theta_2^- S^2_{T_2(x)}\big);\\
&\qquad\qquad T_2(x) \le T_1(x),\, S^1_{T_2(x)} \le -mx\Big]\\
&\quad\ge \big(\exp[(\gamma_1 + \theta_1^-)m - \theta_2^-]x\big)E_{\theta^-}\Big[\exp\big(-\theta_2^-\big(S^2_{T_2(x)} - x\big)\big);\\
&\qquad\qquad\qquad\qquad T_2(x) \le T_1(x),\, S^1_{T_2(x)} \le -mx\Big],
\end{aligned}$$

where the last inequality uses $\gamma_1 + \theta_1^- > 0$. Since $E_{\theta^-}[X_1^1] < 0$ and $E_{\theta^-}[X_1^2] = 0$, we have $T_2(x) < T_1(x) = \infty$ for all sufficiently large x, P_{θ^-}-a.s. Also, $x^{-1}S^1_{T_2(x)} \to -\infty$, so $P_{\theta^-}(T_2(x) \le T_1(x),\ S^1_{T_2(x)} \le -mx) \to 1$. Because the overshoot $S^2_{T_2(x)} - x$ has a limiting distribution under P_{θ^-},

$$E_{\theta^-}\Big[\exp\big(-\theta_2^-\big(S^2_{T_2(x)} - x\big)\big); T_2(x) \le T_1(x),\, S^1_{T_2(x)} \le -mx\Big] \tag{21}$$

converges to a positive constant C' as $x \to \infty$. We have therefore shown that $E_{\gamma_1,0}[\hat{\alpha}^2(x)]$ is bounded below by a quantity asymptotic to $C' \exp([(\gamma_1 + \theta_1^-) m - \theta_2]x)$. Since m may be chosen arbitrarily large, we conclude that $E_{\gamma_1,0}[\hat{\alpha}^2(x)]$ increases to infinity faster than any exponential.

For cases (ii) and (iii), recall that $P_{\gamma_1,0}(T_1 < \infty) = 1$, write

$$E_{\gamma_1,0}\big[\hat{\alpha}^2(x)\big] = E_{\gamma_1,0}\big[\exp\big(-2\gamma_1 S^1_{T_1}\big); T_1 \le T_2\big] + E_{\gamma_1,0}\big[\exp\big(-2\gamma_1 S^1_{T_2}\big); T_2 < T_1\big] \tag{22}$$

and consider the two terms on the right separately. The argument used in Proposition 2(i) shows that

$$E_{\gamma_1,0}\big[\exp\big(-2\gamma_1 S^1_{T_1}\big); T_1 \le T_2\big] \sim C_1' \exp(-2\gamma_1 x) \tag{23}$$

for some constant C_1'. For the second term, we have

$$\begin{aligned} &E_{\gamma_1,0}\big[\exp\big(-2\gamma_1 S^1_{T_2}\big); T_2 < T_1\big] \\ &\quad = E\big[\exp\big(-\gamma_1 S^1_{T_2}\big); T_2 < T_1, T_2 < \infty\big] \\ &\quad = E_{-\gamma_1,\theta_2'}\big[\exp\big(-\gamma_1 S^1_{T_2}\big)\exp\big(\gamma_1 S^1_{T_2} - \theta_2' S^2_{T_2}\big); T_2 < T_1, T_2 < \infty\big] \\ &\quad = E_{-\gamma_1,\theta_2'}\big[\exp\big(-\theta_2' S^2_{T_2}\big); T_2 < T_1, T_2 < \infty\big] \\ &\quad = \exp(-\theta_2' x) E_{-\gamma_1,\theta_2'}\big[\exp\big(-\theta_2'\big(S^2_{T_2} - x\big)\big); T_2 < T_1, T_2 < \infty\big]. \end{aligned} \tag{24}$$

In both cases (ii) and (iii), we have $\partial_{\theta_2}\psi(-\gamma_1, \theta_2') \ge 0$; that is, $E_{-\gamma_1,\theta_2'}[X^2] \ge 0$. In case (ii), we have $\partial_{\theta_2}\psi(-\gamma_1, \theta_2') > \partial_{\theta_1}\psi(-\gamma_1, \theta_2')$ so $P_{-\gamma_1,\theta_2'}(T_2 < T_1, T_2 < \infty) \to 1$ and

$$E_{-\gamma_1,\theta_2'}\big[\exp\big(-\theta_2'\big(S^2_{T_2} - x\big)\big); T_2 < T_1, T_2 < \infty\big]$$

has a limit in $(0, 1)$ as $x \to \infty$. [The details are the same as those leading to (21).] The result in (ii) thus follows from (23) and (24).

In case (iii), $\partial_{\theta_1}\psi(-\gamma_1, \theta_2') \ge \partial_{\theta_2}\psi(-\gamma_1, \theta_2') > 0$, so $E_{-\gamma_1,\theta_2'}[X^2] \ge 0$, and

$$E_{-\gamma_1,\theta_2'}\big[\exp\big(-\theta_2'\big(S^2_{T_2} - x\big)\big); T_2 < T_1, T_2 < \infty\big] \le E_{-\gamma_1,\theta_2'}\big[\exp\big(-\theta_2'\big(S^2_{T_2} - x\big)\big)\big] \to C_2'$$

for some C_2' in $(0, 1)$. Thus, the second term in (22) is bounded above by a quantity asymptotic to $C_2' \exp(-\theta_2' x)$. To compare this with (23), first note that [see (20)]

$$(u + \gamma_1)\partial_{\theta_1}\psi(-\gamma_1, \theta_2') + (v - \theta_2')\partial_{\theta_2}\psi(-\gamma_1, \theta_2') \le 0$$

if $\psi(u, v) \le 0$. Now set $(u, v) = (\gamma_1, 0)$ and rearrange terms to get

$$\theta_2 \ge \frac{\partial_{\theta_1}\psi(-\gamma_1, \theta_2')}{\partial_{\theta_2}\psi(-\gamma_1, \theta_2)} 2\gamma_1 \ge 2\gamma_1.$$

The first term in (22) therefore dominates and we get the conclusion in (iii). □

Acknowledgments. Both authors thank the referees for detailed comments and suggestions and for pointing out some relevant references. The first author thanks Peter Glynn for helpful earlier discussions on twisting and level curves.

REFERENCES

[1] ANANTHARAM, V., HEIDELBERGER, P. and TSOUCAS, P. (1990). Analysis of rare events in continuous time Markov chains via time reversal and fluid approximations. Research Report RC-16280, IBM, Yorktown Heights, NY.

[2] ASMUSSEN, S. (1982). Conditioned limit theorems relating a random walk to its associate. *Adv. in Appl. Probab.* **14** 143–170.

[3] ASMUSSEN, S. (1985). Conjugate processes and the simulation of ruin probabilities. *Stochastic Process. Appl.* **20** 213–229.

[4] ASMUSSEN, S. (1987). *Applied Probability and Queues*. Wiley, Chichester.

[5] ASMUSSEN, S. and BINSWANGER, K. (1995). Simulation of ruin probabilities for subexponential claims. Working paper, Dept. Mathematical Statistics, Univ. Lund, Sweden.

[6] ASMUSSEN, S., RUBINSTEIN, R. and WANG, C. L. (1994). Regenerative rare event simulation via likelihood ratios. *J. Appl. Probab.* **31** 797–815.

[7] BAZARAA, M. S. and SHETTY, C. M. (1979). *Nonlinear Programming: Theory and Algorithms*. Wiley, New York.

[8] BUCKLEW, J. A. (1990). *Large Deviation Techniques in Decision, Simulation, and Estimation*. Wiley, New York.

[9] BUCKLEW, J. A., NEY, P. and SADOWSKY, J. S. (1990). Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *J. Appl. Probab.* **27** 44–59.

[10] CHANG, C. S., HEIDELBERGER, P., JUNEJA, P. and SHAHABUDDIN, P. (1994). Effective bandwidth and fast simulation of ATM intree networks. *Performance Evaluation* **20** 45–65.

[11] CHANG, C. S., HEIDELBERGER, P. and SHAHABUDDIN, P. (1995). Fast simulation of packet loss rates in a shared buffer communications switch. *ACM Trans. Modeling Comp. Simulation* **5** 306–325.

[12] CHEN, J.-C., LU, D., SADOWSKY, J. S. and YAO, K. (1993). On importance sampling in digital communications. I: fundamentals. *IEEE J. Selected Areas in Comm.* **11** 289–307.

[13] DEMBO, A. and ZEITOUNI, O. (1993). *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston.

[14] FRATER, M., LENNON, T. M. and ANDERSON, B. D. O. (1991). Optimally efficient estimation of the statistics of rare events in queuing networks. *IEEE Trans. Automat. Control* **36** 1395–1405.

[15] FRESNEDO, R. D. (1989). Quick simulation of rare events in networks. In *Proceedings of the Winter Simulation Conference* 514–523. Society for Computer Simulation, San Diego.

[16] GLASSERMAN, P. and KOU, S. (1995). Analysis of an importance sampling estimator for tandem queues. *ACM Trans. Modeling Comp. Simulation* **4** 22–42.

[17] GLYNN, P. W. (1995). Large deviations for the infinite server queue in heavy traffic. *IMA Vol. Math. Appl.* **71** 387–394.

[18] HEIDELBERGER, P. (1995). Fast simulation of rare events in queueing and reliability models. *ACM Trans. Modeling Comp. Simulation* **4** 43–85.

[19] HEIDELBERGER, P., SHAHABUDDIN, P. and NICOLA, V. (1994). Bounded relative error in estimating transient measures of highly dependable non-Markovian systems. *ACM Trans. Modeling Comp. Simulation* **4** 137–164.

[20] KESIDIS, G. and WALRAND, J. (1993). Quick simulation of ATM buffers with on-off multiclass Markov fluid sources. *ACM Trans. Modeling Comp. Simulation* **3** 269–276.

[21] LEHTONEN, T. and NYRHINEN, H. (1992). Simulating level-crossing probabilities by importance sampling. *Adv. in Appl. Prob.* **24** 858–874.

[22] MCDONALD, D. R. (1995). Asymptotics of first passage times for random walk in a quadrant. Working paper, Dept. Mathematics and Statistics, Univ. Ottawa.

[23] NAKAYAMA, M. K. (1994). A characterization of the simple failure biasing method for simulations of highly reliable Markovian systems. *ACM Trans. Modeling Comp. Simulation* **4** 52–88.

[24] PAREKH, S. and WALRAND, J. (1989). A quick simulation method for excessive backlogs in networks of queues. *IEEE Trans. Automat. Control* **34** 54–66.

[25] SADOWSKY, J. S. (1991). Large deviations and efficient simulation of excessive backlogs in a $GI/G/m$ queue. *IEEE Trans. Automat. Control* **36** 1383–1394.

[26] SADOWSKY, J. S. (1993). On the optimality and stability of exponential twisting in Monte Carlo estimation. *IEEE Trans. Inform. Theory* **39** 119–128.

[27] SADOWSKY, J. S. (1997). On Monte Carlo estimation of large deviations probabilities. Unpublished manuscript.

[28] SADOWSKY, J. S. and BUCKLEW, J. A. (1990). On large deviations theory and asymptotically efficient Monte Carlo estimation. *IEEE Trans. Inform. Theory* **36** 579–588.

[29] SHAHABUDDIN, P. (1994). Importance sampling for the simulation of highly reliable Markovian systems. *Management Science* **40** 333–352.

[30] SIEGMUND, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* **4** 673–684.

[31] SIEGMUND, D. (1985). *Sequential Analysis*. Springer, New York.

GRADUATE SCHOOL OF BUSINESS
URIS HALL
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027
E-MAIL: pglasser@research.gsb.columbia.edu
ywang@groucho.gsb.columbia.edu