The Annals of Applied Probability
1996, Vol. 6, No. 2, 399–422

ON MONTE CARLO ESTIMATION OF LARGE DEVIATIONS PROBABILITIES[1]

BY JOHN S. SADOWSKY

Arizona State University

Importance sampling is a Monte Carlo technique where random data are sampled from an alternative "sampling distribution" and an unbiased estimator is obtained by likelihood ratio weighting. Here we consider estimation of large deviations probabilities via importance sampling. Previous works have shown, for certain special cases, that "exponentially twisted" distributions possess a strong asymptotic optimality property as a sampling distribution. The results of this paper unify and generalize the previous special case results. The analysis is presented in an abstract setting, so the results are quite general and directly applicable to a number of large deviations problems. Our main motivation, however, is to attack sample path problems. To illustrate the application to this class of problems, we consider Mogulskii type sample path problems in some detail.

1. Introduction and summary. Let $\mathbf{P} = \{P_\varepsilon\colon\ \varepsilon > 0\}$ be a family of Borel probability measures on a topological space $\mathscr{X}$. Suppose that $\mathbf{P}$ satisfies a large deviations principle; that is, $P_\varepsilon \to \delta_\mu$ weakly for some $\mu \in \mathscr{X}$, and for a large class of "continuity sets" we have $-\varepsilon \log(P_\varepsilon(E)) \sim I(E)$, where $I(E) > 0$ whenever $\mu \notin \bar{E}$. This paper considers estimation of $P_\varepsilon(E)$ using the Monte Carlo method commonly known as *importance sampling*. Let $X_\varepsilon^{(1)}, \dots, X_\varepsilon^{(L_\varepsilon)}$ be independent samples from a *sampling distribution* Q_ε. Then the estimator is

$$\hat{P}_\varepsilon(E) = \frac{1}{L_\varepsilon} \sum_{l=1}^{L_\varepsilon} 1_E\big(X_\varepsilon^{(l)}\big) \frac{dP_\varepsilon}{dQ_\varepsilon}\big(X_\varepsilon^{(l)}\big), \tag{1}$$

which is well defined and unbiased [i.e., $\mathrm{E}_{Q_\varepsilon}[\hat{P}_\varepsilon(E)] = P_\varepsilon(E)$] if and only if $1_E P_\varepsilon \ll Q_\varepsilon$ [where $1_E P_\varepsilon(\cdot) = P_\varepsilon(\cdot \cap E)$].

Set $L_\varepsilon(Q_\varepsilon) = \min\{L_\varepsilon\colon\ \mathrm{var}_{Q_\varepsilon}[\hat{P}_\varepsilon(E)] \le cP_\varepsilon(E)^2\}$ for some $0 < c < \infty$. Then $L_\varepsilon(Q_\varepsilon) = \lceil v_\varepsilon(E; Q_\varepsilon)/(cP_\varepsilon(E)^2)\rceil$, where

$$v_\varepsilon(E; Q_\varepsilon) \overset{\mathrm{def}}{=} \mathrm{var}_{Q_\varepsilon}\left[1_E(X_\varepsilon)\frac{dP_\varepsilon}{dQ_\varepsilon}(X_\varepsilon)\right] \ge 0$$

and, hence, a good sampling distribution will tend to minimize this single sample variance. Of course, $v_\varepsilon(E; Q_\varepsilon) = 0$ if and only if $1_E\, dP_\varepsilon/dQ_\varepsilon =$ constant

Received January 1994; revised May 1995.

[1] Work supported by NSF Grant NCR-90-03007.

AMS 1991 *subject classifications*. Primary 60F10, 65C05; secondary 93E30.

Key words and phrases. Large deviations, Monte Carlo methods, computer simulations of stochastic systems.

with Q_ε probability 1, and after normalization we find that the zero variance sampling distribution is just the conditional law $Q_\varepsilon(\cdot) = 1_E P_\varepsilon(\cdot)/P_\varepsilon(E) = P_\varepsilon(\cdot|E)$. However, this is not a practical solution. To see why, realize that total cost of computation is actually $L_\varepsilon(Q_\varepsilon) \times$ "per sample costs," where the latter factor includes the cost of numerically generating the samples $X_\varepsilon^{(l)}$ and then evaluating the likelihood ratios in (1). The conditional law blindly minimizes $L_\varepsilon(Q_\varepsilon)$ without regard to per sample costs. The likelihood ratio becomes $(dP_\varepsilon/dQ_\varepsilon)(X_\varepsilon) = P_\varepsilon(E)$ with probability 1 and, hence, the estimator reduces to direct evaluation of $P_\varepsilon(E)$. This is why the variance is zero, but prohibitive numerical cost of direct computation is presumably what leads to consideration of Monte Carlo methods in the first place.

Practical importance sampling must seek a tradeoff between sampling efficiency and implementation complexity. A common strategy is to impose "ease of implementation" by constraining candidate sampling distributions to lie in some natural family. One then attempts to minimize $v_\varepsilon(E; Q_\varepsilon)$ within the constrained family. For example, if P_ε is a Gaussian distribution, then it is natural to constrain Q_ε to also be Gaussian.

Unlike the unconstrained problem, constrained minimization of $v_\varepsilon(E; Q_\varepsilon)$ can be quite difficult. This has led to consideration of the asymptotics, particularly in the context of large deviations problems. Since $P_\varepsilon(E)$ is exponentially small, it is reasonable to suspect that $L_\varepsilon(Q_\varepsilon)$ will be exponentially large as $\varepsilon \downarrow 0$. Thus, the exponential growth rate of $L_\varepsilon(Q_\varepsilon)$ is the natural asymptotic characterization of sampling efficiency within the context of large deviations.

The literature contains several examples of constrained optimizations using this asymptotic efficiency measure. For example, consider $P_n(\sum_{k=1}^n Z_k \geq n\gamma)$, where P_n is the i.i.d. distribution for $X_n = (Z_1, \dots, Z_n)$ determined by the marginal $p(\cdot) = \mathscr{L}(Z_k)$, and $\varepsilon = 1/n$. From an implementation point of view it is desirable to constrain the sampling distribution Q_n to be also of i.i.d. form as determined by a marginal q (such that $q \ll p$). This determines a nonparametric candidate family of sampling distributions. Embedded within is the parametric family determined by the exponentially twisted marginal $p^\alpha(dz) = \exp(\alpha z - \Lambda_Z(\alpha))p(dz)$, where $\Lambda_Z(\alpha) = \log(\mathrm{E}_p[e^{\alpha Z}])$. Early works, particularly Siegmund (1976), had showed that within the one-dimensional exponentially twisted family there is a unique asymptotically optimal solution $q = p^\theta$, where $\Lambda'_Z(\theta) = \gamma$. Bucklew, Ney and Sadowsky (1990) extended the unique optimality of $q = p^\theta$ to the entire nonparametric family (actually, in a Markov rather than i.i.d. framework). Lehtonen and Nyrhinen (1992a, b) proved a similar result for level crossing problems. Other examples and practical applications are found in Asmussen (1985), Ben Letaief and Sadowsky (1992, 1994), Bucklew (1990), Chang, Heidelberger, Juneja and Shahabuddin (1992), Chen, Lu, Sadowsky and Yao (1993), Sadowsky and Bucklew (1990), Sadowsky (1991, 1993) and Sadowsky and Bahr (1991).

Before proceeding with a formal problem statement, we must extend the discussion of practical issues in one last direction. It is typically impossible to evaluate $L_\varepsilon(Q_\varepsilon)$ a priori. For this reason, it is standard practice to implement the sample variance estimator, denoted V_ε, to estimate $v_\varepsilon(E; Q_\varepsilon)$ in tandem with the sample mean estimator (1). As $\hat{P}_\varepsilon(E)$ and V_ε are both implicit functions of L_ε, sampling proceeds up to

$$L_\varepsilon^* = \min\left\{L_\varepsilon : \hat{P}_\varepsilon(E)^2 \le cV_\varepsilon/L_\varepsilon\right\},$$

which is a stopping time random variable with respect to the sequence $X_\varepsilon^{(1)}, X_\varepsilon^{(2)}, \dots$. The behavior of this sequential estimator has been well studied and characterized in the statistical literature. See Woodroofe (1982) and references therein. The stability of the estimator is determined by both the variance of the sample mean and the variance of the sample variance, and the latter depends on the fourth moment of $1_E\, dP_\varepsilon/dQ_\varepsilon$. The implication in the large deviations context is as follows. It would be fruitless to optimize Q_ε by simply minimizing the exponential growth of the second moment if the ratio of the fourth to second moments blows up exponentially. If such an exponential divergence of moments occurs, the sequential estimator will be hopelessly unstable for small values of ε. From another point of view, instability of higher order moments is manifest as an increasingly heavy tailed error distribution as $\varepsilon \downarrow 0$. The reader is directed to Sadowsky (1993) for a more complete discussion of this stability issue, including a numerical example of a low variance yet highly unstable importance sampling estimator. Fortunately, it happens that the asymptotic analysis of any integral moment of the estimator error is no more difficult than variance analysis. For these reasons, we will formulate an asymptotic performance criteron in terms of an arbitrary integral order moment.

The original goal for this work was to obtain a general asymptotic result for "sample path" large deviation problems. This was motivated by efforts to apply the importance sampling method to estimate probabilities of rare trajectories in semiconductor devices: $X(t) = \int_0^t V(s)\, ds$, where the velocity process $V(t)$ is a Markov jump process with drift. In this paper, we will illustrate the sample path issues using the simpler process $X_n(t) = (1/n)S_{\lfloor nt \rfloor}$, where $S_k = \sum_{\kappa=1}^k Z_\kappa$ is a random walk. One type of twisted distribution would make the Z_k independent with twisted marginals $\mathscr{L}(Z_k) = p^{\alpha_k}$, where α_k is a time varying (but deterministic) sequence. We will call this *simple* exponential twisting. An alternative is *sequential* exponential twisting determined by the conditional distributions $\mathscr{L}(Z_k | X_n(s)\colon\ s < k/n) = p^{\alpha_k}$, where $\alpha_k = \alpha_k(X_n(s)\colon\ x < k/n)$. (The history $\{X_n(s)\colon\ s < k/n\}$ depends only on $Z_1, \dots, Z_{k-1}$.) In fact, the sampling distribution used in Asmussen (1985), Lehtonen and Nyrhinen (1992a, b), Sadowsky (1991) and Siegmund (1976) for the level crossing problem is actually a sequential exponential twisting where the α_k depend on the sample path only through the level crossing stopping time. These previous works have attacked importance sampling analysis using Wald's identity. The proofs, however, are cumbersome (several cases must be eliminated), and it is not likely that that approach can be readily generalized, say, to problems involving multidimensional and/or time varying boundaries or position dependent acceleration fields. For example, Asmussen (1985) discussed difficulties when simulating a modified gambler's ruin problem where the betting strategy changes as the gambler's fortune approaches the ruin boundary.

This paper presents an entirely new approach that is generally applicable to large deviations problems involving a convex rate function. As is consistent with modern large deviations theory, we cast the analysis in the abstract setting of locally convex Hausdorff topological vector space. Our main result, Theorem 2, establishes a general necessary condition for asymptotic optimality. Roughly stated, the sequence $\mathbf{Q} = \{Q_\varepsilon\}$ must be locally similar (at certain "points of continuity") to the simple exponentially twisted distribution. This result is similar in character (and proof) to the local nature of large deviations lower bounds. A significant departure from previous results is that Theorem 2 is applicable to *any* sequence $\mathbf{Q}$; we do not consider a particular constrained candidate family. Thus, a candidate family may be tailored to match the specifics of the particular application at hand and then Theorem 2 can be used to identify possible optimal solutions within the selected family.

In addition to the general necessary condition, Theorem 3 presents necessary and sufficient conditions for optimality of simple exponential twisting. It turns out that *simple* exponential twisting is generally *not* asymptotically optimal, except for problems with a special "dominating point" geometry.

The paper is organized as follows. Section 2 is the formal presentation of results. Section 3 then demonstrates the procedure for application of Theorems 2 and 3 to time varying level crossing problems for the process $X_n(t) = (1/n)S_{\lfloor nt \rfloor}$. Proof are deferred to Section 4, and further discussion is found in Section 5.

2. Presentation of results. In this section we present the new results. To do this, we must first review the large deviations setting. Given this background, we then formally state the asymptotic optimality criteria and the new results, Theorems 2 and 3.

Let $\mathscr{X}$ be a locally convex regular Hausdorff topological vector space and let $\mathscr{X}^*$ denote its topological dual endowed with the weak-$*$ topology. For a function $g\colon \mathscr{X}^* \to [-\infty, \infty]$ the *Fenchel transform* is defined as $g^*(x) \overset{\text{def}}{=} \sup_{\lambda \in \mathscr{X}^*}\{\langle \lambda, x\rangle - g(\lambda)\}$ for all $x \in \mathscr{X}$, which is convex and lower semicontinuous. Likewise, for $f\colon \mathscr{X} \to [-\infty, \infty]$, define $f^*(\lambda) \overset{\text{def}}{=} \sup_{x \in \mathscr{X}}\{\langle \lambda, x\rangle - f(x)\}$ for each $\lambda \in \mathscr{X}^*$. If $f(\cdot)$ is convex, lower semicontinuous and $f(\cdot) > -\infty$, then $f^{**}(\cdot) = f(\cdot)$. A point $x \in \mathscr{X}$ is called an *exposed point* of $f(\cdot)$ if there exists a $\lambda_x \in \mathscr{X}^*$ such that

$$f(y) > f(x) + \langle \lambda_x, y - x\rangle \quad \text{for all } y \neq x;$$

λ_x is called an *exposing hyperplane* of $f(\cdot)$ at x. If $g^*(\cdot)$ has an exposed point x with exposing hyperplane λ_x, then we may easily evaluate $g^*(x) = \langle \lambda_x, x\rangle - g(\lambda_x)$ and $g^{**}(\lambda_x) = \langle \lambda_x, x\rangle - g^*(x)$.

Let $\mathbf{P} = \{P_\varepsilon, \varepsilon > 0\}$ be a family of probability measures on $\mathscr{X}$. The family $\mathbf{P}$ is said to satisfy a *large deviations principle* with *rate function* $I\colon \mathscr{X} \to [0, \infty]$ if

$I(\cdot)$ is lower semicontinuous,

(2) $$\liminf_{\varepsilon\to 0} \varepsilon \log(P_\varepsilon(O)) \geq -I(O)$$

for all open $O \subset \mathscr{X}$ and

(3) $$\liminf_{\varepsilon\to 0} \varepsilon \log(P_\varepsilon(C)) \leq -I(C)$$

for all closed $C \subset \mathscr{X}$, where $I(E) \stackrel{\text{def}}{=} \inf_{x\in E} I(x)$. If the level sets $\{x: I(x) \leq l\}$ are compact, then $I(\cdot)$ is called a *good* rate function.

Let X_ε be an $\mathscr{X}$-valued random element and define

(4) $$\Lambda_\varepsilon(\lambda) \stackrel{\text{def}}{=} \varepsilon \log\big(\mathrm{E}_{P_\varepsilon}[\exp(\langle \lambda, X_\varepsilon\rangle/\varepsilon)]\big)$$

and $\Lambda(\lambda) \stackrel{\text{def}}{=} \lim_{\varepsilon\to 0} \Lambda_\varepsilon(\lambda)$ for $\lambda \in \mathscr{X}^*$. The following is taken from Theorem 4.5.10 in Dembo and Zeitouni (1993).

THEOREM 1. *Assume that* $\mathbf{P}$ *satisfies a large deviations principle with convex good rate function* $I(\cdot)$ *and that* $\limsup_{\varepsilon\to 0} \Lambda_\varepsilon(\lambda) < \infty$ *for all* $\lambda \in \mathscr{X}^*$. *Then the limit* $\Lambda(\lambda)$ *exists for all* $\lambda \in \mathscr{X}^*$, *and* $I(\cdot) = \Lambda^*(\cdot)$.

Let $\mathscr{F}$ denote the set of all exposed points of $I(\cdot)$. We will say a Borel set E is a *continuity set* if

(5) $$0 < \inf_{x\in E} I(x) = \inf_{x\in \overline{E}} I(x) = \inf_{x\in E^o\cap\mathscr{F}} I(x) < \infty,$$

and $\gamma \in \overline{E}$ is a *point of continuity* if $I(\gamma) = I(E)$ and there is a sequence $x_n \in E^o \cap \mathscr{F}$ such that $x_n \to \gamma$. A continuity set always has at least one point of continuity.

For any $\lambda \in \mathscr{X}^*$ such that $\Lambda_\varepsilon(\lambda) < \infty$, the *simple twisted distribution* is

(6) $$P_\varepsilon^\lambda(dx) \stackrel{\text{def}}{=} \exp\left(\frac{1}{\varepsilon}[\langle \lambda, x\rangle - \Lambda_\varepsilon(\lambda)]\right) P_\varepsilon(dx).$$

Next, we move on to importance sampling issues. We say that Q_ε is a *candidate* sampling distribution if $1_E P \ll Q_\varepsilon$. Recall that this is the minimal requirement for (1) to be well defined. For a candidate sequence $\mathbf{Q}$ and an integer $\nu \geq 2$, define

(7) $$a_\nu(E,\mathbf{Q}) \stackrel{\text{def}}{=} \limsup_{\varepsilon\to 0} \varepsilon \log(L_\varepsilon(Q_\varepsilon)),$$

where $L_\varepsilon(Q_\varepsilon) = \min\{L_\varepsilon: \mathrm{E}_{Q_\varepsilon}[\hat{P}_\varepsilon(E)^\nu] \leq cP_\varepsilon(E)^\nu\}$ for some $1 < c < \infty$. Clearly $a_\nu(E,\mathbf{Q}) \geq 0$. We say $\mathbf{Q}$ is *ν-efficient*, for a given set E, if $a_\nu(E,\mathbf{Q}) = 0$. This definition of $a_\nu(E,\mathbf{Q})$ neglects the subexponential behavior of L_ε; hence, ν-efficiency is a weaker optimization criteria than the strict minimization. Unlike strict optimization, there may be many sequences $\mathbf{Q}$ that are ν-efficient. As discussed in the Introduction, our goal is to characterize the candidate sequences $\mathbf{Q}$ that are at least 2-efficient, both 2- and 4-efficient when the sequential estimator is used and perhaps also ν-efficient for arbitrary $\nu \geq 2$.

For nontrivial substochastic measures $\mu \ll \xi$, define *differential entropy* (or Kullback–Leibler information) as

$$D(\mu \| \xi) \overset{\text{def}}{=} \int \log\left(\frac{d\mu}{d\xi}\right) d\mu. \tag{8}$$

Let ξ^a be the absolutely continuous Lebesgue component of ξ with respect to μ, so $\xi^a \sim \mu$, and put $\tilde{\xi}^a = \xi^a / \xi^a(\mathscr{X})$. Since $u \log(u)$ is convex, by Jensen's inequality,

$$\begin{aligned} D(\mu \| \xi) &= \xi^a(\mathscr{X}) \int \frac{d\mu}{d\xi} \log\left(\frac{d\mu}{d\xi}\right) d\tilde{\xi}^a \\ &\geq \mu(\mathscr{X}) \log\left(\frac{\mu(\mathscr{X})}{\xi^a(\mathscr{X})}\right) \geq \mu(\mathscr{X}) \log(\mu(\mathscr{X})). \end{aligned}$$

In particular, if μ is a probability measure, then $D(\mu \| \xi) \geq 0$ with equality if and only if $\mu = \xi$. For strictly substochastic measures, the lower bound $D(\mu \| \xi) \geq \mu(\mathscr{X}) \log(\mu(\mathscr{X}))$ is negative.

We are now ready to state the main result.

THEOREM 2. *Assume* $\mathbf{P}$ *satisfies the conditions of Theorem* 1 *and* E *is a continuity set. Let* γ *be a point of continuity. A necessary condition for* ν*-efficiency of any candidate sequence* $\mathbf{Q}$ *is*

$$\limsup_{\substack{x \to \gamma \\ x \in E^o \cap \mathscr{F}}} \liminf_{\varepsilon \to 0} \varepsilon D\left(1_E P_\varepsilon^{\lambda_x} \| Q_\varepsilon\right) = 0, \tag{9}$$

where λ_x *is an exposing hyperplane at each* x.

REMARK 1. The bound $D(1_E P_\varepsilon^{\lambda_x} | Q_\varepsilon) \geq P_\varepsilon^{\lambda_x}(E) \log(P_\varepsilon^{\lambda_x}(E))$ will generally be negative for fixed $\varepsilon > 0$ because $P_\varepsilon^{\lambda_x}(E) < 1$. However, in Section 4 we will show that $P_\varepsilon^{\lambda_x}(E) \to 1$ for any $x \in E^o \cap \mathscr{F}$, and hence, the left-hand side of (9) is always nonnegative.

REMARK 2. In the Introduction it was stated that the necessary condition is a "local condition" that must be checked in the vicinity of points of continuity. Suppose there are two points of continuity $\gamma_1, \gamma_2 \in \mathscr{F}$. Let $\mathbf{Q}_1$ and $\mathbf{Q}_2$ be candidate sequences such that each $\mathbf{Q}_i$ satisfies (9) with $\gamma = \gamma_i$, but not with $\gamma = \gamma_j$, $j \neq i$. Put $Q_\varepsilon(Q_{1,\varepsilon} + Q_{2,\varepsilon})/2$. Then

$$\begin{aligned} D\left(1_E P_\varepsilon^{\lambda_x} \| Q_\varepsilon\right) &= \mathrm{E}_{P_\varepsilon^{\lambda_x}}\left[\log\left(\frac{dP_\varepsilon^{\lambda_x}}{d(Q_{1,\varepsilon} + Q_{2,\varepsilon})/2}\right); E\right] \\ &\leq D\left(1_E P_\varepsilon^{\lambda_x} \| Q_{i,\varepsilon}\right) + \log(2) P_\varepsilon^{\lambda_x}(E) \end{aligned}$$

for both $i = 1$ and 2 and, hence, $\mathbf{Q} = (\mathbf{Q}_1 + \mathbf{Q}_2)/2$ satisfies (9) at *both* points γ_1 and γ_2.

Finally, we turn to the issue of necessary and sufficient conditions for *simple* exponential twisting to be ν-efficient. A point of continuity $\gamma \in \overline{E}$ is called a *dominating point* if there is a $\lambda^* \in \mathscr{X}^*$ such that $I(\gamma) = \langle \lambda^*, \gamma \rangle - \Lambda(\lambda^*)$ and $\overline{E} \subset \mathscr{H}(\gamma, \lambda^*) \stackrel{\text{def}}{=} \{x: \langle \lambda^*, (x - \gamma) \rangle \geq 0\}$. [A dominating point need not be an exposed point, except when $I(\cdot)$ is strictly convex.] Suppose that there is a point of continuity γ such that $I(\gamma) = \langle \lambda^*, \gamma \rangle - \Lambda(\lambda^*)$, and define

$$E_\nu(\gamma, \lambda^*) \stackrel{\text{def}}{=} \{x: I(x) + (\nu - 1)\langle \lambda^*, x - \gamma \rangle \geq I(\gamma)\}. \tag{10}$$

Since $I(x) = \Lambda^*(x) \geq \langle \lambda^*, x \rangle - \Lambda(\lambda^*) = \langle \lambda^*, (x - \gamma) \rangle - I(\gamma)$, we observe that γ is a boundary point of $E_\nu(\gamma, \lambda^*)$. Moreover,

$$E_2(\gamma, \lambda^*) \supset E_3(\gamma, \lambda^*) \supset \cdots \quad \text{and} \quad \bigcap_{\nu=2}^{\infty} E_\nu(\gamma, \lambda^*) = \mathscr{H}(\gamma, \lambda^*).$$

THEOREM 3. *Assume* **P** *satisfies the conditions of Theorem* 1 *and* E *is a continuity set. Fix* $\lambda^* \in \mathscr{X}^*$ *and an integer* $\nu \geq 2$.

(i) *The following conditions are sufficient for* ν*-efficiency of* $\mathbf{P}^{\lambda^*}$: (a) *there is a point of continuity* γ *such that* $I(E) = I(\gamma) = \langle \lambda^*, \gamma \rangle - \Lambda(\lambda^*)$; (b) $\overline{E} \subset E_\nu(\gamma, \lambda^*)$; (c) *either* $\overline{E} \subset \mathscr{H}(\gamma, \lambda^*)$ (*in which case* γ *is a dominating point*) *or* $(1 - \nu)\lambda^*$ *is an exposing hyperplane*.

(ii) *Conditions* (a) *and* (b′) $E^o \cap \mathscr{F} \subset E_\nu(\gamma, \lambda^*)$ *are necessary for* ν*-efficiency of* $\mathbf{P}^{\lambda^*}$.

COROLLARY 1. *Fix* $\lambda^* \in \mathscr{X}^*$ *and suppose that* $\overline{E} = \overline{E^o \cap \mathscr{F}}$. *Then* $\mathbf{P}^{\lambda^*}$ *is* ν*-efficient for all integers* $\nu \geq 2$ *if and only if* E *has dominating point* γ *and* $I(\gamma) = \langle \lambda^*, \gamma \rangle - \Lambda(\lambda^*)$.

When the necessary condition of Theorem 3 fails, it may be possible to construct ν-efficient sequences as convex combinations of exponentially twisted sequences. This avenue is explored in Sadowsky and Bucklew (1990) for the finite-dimensional case with $\nu = 2$. Chen, Lu, Sadowsky and Yao (1993) first proved this type of result in the setting of the finite-dimensional Gaussian case with $\nu = 2$, in which case the boundary $\partial E_2(\gamma, \lambda^*)$ is an ellipsoid.

3. Mogulskii sample path probabilities. Let $\{Z_k\}$ be an i.i.d. sequence of bounded random variables with marginal distribution $p(\cdot)$ and define

$$X_n(t) = \frac{1}{n}\{S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)Z_{\lfloor nt \rfloor + 1}\} \tag{11}$$

for $t \in [0, 1]$, where $S_k = \sum_{\kappa=1}^{k} Z_\kappa$. Take $\mathscr{X}$ to be the Banach space of continuous functions $x: [0, 1] \to \mathbf{R}$ with $x(0) = 0$, endowed with the sup norm topology. Throughout this section, we replace ε with $1/n$ and modify all notation accordingly.

The process X_n really depends on finitely many random variables $Z_1, \dots, Z_n$. The Banach space $\mathscr{X}$ simply provides a convenient setting for studying the large deviations asymptotics as $n \to \infty$. Observe that (11) defines a continuous mapping $\mathbf{R}^n \to \mathscr{X}$. Thus, P_n = the distribution of X_n on $\mathscr{X}$ is induced by the i.i.d. product form distribution

$$p_n(dz_1 \times \cdots \times dz_n) \stackrel{\text{def}}{=} p(dz_1) \times \cdots \times p(dz_n)$$

for $Z_1, \dots, Z_n$. It is this finite-dimensional distribution to which any Monte Carlo method would be applied; hence, we really seek a finite-dimensional sampling distribution $q_n(dz_1 \times \cdots \times dz_n)$, which may or may not have product form. Of course, q_n induces a distribution Q_n on $\mathscr{X}$ via the transformation (11).

Write $\bar{z} = \operatorname{ess\,sup} Z_k < \infty$ and $\underline{z} = \operatorname{ess\,inf} Z_k > -\infty$, and assume that $\underline{z} < \bar{z}$ [so $p(\cdot)$ is not degenerate]. Define $\Lambda_Z(\alpha) \stackrel{\text{def}}{=} \log(\mathrm{E}_p[e^{\alpha Z}])$. Then, under the current assumptions, $\Lambda_Z(\cdot)$ is strictly convex and analytic on $\mathbf{R}$ and $\theta(z) \stackrel{\text{def}}{=} \Lambda_Z'^{-1}(z)$ defines a 1-to-1, strictly increasing and continuous map $(\underline{z}, \bar{z}) \to \mathbf{R}$. Next, define $\Lambda_Z^*(z) = \sup_\alpha\{\alpha z - \Lambda_Z(\alpha)\}$. Then $\Lambda_Z^*(\cdot)$ is a lower semicontinuous convex function and

$$\Lambda_Z^*(z) = \begin{cases} \theta(z)z - \Lambda_Z(\theta(z)), & \text{for } z \in (\underline{z}, \bar{z}), \\ +\infty, & \text{for } z \notin [\underline{z}, \bar{z}]. \end{cases}$$

Write $\operatorname{dom}(\Lambda_Z^*) = \{z\colon \Lambda_Z^*(z) < \infty\}$, so $(\underline{z}, \bar{z}) \subset \operatorname{dom}(\Lambda_Z^*) \subset [\underline{z}, \bar{z}]$. It happens that $\Lambda_Z^*(\bar{z}) < \infty$ if and only if $\mathscr{P}(Z = \bar{z}) > 0$, and likewise for $\underline{z}$. By the strict convexity of $\Lambda_Z(\cdot)$, it turns out that $|\Lambda_Z^{*\prime}(z)| \uparrow \infty$ as $z \uparrow \bar{z}$ or $z \downarrow \underline{z}$; that is, $\Lambda_Z^*(\cdot)$ is "steep." The set of exposed points is precisely the interior interval $(\underline{z}, \bar{z})$, and on this interval we may evaluate the derivatives $\Lambda_Z^{*\prime}(z) = \theta(z)$ and $\Lambda_Z^{*\prime\prime}(z) = 1/\Lambda_Z''(\theta(z))$. Finally, define

$$p^\alpha(dz) \stackrel{\text{def}}{=} \exp(\alpha z - \Lambda_Z(\alpha))p(dz).$$

Then $\mathrm{E}_{p^\alpha}[Z] = \Lambda_Z'(\alpha)$. In particular, $\mathrm{E}_{p^{\theta(z)}}[Z] = z$ for $z \in (\underline{z}, \bar{z})$.

Let $\mathscr{AC} \subset \mathscr{X}$ denote the set of absolutely continuous $x\colon [0,1] \to \mathbf{R}$ such that $x(0) = 0$.

THEOREM 4 (Molgulskii). *The family* $\mathbf{P}$ *satisfies a large deviations principle with good rate function*

$$I(x) = \begin{cases} \int_0^1 \Lambda_Z^*(\dot{x}(t))\,dt, & \text{for } x \in \mathscr{AC}, \\ \infty, & \text{for } x \notin \mathscr{AC}. \end{cases} \tag{12}$$

Moreover, the set of exposed points $\mathscr{F}$ *is the set of all absolutely continuous* $x \in \operatorname{dom}(I) \stackrel{\text{def}}{=} \{x \in \mathscr{AC}\colon I(x) < \infty\}$ *such that* $\dot{x}$ *is a function of bounded variation, and* $I(\cdot)$ *is strictly convex on* $\operatorname{dom}(I)$.

For a proof of Molgulskii's large deviations principle, see Dembo and Zeitouni (1993). The characterization of exposed points is established below.

The topological dual space $\mathscr{X}^*$ consists of the signed Borel measures of finite variation and $\langle \lambda, x \rangle = \int_0^1 x(t)\lambda(dt)$. For $\lambda \in \mathscr{X}^*$, we directly evaluate

$$\langle \lambda, X_n \rangle = \int_0^1 X_n(t)\lambda(dt) = \frac{1}{n}\sum_{k=1}^{n} \alpha_{n,k}(\lambda) Z_k,$$

where

$$\alpha_{n,k}(\lambda) \stackrel{\text{def}}{=} \lambda((k/n, 1]) + \int_{(k-1)/n}^{k/n} (nt - k + 1)\lambda(dt)$$

and, hence,

$$\Lambda_n(\lambda) \stackrel{\text{def}}{=} \frac{1}{n}\log\big(\mathrm{E}_{P_n}[\exp(n\langle \lambda, X_n \rangle)]\big) = \frac{1}{n}\sum_{k=1}^{n} \Lambda_Z(\alpha_{n,k}(\lambda)).$$

It immediately follows that the simple twisted distribution P_n^λ is induced by the product form distribution

(13) $$p_n^\lambda(dz_1 \times \cdots \times dz_n) \stackrel{\text{def}}{=} p^{\alpha_{n,1}(\lambda)}(dz_1) \times \cdots \times p^{\alpha_{n,n}(\lambda)}(dz_n)$$

via the transformation (11). Moreover, letting $n \to \infty$ we obtain the limit

(14) $$\Lambda(\lambda) = \lim_{n \to \infty} \Lambda_n(\lambda) = \int_0^1 \Lambda_Z(\lambda((t, 1]))\, dt.$$

Clearly, $\Lambda(\lambda) < \infty$ for each $\lambda \in \mathscr{X}^*$. Thus, by Theorem 1 we have $I(\cdot) = \Lambda^*(\cdot)$.

Observe that $I(\cdot)$ is strictly convex on $\mathrm{dom}(I)$ because of the strict convexity of $\Lambda_Z^*(\cdot)$. Using $\Lambda_Z^{*\prime}(z) = \theta(z)$, we find that $I(\cdot)$ is Gateaux differentiable on $\mathrm{dom}(I)$ with directional derivative

$$D_{\delta x} I(x) \stackrel{\text{def}}{=} \lim_{\varepsilon \downarrow 0} \frac{I(x + \varepsilon\,\delta x) - I(x)}{\varepsilon} = \int_0^1 \theta(\dot{x}(t))\, \dot{\delta x}(t)\, dt.$$

The function $I(\cdot)$ has Fréchet derivative $\lambda_x \in \mathscr{X}^*$ at x if $D_{\delta x} I(x) = \langle \lambda_x, \delta x \rangle$ for all δx such that $x + \varepsilon\,\delta x \in \mathrm{dom}(I)$ for some $\varepsilon > 0$. Let $\mathscr{F}$ be as stated in Theorem 4. Then by integration by parts, we find that $I(\cdot)$ is Fréchet differentiable on $\mathscr{F}$ and

(15) $$\lambda_x(dt) = \theta(\dot{x}(1))\delta_1(dt) - d\theta(\dot{x}(t))$$

determines a 1-to-1 relationship between $\mathscr{F}$ and $\mathscr{X}^*$. By strict convexity, it follows that $\mathscr{F}$ is precisely the set of exposed points and the measure λ_x given by (15) is the unique exposing hyperplane.

There are a number of problems that can be attacked within the framework of Mogulskii's theorem. Here we consider only the case of a time varying level crossing:

$$E = \{x \in \mathscr{X}: x(t) \geq e(t) \text{ for at least one } t \in [0, 1]\},$$

where $e(t)$ is lower semicontinuous (so E is closed) and we assume $\min_{t \in [0,1]} e(t)/t > \mathrm{E}_p[Z]$. The simple level crossing problem (as considered in the references) is the case $e(t) \equiv a > 0$ and $\mathrm{E}_p[Z] < 0$.

For $\tau \in (0,1]$, define

$$\gamma_\tau(t) \stackrel{\text{def}}{=} \begin{cases} (e(\tau)/\tau)t, & \text{for } t \le \tau, \\ e(\tau) + \mathrm{E}_p[Z](t-\tau), & \text{for } t > \tau. \end{cases}$$

Then $\gamma_\tau \in \mathscr{F}$ whenever $\underline{z} < e(\tau)/\tau < \bar{z}$, and from (15) the exposing hyperplane is $\lambda_\tau(dt) \stackrel{\text{def}}{=} \lambda_{\gamma_\tau}(dt) = \theta(e(\tau)/\tau)\delta_\tau(dt)$ [because $\theta(\mathrm{E}_p[Z]) = 0$]. Thus, $\langle \lambda_\tau, x \rangle = \theta(e(\tau)/\tau)x(\tau)$. By convexity, γ_τ minimizes $I(\cdot)$ over the half-space $\mathscr{H}(\gamma_\tau, \lambda_\tau) = \{x\colon\ x(\tau) \ge e(\tau)\}$, and $I(\mathscr{H}(\gamma_\tau, \lambda_\tau)) = I(\gamma_\tau) = \tau\Lambda_Z^*(e(\tau)/\tau)$. Since $E = \bigcup_{\tau \in (0,1]} \mathscr{H}(\gamma_\tau, \lambda_\tau)$, the points of continuity for E are the sample paths γ_{τ^*}, where

$$\tau^* = \arg\min \tau\Lambda_Z^*(e(\tau)/\tau),$$

which may not be unique. We will write $\gamma = \gamma_{\tau^*}$, $\lambda^* = \lambda_{\tau^*}$ and $\theta^* = \theta(e(\tau^*)/\tau^*)$. The simple twisted distribution is as in (13) with

$$\alpha_{n,k}(\lambda^*) = \begin{cases} \theta^*, & \text{for } k < \lceil n\tau^* \rceil, \\ (n\tau^* - \lfloor n\tau^* \rfloor)\theta^*, & \text{for } k = \lceil n\tau^* \rceil, \\ 0, & \text{for } k > \lceil n\tau^* \rceil. \end{cases}$$

Now consider the Monte Carlo estimation problem. For each $n < \infty$, let $q_n(\cdot)$ denote a candidate joint sampling distribution for $(Z_1, \dots, Z_n)$; in particular, $q_n(\cdot)$ need not be of product form.

Unfortunately, γ is not a dominating point for E. This is clear from the representation $E = \bigcup_{\tau \in (0,1]} \mathscr{H}(\gamma_\tau, \lambda_\tau)$. This argument can be extended to show that the necessary condition of Theorem 3 is violated for any $\nu \ge 2$ and, hence, the simple twisted distribution $p_n^{\lambda^*}$ is *not* ν-efficient.

Thus, we consider the following sequential alternative. On the event $\{X_n \in E\}$, define $T_n \stackrel{\text{def}}{=} \min\{t \in [0,1]\colon X_n(t) \ge e(t)\}$ and $K_n \stackrel{\text{def}}{=} \lceil nT_n \rceil$. On $\{X_n \notin E\}$ put $K_n = n$. Sample the Z_k from an alternative marginal q up to the random stopping time K_n, instead of up to the deterministic time $\lfloor n\tau^* \rfloor$ (as $p_n^{\lambda^*}$ does). After time K_n, revert to sampling from p. At this point, consider any marginal q such that $q \ll p$. Then the joint distribution is

$$\begin{aligned} &q_n(dz_1 \times \cdots \times dz_n) \\ &\quad = \sum_{k=1}^{n} \left[\prod_{\kappa=1}^{k} q(dz_\kappa) \prod_{\kappa=k+1}^{n} p(dz_\kappa) \right] 1_{\{K_n = k\}}(z_1, \dots, z_k). \end{aligned} \tag{16}$$

We now apply the necessary condition of Theorem 2 to prove the following proposition.

PROPOSITION 1. *The sequence* $\mathbf{Q}$ *defined in* (16) *can be* ν*-efficient only if* τ^* *is unique and* $q = p^{\theta^*}$.

PROOF. Let $\gamma = \gamma_{\tau^*}$ denote an "extremal trajectory" as identified above. To apply Theorem 2 we first must construct a sequence of "interior trajectories" $x \to \gamma$, $x \in E^o \cap \mathscr{F}$. Fix $(t_0, x_0) \in [0,1] \times \mathbf{R}$ such that $x_0 > e(t_0)$ and $\mathrm{E}_p[Z]$

$< x_0/t_0 < \bar{z}$, and put

$$x(t) = \begin{cases} (x_0/t_0)t, & \text{for } t \le t_0, \\ x_0 + \mathrm{E}_p[Z](t - t_0), & \text{for } t > t_0. \end{cases}$$

Then $x \in E^o \cap \mathscr{F}$ and the exposing hyperplane is $\lambda_x(dt) = \alpha_0 \delta_{t_0}(dt)$, where $\alpha_0 = \theta(x_0/t_0)$. This construction is illustrated in Figure 1. Observe that the outer limit of (9), $x \to \gamma$, is achieved by letting $(t_0, x_0) \to (\tau^*, e(\tau^*))$.

The argument goes essentially as follows. Under the simple twisted distribution $P_n^{\lambda_x}$ the process $X_n(t)$ tends to follow the trajectory $x(t)$. By the law of large numbers, $K_n \sim n\tau_0$, where $\tau_0 = \min\{t\colon\ x(t) \ge e(t)\}$, as illustrated in Figure 1. Thus, replacing K_n by $\lfloor n\tau_0 \rfloor$,

$$\frac{1}{n} D\big(1_E P_n^{\lambda_x} \| Q_n\big) \sim \mathrm{E}_{P_n^{\lambda_x}}\left[\frac{1}{n}\sum_{k=1}^{\lfloor n\tau_0\rfloor} \log\left(\frac{dp^{\alpha_0}}{dq}(Z_k)\right); E_n\right] \sim \tau_0 D(p^{\alpha_0}\|q).$$

Letting $x \to \gamma$, we have $\tau_0 \to \tau^*$ and $\alpha_0 \to \theta^*$ and, hence, the necessary condition for ν-efficiency boils down to $D(p^{\theta^*}\|q) = 0$, which implies $q = p^{\theta^*}$.

We now formalize the above sketch. In order to apply the law of large numbers, we consider the expectations indexed by n to be evaluated on a common probability space. Let $\{Z_k^a\}$ and $\{Z_k^b\}$ be independent i.i.d. sequences with respective marginals p^{α_0} and p. In the nth expectation, put $Z_k = Z_k^a$ for $k < \lceil nt_0 \rceil$ and $Z_k = Z^b_{k - \lceil nt_0 \rceil}$ for $k > \lceil nt_0 \rceil$, and on this probability space let E_n denote the event $\{X_n \in E\}$. Define $\tilde{K}_n \stackrel{\text{def}}{=} \min\{K_n, \lfloor nt_0 \rfloor\}$. Then

$$\begin{aligned} (17)\qquad & D\big(1_E P_n^{\lambda_n}\|Q_n\big) \\ &= \mathrm{E}\left[\sum_{k=1}^{\tilde{K}_n} \log\left(\frac{dp^{\alpha_0}}{dq}(Z_k^a)\right); E_n\right] \\ &\quad + \mathrm{E}\left[\sum_{k=K_n+1}^{\lfloor nt_0\rfloor} \log\left(\frac{dp^{\alpha_0}}{dp}(Z_k^a)\right) + \log\left(\frac{dp^{\alpha_0^n}}{dp}(Z_{\lceil nt_0\rceil})\right); \right. \\ &\qquad\qquad \left. K_n \le \lfloor nt_0 \rfloor \right] \\ &\quad + \mathrm{E}\left[\log\left(\frac{dp^{\alpha_0^n}}{dq}(Z_{\lceil nt_0\rceil})\right) + \sum_{k=1}^{K_n - \lceil nt_0\rceil} \log\left(\frac{dp}{dq}(Z_k^b)\right); \right. \\ &\qquad\qquad \left. \{K_n > \lfloor nt_0 \rfloor\} \cap E_n \right], \end{aligned}$$

where $\alpha_0^n = (nt_0 - \lfloor nt_0 \rfloor)\alpha_0$. We will consider the preceding three terms separately below. However, hereafter we neglect two two asymptotically negligible terms involving the single random variable $Z_{\lceil nt_0 \rceil}$. Also, observe that $\{K_n \le \lfloor nt_0 \rfloor\} = \{\tilde{K}_n = K_n\} \subset E_n$ in the middle term above.

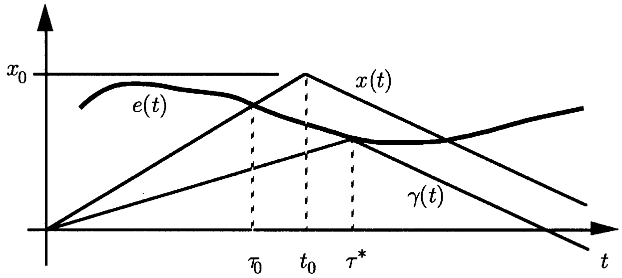

FIG. 1. *Illustration of the extremal trajectory $\gamma = \gamma_{\tau^*}$ and the interior trajectory x.*

The dominant term on the right-hand side of (17) is the first term. Write $\log((dp^{\alpha_0}/dq)(z)) = f^+(z) - f^-(z)$, where $f^+(z)$ and $f^-(z)$ are the positive and negative parts. Using $\tilde{K}_n \le \lfloor nt_0 \rfloor$, we have

$$\frac{1}{n}\mathrm{E}\left[\sum_{k=1}^{\tilde{K}_n} f^-(Z_k^a);\, E_n\right] \le \frac{1}{n}\mathrm{E}\left[\sum_{k=1}^{\lfloor nt_0 \rfloor} f^-(Z_k^a)\right] \le t_0 \mathrm{E}_{p^{\alpha_0}}[f^-(Z)].$$

Next, by the $\{Z_k^a\}$ strong law of large numbers, we have $\mathscr{P}(E_n) \to 1$, $K_n/n \to \tau_0$ a.s., $\tilde{K}_n/n \to \tau_0$ a.s. and, hence,

$$\frac{1}{n}\sum_{k=1}^{\tilde{K}_n} f^+(Z_k^a) = \frac{\tilde{K}_n}{n}\frac{1}{\tilde{K}_n}\sum_{k=1}^{\tilde{K}_n} f^+(Z_k^a) \to \tau_0 \mathrm{E}_{p^{\alpha_0}}[f^+(Z)] \quad \text{a.s.}$$

even when $\mathrm{E}_{p^{\alpha_0}}[f^+(Z)] = +\infty$. Thus, by Fatou's lemma and the previous upper bound, we have

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n}\mathrm{E}&\left[\sum_{k=1}^{\tilde{K}_n} \log\left(\frac{dp^{\alpha_0}}{dq}(Z_k^a)\right);\, E_n\right]\\ &\ge \tau_0 \mathrm{E}_{p^{\alpha_0}}[f^+(Z)] - t_0 \mathrm{E}_{p^{\alpha_0}}[f^-(Z)]\\ &= \tau_0 D(p^{\alpha_0}\|q) - (t_0 - \tau_0)\mathrm{E}_{p^{\alpha_0}}[f^-(Z)].\end{aligned}$$

Next, since the Z_k for $k > K_n$ are independent of K_n, the middle term in (17) is easily evaluated as

$$\begin{aligned}\mathrm{E}&\left[\sum_{k=K_n+1}^{\lfloor nt_0 \rfloor} \log\left(\frac{dp^{\alpha_0}}{dp}(Z_k^a)\right);\, K_n \le \lfloor nt_0 \rfloor\right]\\ &= D(p^{\alpha_0}\|p)\mathrm{E}[(\lfloor nt_0 \rfloor - K_n);\, K_n \le \lfloor nt_0 \rfloor] \ge 0.\end{aligned}$$

Finally, consider the third term in (17). Let

$$g^-(z) = \max\{0, -\log((dp/dq)(x))\}.$$

Since $-u\log u \le 1/e$ we have $\mathrm{E}_p[g^-(Z)] = \mathrm{E}_q[(dp/dq)(Z)g^-(Z)] \le 1/e$. Then

$$\left(\frac{1}{n}\sum_{k=1}^{K_n-\lceil nt_0\rceil} g^-\left(Z_k^b\right)\right)1_{\{K_n>\lfloor nt_0\rfloor\}\cap E_n} \le \frac{1}{n}\sum_{k=1}^{n+\lceil nt_0\rceil} g^-\left(Z_k^b\right)$$

and the random variable on the right-hand side has expectation less than or equal to $(1-t_0)\mathrm{E}_p[g^-(Z)] < \infty$. By the $\{Z_k^a\}$ law of large numbers, we have $P(K_n > \lfloor nt_0\rfloor) \to 0$ and, hence, by dominated convergence, we have

$$\lim_{n\to\infty} \mathrm{E}\left[\frac{1}{n}\sum_{n=\lfloor nt_0\rfloor+1}^{K_n} g^-\left(Z_k^b\right); \{K_n > \lfloor nt_0\rfloor\}\cap E_n\right] = 0.$$

Since $\log((dp/dq)(z)) \ge -g^-(z)$, we conclude that as $n\to\infty$ the limit inferior of the third term in (17) is greater than or equal to 0. [For the purpose of getting a lower bound greater than or equal to 0, we do not even have to consider the positive part $g^+(z)$. However, that limit is also zero whenever $D(p\|q) < \infty$.]

Applying the above results to the three terms of (17), we have

$$\liminf_{n\to\infty}\frac{1}{n}D\left(1_E P_n^{\lambda_n}\|Q_n\right) \ge \tau_0 D(p^{\alpha_0}\|q) - (t_0-\tau_0)\mathrm{E}_{p^{\alpha_0}}[f^-(Z)].$$

Let $x\to\gamma$, or equivalently, $(t_0, x_0) \to (\tau^*, e(\tau^*))$. Then $\alpha_0 = \theta(x_0/t_0) \to \theta(e(\tau^*)/\tau^*) = \theta^*$ and $\tau_0 \to \tau^*$ and, hence,

$$\limsup_{x\to\gamma}\liminf_{n\to\infty}\frac{1}{n}D\left(1_E P_n^{\lambda_x}\|Q_n\right) \ge \tau^* D(p^{\theta^*}\|q).$$

Thus, by Theorem 2, $D(p^{\theta^*}\|q) = 0$ is necessary for ν-efficiency of a sequence $\mathbf{Q}$ of the form (16), which occurs if and only if $q = p^{\theta^*}$. □

Next we address the sufficiency of the condition $q = p^{\theta^*}$ to yields a ν-efficient sequential sampling distribution q_n as defined in (16). This requires additional conditions. Observe that the pair (τ^*, θ^*) is the solution of a min-max problem:

$$I(E) = \min_{\tau\in[0,1]} \tau\Lambda_Z^*\left(\frac{e(\tau)}{\tau}\right) = \min_{\tau\in[0,1]}\sup_{\alpha\in\mathbf{R}}\{\alpha e(\tau) - \tau\Lambda_Z(\alpha)\}.$$

Thus, consider the possibility that (τ^*, θ^*) satisfies the saddle point inequalities

$$\alpha e(\tau^*) - \tau^*\Lambda_Z(\alpha) \le I(E) \le \theta^* e(\tau) - \tau\Lambda_Z(\theta^*) \tag{18}$$

for all $(\tau, \alpha) \in [0,1]\times\mathbf{R}$. The left-hand inequality above follows immediately from Fenchel's inequality: $\Lambda_Z^*(z) \ge \alpha z - \Lambda_Z(\alpha)$ for all $\alpha\in\mathbf{R}$. However, the right-hand saddle point inequality may fail. By the min-max theorem a sufficient (but not necessary) condition is that $e(t)$ be convex.

PROPOSITION 2. *Assume that the saddle point inequalities* (18) *hold. Then the sampling distribution* q_n *defined in* (16) *with* $q = p^{\theta^*}$ *is* ν*-efficient.*

The proof of Proposition 2 is postponed to the next section.

We remark that if the minimizer τ^* is not unique, then one might construct efficient sampling distributions as convex combinations of those of the form (16). Alternatively, the set E might be partitioned into subsets each satisfying the conditions of Proposition 2.

4. Proofs. In this section we prove the results presented in Section 2. Throughout this section we assume that $\mathbf{P}$ satisfies the conditions of Theorem 1 and that E is a continuity set.

Our first task is to relate $a_\nu(E, \mathbf{Q})$ [defined in (7)] to

$$I_\nu(E, \mathbf{Q}) \stackrel{\text{def}}{=} -\limsup_{\varepsilon \to 0} \varepsilon \log\left(\mathrm{E}_{Q_\varepsilon}\left[\left(1_E \frac{dP_\varepsilon}{dQ_\varepsilon}\right)^\nu\right]\right). \tag{19}$$

Recall that $a_\nu(E, \mathbf{Q})$ was defined as the exponential rate of growth of $L_\varepsilon(Q_\varepsilon)$ subject to $\mathrm{E}_{Q_\varepsilon}[\hat{P}_\varepsilon(E)^\nu] = O(P_\varepsilon(E)^\nu)$. The quantity $I_\nu(E, \mathbf{Q})$ is more fundamental because it does not depend on the precise procedure for setting L_ε. By Jensen's inequality, $\mathrm{E}_{Q_\varepsilon}[(1_E\, dP_\varepsilon/dQ_\varepsilon)^\nu] \geq \mathrm{E}_{Q_\varepsilon}[1_E\, dP_\varepsilon/dQ_\varepsilon]^\nu = P_\varepsilon(E)^\nu$. Our first lemma establishes the satisfying fact that ν-efficiency is equivalent to $I_\nu(E, \mathbf{Q}) = \nu I(E)$. That is, the exponential behavior of the νth moment matches that of $P_\varepsilon(E)^\nu$.

LEMMA 1. (i) *The inequality* $I_\nu(E, \mathbf{Q})/\nu \leq I_{\tilde{\nu}}(E, \mathbf{Q})/\tilde{\nu}$ *holds for each* $1 \leq \tilde{\nu} < \nu < \infty$, $I_1(E, \mathbf{Q}) = I(E)$ *and* $I_\nu(E, \mathbf{Q})$ *is a concave function of* ν *on* $[1, \infty)$.

(ii) *For any integer* $\nu \geq 2$,

$$0 \leq a_\nu(E, \mathbf{Q}) = \frac{1}{\nu - 1}[\nu I(E) - I_\nu(E, \mathbf{Q})] \tag{20}$$

and, moreover, $a_\nu(E, \mathbf{Q})$ *is a nondecreasing function of* ν. *In particular,* $I_\nu(E, \mathbf{Q}) \leq \nu I(E)$ *and* $\mathbf{Q}$ *is* ν*-efficient if and only if* $I_\nu(E, \mathbf{Q}) = \nu I(E)$. *Moreover,* ν*-efficiency implies* $\tilde{\nu}$*-efficiency for* $\tilde{\nu} \leq \nu$.

SKETCH OF THE PROOF. For $1 \leq \tilde{\nu} < \nu < \infty$ and $Z \geq 0$, by Jensen's inequality we have $\mathrm{E}[Z^\nu] \geq \mathrm{E}[Z^{\tilde{\nu}}]^{\nu/\tilde{\nu}}$. Applying this to (19) yields $I_\nu(E, \mathbf{Q})/\nu \leq I_{\tilde{\nu}}(E, \mathbf{Q})/\tilde{\nu}$. Likewise, the Hölder inequality

$$\mathrm{E}[Z^{\sigma\nu_1 + (1-\sigma)\nu_2}] \leq \mathrm{E}[Z^{\nu_1}]^\sigma \mathrm{E}[Z^{\nu_2}]^{(1-\sigma)},$$

for $\sigma \in (0, 1)$, yields concavity in ν. That $I_1(E, \mathbf{Q}) = I(E)$ follows immediately from definition (19).

With a little work one finds that

$$\mathrm{E}_{Q_\varepsilon}\left[\hat{P}_\varepsilon(E)^\nu\right] = \sum_{m=1}^{\nu} O(L_\varepsilon^{m-\nu}) \sum_{(j_1, \dots, j_m)} \prod_{i=1}^{m} \mathrm{E}_{Q_\varepsilon}\left[\left(1_E \frac{dP_\varepsilon}{dQ_\varepsilon}\right)^{j_i}\right],$$

where the inner sum is over all integer m-tuples $(j_1, \dots, j_m)$ such that $j_1 \geq j_2 \geq \cdots \geq j_m \geq 1$ and $\sum_{i=1}^m j_i = \nu$. The properties established in part (i) can be applied to show that the $m = 1$ term is exponentially dominant, and setting L_ε so that $O(L_\varepsilon^{1-\nu})\mathrm{E}_{Q_\varepsilon}[(1_E(dP_\varepsilon/dQ_\varepsilon))^\nu] = O(P_\varepsilon(E)^\nu)$ yields (20). See the proofs of Corollary 3 and Theorem 3 in Sadowsky (1993) for a complete development of these arguments. □

REMARK 3. In definition (7) we set L_ε to stabilize the noncentral moment $\mathrm{E}_{Q_\varepsilon}[\hat{P}_\varepsilon(E)^\nu]$, rather than the central moment $\mathrm{E}_{Q_\varepsilon}[(\hat{P}_\varepsilon(E) - P_\varepsilon(E))^\nu]$. Had we used the latter definition, the factors in the expansion of $\mathrm{E}_{Q_\varepsilon}[(\hat{P}_\varepsilon(E) - P_\varepsilon(E))^\nu]$ would be $\mathrm{E}_{Q_\varepsilon}[(1_E(dP_\varepsilon/dQ_\varepsilon) - P_\varepsilon(E))^{j_i}]$, which may have alternating signs. When $I_1(E, \mathbf{Q}) > I_2(E, \mathbf{Q})/2 > \cdots > I_\nu(E, \mathbf{Q})/\nu$, the exponential dominance of the $m = 1$ term is unchanged. [See Corollary 2 in Sadowsky (1993).] When $I_{\tilde{\nu}}(E, \mathbf{Q})/\tilde{\nu} = I_1(E, \mathbf{Q})$ for each $\tilde{\nu} = 2, \dots, \nu$, however, there is the possibility of cancellation due to negative odd order moments. In Sadowsky (1993), subexponential asymptotics were applied to establish dominance of the $m = 1$ when the exponential rates are the same. The use of noncentral moments here is simply a convenient way to avoid the cancellation issue, as in this more general setting we do not have subexponential asymptotics to work with.

REMARK 4. Let V_ε denote the same variance estimator that complements the sample mean (1); that is, $V_\varepsilon \to v_\varepsilon(E; Q_\varepsilon) \stackrel{\text{def}}{=} \mathrm{var}_{Q_\varepsilon}[1_E(dP_\varepsilon/dQ_\varepsilon)]$ as $L_\varepsilon \to \infty$. Suppose that we set L_ε to stabilize the νth moment of the sample variance V_ε, that is, so that $\mathrm{E}_{Q_\varepsilon}[V_\varepsilon^\nu] \sim O(v_\varepsilon(E; Q_\varepsilon)^\nu)$. Following the arguments of Lemma 1, we find that

$$b_\nu(\gamma; \mathbf{Q}) \stackrel{\text{def}}{=} \limsup_{\varepsilon \to 0} \varepsilon \log(L_\varepsilon) = \frac{1}{\nu - 1}\left[\nu I_2(E, \mathbf{Q}) - I_{2\nu}(E, \mathbf{Q})\right] \geq 0.$$

See Section IV of Sadowsky (1993). Now, let us relate the asymptotic optimality $b_\nu(\gamma; \mathbf{Q}) = 0$ for variance estimation to our previous definition of 2ν-efficiency. The 2ν-efficiency is equivalent to $I_{\tilde{\nu}}(E, \mathbf{Q}) = \tilde{\nu} I(E)$ for $1 \leq \tilde{\nu} \leq 2\nu$, and, clearly, this is sufficient to achieve $b_\nu(\gamma; \mathbf{Q}) = 0$. To show necessity, assume $I_2(E, \mathbf{Q})/2 = I_{2\nu}(E, \mathbf{Q})/(2\nu)$. Then, since $I_{\tilde{\nu}}(E, \mathbf{Q})$ is concave in $\tilde{\nu}$, we must have $I_{\tilde{\nu}}(E, \mathbf{Q}) \leq [I_{2\nu}(E, \mathbf{Q})/(2\nu)]\tilde{\nu}$ for all $\tilde{\nu} \geq 1$, in particular, $I(E) = I_1(E, \mathbf{Q}) \leq I_{2\nu}(E, \mathbf{Q})/(2\nu)$. By Lemma 1, $I_{2\nu}(E, \mathbf{Q}) \leq 2\nu I(E)$. Thus, in light of the inequality $I(E) \leq I_{2\nu}(E, \mathbf{Q})/(2\nu)$, we have $I_{2\nu}(E, \mathbf{Q}) = 2\nu I(E)$, that is, $\mathbf{Q}$ is 2ν-efficient. In particular, as noted in the Introduction, 4-efficiency is necessary (and sufficient) to minimize asymptotically the variance of the sample variance.

Our next lemma provides basic upper and lower bounds for $I_\nu(E, \mathbf{Q})$. For any candidate Q_ε, define

$$\Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon) \stackrel{\text{def}}{=} \varepsilon \log\left(\mathrm{E}_{Q_\varepsilon}\left[\exp\left(\frac{\langle \lambda, X_\varepsilon \rangle}{\varepsilon}\right)\left(\frac{dP_\varepsilon}{dQ_\varepsilon}\right)^\nu; E\right]\right) \tag{21}$$

for each $\lambda \in \mathscr{X}^*$. Also define the extended real-valued functions

$$\overline{\Lambda}_\nu(\lambda; E, \mathbf{Q}) \stackrel{\text{def}}{=} \limsup_{\varepsilon \to 0} \Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon)$$

and

$$\underline{\Lambda}_\nu(\lambda E, \mathbf{Q}) \stackrel{\text{def}}{=} \liminf_{\varepsilon \to 0} \Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon),$$

and the measures

$$\xi_{\nu,\varepsilon}(dx; E, Q_\varepsilon) \stackrel{\text{def}}{=} 1_E(x)\left(\frac{dP_\varepsilon}{dQ_\varepsilon}(x)\right)^\nu Q_\varepsilon(dx). \tag{22}$$

A set of measures $\{\xi_\varepsilon(\cdot)\}$ is said to be *exponentially tight* if for any $\alpha > 0$ there exists a compact set K such that $\limsup_{\varepsilon\to 0} \varepsilon \log(\xi_\varepsilon(K^c)) \le -\alpha$.

LEMMA 2. *Let* $\mathbf{Q}$ *be any candidate sequence.*

(i) *We have*

$$I_\nu(E, \mathbf{Q}) \le \inf_{x \in \mathscr{X}} \underline{\Lambda}^*_\nu(x; E, \mathbf{Q}). \tag{23}$$

(ii) *If the measures* $\{\xi_{\nu,\varepsilon}(\cdot)\}$ *are exponentially tight or if E has a dominating point, then*

$$I_\nu(E, \mathbf{Q}) \ge \inf_{x \in \overline{E}} \overline{\Lambda}^*_\nu(x; E, \mathbf{Q}). \tag{24}$$

PROOF. To prove (23), from (21) we get $\varepsilon \log(\mathrm{E}_{Q_\varepsilon}[(1_E(dP_\varepsilon/dQ_\varepsilon))^\nu]) = \Lambda_{\nu,\varepsilon}(0; E, Q_\varepsilon)$ for each $\varepsilon > 0$. Thus,

$$\begin{aligned}-I_\nu(E, \mathbf{Q}) &\ge \liminf_{\varepsilon\to 0} \varepsilon \log\left(\mathrm{E}_{Q_\varepsilon}\left[\left(1_E \frac{dP_\varepsilon}{dQ_\varepsilon}\right)^\nu\right]\right)\\ &= \underline{\Lambda}_\nu(0; E, \mathbf{Q}) \ge \underline{\Lambda}^{**}_\nu(0; E, \mathbf{Q}) = -\inf_{x\in\mathscr{X}} \underline{\Lambda}^*_\nu(x; E, \mathbf{Q}).\end{aligned}$$

To prove part (ii), apply the standard upper bound proof to

$$\limsup_{\varepsilon\to 0} \varepsilon \log(\xi_{\nu,\varepsilon}(C; E, \mathbf{Q})).$$

See Theorem 4.5.3 in Dembo and Zeitouni (1993). The fact that the measures $\{\xi_{\nu,\varepsilon}(\cdot)\}$ are not necessarily probability measures makes no difference. Part (ii) follows by taking $C = \overline{E}$. □

REMARK 5. Part (ii) of Lemma 2 is standard large deviations (upper bound) theory, but part (i) is not. Even if the measures $\{\xi_{\nu,\varepsilon}\}$ do satisfy a logarithmic lower bound for arbitrary open sets, since each $\xi_{\nu,\varepsilon}$ is concentrated on the not necessarily convex set E, the lower bound rate function need not be convex. Thus, one should not expect that a general lower bound will hold with the convex rate function $\underline{\Lambda}^*(\cdot; E, \mathbf{Q})$, but this is no matter here. We only need a lower bound for the fixed set E.

REMARK 6. If $dP_\varepsilon/dQ_\varepsilon \le M_\varepsilon$ for all $x \in E$, where $\limsup_{\varepsilon\to 0} \varepsilon \log(M_\varepsilon) < \infty$, then $\xi_{\nu,\varepsilon}(\cdot) \le M_\varepsilon^{(\nu-1)} 1_E P_\varepsilon(\cdot)$ and exponential tightness of $\{\xi_{\nu,\varepsilon}\}$ follows from the exponential tightness of $\mathbf{P}$. With this observation, we may now prove Proposition 2 from the previous section as a consequence of Lemma 2.

PROOF OF PROPOSITION 2. Observe that

$$\frac{dP_n}{dQ_n} = \exp\big(-\theta^* S_{K_n} + K_n \Lambda_Z(\theta^*)\big).$$

On the event $\{X_n \in E\}$ we have $S_{K_n} \ge ne(K_n/n)$. Thus, since we assume that the right inequality in (18) holds, we have the bound

$$\frac{dP_n}{dQ_n} \le \exp\left(-\left[\theta^* e\left(\frac{K_n}{n}\right) - \left(\frac{K_n}{n}\right)\Lambda_Z(\theta^*)\right]n\right) \le \exp(-I(E)n)$$

on the event $\{X_n \in E\}$. Also, applying this bound to definition (21) (in the next section), we have

$$\begin{aligned}\Lambda_{\nu,n}(\lambda; E, Q_n) &= \frac{1}{n}\log\left(\mathrm{E}_{P_\varepsilon}\left[\exp(\langle \lambda, X_\varepsilon\rangle n)\left(\frac{dP_\varepsilon}{dQ_\varepsilon}\right)^{(\nu-1)}; E\right]\right)\\ &\le \Lambda_n(\lambda) - (\nu-1)I(E).\end{aligned}$$

Letting $n \to \infty$ and applying the resulting bound to the Fenchel transform, we obtain $\overline{\Lambda}_\nu^*(x; E, \mathbf{Q}) \ge \Lambda^*(x) + (\nu - 1)I(E)$ and, hence, $\inf_{x\in\overline{E}} \overline{\Lambda}_\nu^*(x; E, \mathbf{Q}) \ge \nu I(E)$. The proposition now follows by Lemma 2. □

Next, we will need to work with the following twisted measures. First, define

$$Q^\lambda_{\nu,\varepsilon}(dx) \stackrel{\text{def}}{=} \exp\left(\frac{1}{\varepsilon}\left[\langle\lambda, x\rangle - \Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon)\right]\right)1_E(x)\left(\frac{dP_\varepsilon}{dQ_\varepsilon}(x)\right)^{(\nu-1)} P_\varepsilon(dx), \tag{25}$$

which can be viewed as an exponentially twisted version of the measure $\xi_{\nu,\varepsilon}(\cdot)$. Notice that $Q^\lambda_{\nu,\varepsilon}$ is a probability measure even though $\varepsilon_{\nu,\varepsilon}(\cdot)$ generally is not, and that $Q^\lambda_{\nu,\varepsilon}$ and $1_E P_\varepsilon$ are mutually absolutely continuous. Recall the definition of the simple twisted distribution (6). Its conditional law is $\tilde{P}^\lambda_\varepsilon = 1_E P^\lambda_\varepsilon / P^\lambda_\varepsilon(E)$ or, equivalently,

$$\tilde{P}^\lambda_\varepsilon(dx) \stackrel{\text{def}}{=} \exp\left(\frac{1}{\varepsilon}\left[\langle\lambda, x\rangle - \tilde{\Lambda}_\varepsilon(\lambda; E)\right]\right)1_E(x)P_\varepsilon(dx), \tag{26}$$

where

$$\tilde{\Lambda}_\varepsilon(\lambda; E) \stackrel{\text{def}}{=} \varepsilon \log\big(\mathrm{E}_{P_\varepsilon}\left[\exp(\langle\lambda, X_\varepsilon\rangle); E\right]\big). \tag{27}$$

Clearly, $\tilde{\Lambda}_\varepsilon(\lambda; E) \le \Lambda_\varepsilon(\lambda)$. Also, observe from (21) and (25) with $\nu = 1$ that $\tilde{\Lambda}_\varepsilon(\cdot\,; E) = \Lambda_{1,\varepsilon}(\cdot\,; E, Q_\varepsilon)$ and $\tilde{P}^\lambda_\varepsilon = Q^\lambda_{1,\varepsilon}$. (So $Q^\lambda_{1,\varepsilon}$ does not actually depend on Q_ε.)

LEMMA 3. *Fix $x \in \mathscr{F}$ and let λ_x be an exposing hyperplane at x.*

(i) *The sequence $\mathbf{P}^{\lambda_x} \stackrel{\text{def}}{=} \{P_\varepsilon^{\lambda_x}\}$ satisfies the large deviations principle [bounds (2) and (3)] with rate function $I(y;x) \stackrel{\text{def}}{=} I(y) - I(x) - \langle \lambda_x, y - x\rangle$. In particular, we have the exponential convergence $P_\varepsilon^{\lambda_x} \to \delta_x$.*

(ii) *For $x \in E^o \cap \mathscr{F}$, we have $P_\varepsilon^{\lambda_x}(E) \to 1$. For $x \in (\overline{E^o \cap \mathscr{F}}) \cap \mathscr{F}$ (in particular, $x \in \mathscr{F}$ may be a point of continuity), we have $\varepsilon \log(P_\varepsilon^{\lambda_x}(E)) \to 0$. In either case, the sequence $\tilde{\mathbf{P}}^{\lambda_x} \stackrel{\text{def}}{=} \{\tilde{P}_\varepsilon^{\lambda_x}\}$ satisfies the large deviations principle with rate function*

$$\hat{I}(y;x) \stackrel{\text{def}}{=} \begin{cases} I(y;x), & \text{for } y \in \overline{E}, \\ \infty, & \text{otherwise}. \end{cases}$$

In particular, we have the exponential convergence $\tilde{P}_\varepsilon^{\lambda_x} \to \delta_x$.

PROOF. Part (i) is established as part of the proof Baldi's theorem, which holds under the hypothesis of Theorem 1. See Theorem 4.5.20 in Dembo and Zeitouni (1993). We elaborate slightly. Define

$$\Lambda(\lambda;x) \stackrel{\text{def}}{=} \lim_{\varepsilon \to 0} \varepsilon \log\bigl(\mathrm{E}_{P_\varepsilon^{\lambda_x}}[\exp(\langle \lambda, X_\varepsilon\rangle/\varepsilon)]\bigr).$$

It is easily shown that $\Lambda(\lambda;x) = \Lambda(\lambda + \lambda_x) - \Lambda(\lambda_x)$ for all $\lambda \in \mathscr{X}^*$. The expression for $I(\cdot\,;x)$ follows. Moreover, the exposed points of $I(\cdot\,;x)$, are precisely the exposed points of $I(\cdot) = \Lambda^*(\cdot)$, and for an exposed point y, if λ_y is a exposing hyperplane for $I(\cdot)$, then $\lambda_y - \lambda_x$ is a exposing hyperplane for $I(\cdot\,;x)$. Thus, $I(\cdot\,;x)$ and $I(\cdot)$ share the same set of exposed points $\mathscr{F}$.

Since $\tilde{P}_\varepsilon^{\lambda}(B) = P_\varepsilon^{\lambda}(B \cap E)/P_\varepsilon^{\lambda}(E)$, clearly $\varepsilon \log(\tilde{P}_\varepsilon^{\lambda_x}(B)) = \varepsilon \log(P_\varepsilon^{\lambda_x}(B \cap E)) - \varepsilon \log(P_\varepsilon^{\lambda_x}(E))$. For $x \in (\overline{E^o \cap \mathscr{F}}) \cap \mathscr{F}$, the large deviations lower bound of part (i) yields $\varepsilon \log(P_\varepsilon^{\lambda_x}(E)) \to 0$ and for $x \in E^0 \cap \mathscr{F}$, the large deviations upper bound yields $P_\varepsilon^{\lambda_x}(E) \to 1$. Thus, the conclusions of part (ii) follow by part (i). □

LEMMA 4. *Let λ_x denote an exposing hyperplane for $I(\cdot)$ at a point $x \in \mathscr{F}$.*

(i) *For $x \in (\overline{E^o \cap \mathscr{F}}) \cap \mathscr{F}$ we have the limit*

$$\lim_{\varepsilon \to 0} \tilde{\Lambda}_\varepsilon(\lambda_x; E) = \Lambda(\lambda_x). \tag{28}$$

(ii) *For $x \in E^o \cap \mathscr{F}$ we have*

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\bigl[\tilde{\Lambda}_\varepsilon(\lambda_x; E) - \Lambda_\varepsilon(\lambda_x)\bigr] = 0 \tag{29}$$

and

$$\liminf_{\varepsilon \to 0} \tilde{\Lambda}_\varepsilon(\lambda; E) \geq \langle \lambda - \lambda_x, x\rangle + \Lambda(\lambda_x) \tag{30}$$

for all $\lambda \in \mathscr{X}^*$.

PROOF. By (6) and (27), we have

$$P_\varepsilon^{\lambda}(E) = \mathrm{E}_{P_\varepsilon}\bigl[\exp([\langle \lambda, X_\varepsilon\rangle - \Lambda_\varepsilon(\lambda)]/\varepsilon); E\bigr] = \exp\bigl(\bigl[\tilde{\Lambda}_\varepsilon(\lambda;E) - \Lambda_\varepsilon(\lambda)\bigr]/\varepsilon\bigr).$$

For $\lambda = \lambda_x$, (28) and (29) follow by Lemma 3. Since we assume that E is a continuity set, it is evident that the sequence of substochastic distributions $\{1_E P_\varepsilon\}$ satisfies the large deviations lower bound with rate function

$$I(x;E) = \begin{cases} I(x), & \text{for } x \in E^o, \\ \infty, & \text{for } x \notin E^0. \end{cases}$$

Thus, from the proof of Lemma 4.3.4 in Dembo and Zeitouni (1993) (the lower bound part of Varadhan's integral lemma), for any $\lambda \in \mathscr{X}^*$ we have

$$\liminf_{\varepsilon \to 0} \tilde{\Lambda}_\varepsilon(\lambda; E) \geq \sup_{x \in \mathscr{F}} \{\langle \lambda, x\rangle - I(x;E)\} \geq \langle \lambda, x \rangle - I(x),$$

where the last bound holds for $x \in E^o \cap \mathscr{F}$. Let λ_x be an exposing hyperplane at x. By Theorem 1, $I(x) = \Lambda^*(x) = \langle \lambda_x, x\rangle - \Lambda(\lambda_x)$, which in the last display yields (30). □

LEMMA 5. *For $x \in E^o \cap \mathscr{F}$,*

$$\liminf_{\varepsilon \to 0} \varepsilon D\big(\tilde{P}_\varepsilon^{\lambda_x} \| Q_\varepsilon\big) = \liminf_{\varepsilon \to 0} \varepsilon D\big(1_E P_\varepsilon^{\lambda_x} \| Q_\varepsilon\big), \tag{31}$$

where λ_x is an exposing hyperplane.

PROOF. From (6), (26) and (8), we have

$$\begin{aligned} \varepsilon D\big(\tilde{P}_\varepsilon^{\lambda_x} \| Q_\varepsilon\big) &= \varepsilon \mathrm{E}_{P_\varepsilon^{\lambda_x}}\left[\left(\frac{d\tilde{P}_\varepsilon^{\lambda_x}}{dP_\varepsilon^{\lambda_x}}\right)\log\left(\frac{d\tilde{P}_\varepsilon^{\lambda_x}}{dP_\varepsilon^{\lambda_x}}\frac{dP_\varepsilon^{\lambda_x}}{dQ_\varepsilon}\right); E\right] \\ &= \exp\left(\frac{\big[\Lambda_\varepsilon(\lambda_x) - \tilde{\Lambda}_\varepsilon(\lambda_x;E)\big]}{\varepsilon}\right) \\ &\quad \times\Big\{\varepsilon D\big(1_E P_\varepsilon^{\lambda_x} \| Q_\varepsilon\big) + \big(\Lambda_\varepsilon(\lambda_x) - \tilde{\Lambda}_\varepsilon(\lambda_x;E)\big) P_\varepsilon^{\lambda_x}(E)\Big\} \end{aligned}$$

Lemma 5 now follows directly from Lemma 4. □

Next, define

$$\mathscr{M}_\varepsilon(E) \stackrel{\text{def}}{=} \{\mu \in \mathscr{M} \colon \mu \sim 1_E P_\varepsilon\},$$

where $\mathscr{M}$ is the set of all Borel probability measures on $\mathscr{X}$. From definitions (25) and (26) observe that $\tilde{P}_\varepsilon^\lambda, Q_{\nu,\varepsilon}^\lambda \in \mathscr{M}_\varepsilon(E)$. The next lemma provides a key representation.

LEMMA 6. *Let Q_ε be a candidate sampling distribution. Then*

$$\begin{aligned} &\Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon) \\ &\quad = \sup_{\mu \in \mathscr{M}_\varepsilon(E)} \big\{\mathrm{E}_\mu[\langle \lambda, X_\varepsilon\rangle] - \nu\varepsilon D(\mu \| P_\varepsilon) + (\nu - 1)\varepsilon D(\mu \| Q_\varepsilon)\big\} \end{aligned} \tag{32}$$

and when $\Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon) < \infty$, the supremum is obtained at $\mu = Q_{\nu,\varepsilon}^\lambda \in \mathscr{M}_\varepsilon(E)$.

PROOF. For any $\mu \in \mathscr{M}_\varepsilon(E)$, we have $\mu(E) = 1$ and $\mu \ll Q_\varepsilon$ and, hence, by Jensen's inequality,

$$\begin{aligned}\Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon) &= \varepsilon \log\left(\mathrm{E}_{P_\varepsilon}\left[\exp\left(\frac{\langle\lambda, X_\varepsilon\rangle}{\varepsilon}\right)\left(\frac{dP_\varepsilon}{dQ_\varepsilon}\right)^{(\nu-1)}; E\right]\right)\\ &= \varepsilon \log\left(\mathrm{E}_{\mu}\left[\exp\left(\frac{\langle\lambda, X_\varepsilon\rangle}{\varepsilon}\right)\left(\frac{dP_\varepsilon}{d\mu}\right)^{\nu}\left(\frac{d\mu}{dQ_\varepsilon}\right)^{(\nu-1)}\right]\right)\\ &\geq \mathrm{E}_\mu[\langle\lambda, X_\varepsilon\rangle] - \nu\varepsilon\mathrm{E}_\mu\left[\log\left(\frac{d\mu}{dP_\varepsilon}\right)\right] + (\nu-1)\varepsilon\mathrm{E}_\mu\left[\log\left(\frac{d\mu}{dQ_\varepsilon}\right)\right]\\ &= \mathrm{E}_\mu[\langle\lambda, X_\varepsilon\rangle] - \nu\varepsilon D(\mu\|P_\varepsilon) + (\nu-1)\varepsilon D(\mu\|Q_\varepsilon).\end{aligned}$$

Thus,

$$\Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon) \geq \sup_{\mu\in\mathscr{M}_\varepsilon(E)}\left\{\mathrm{E}_\mu[\langle\lambda, X_\varepsilon\rangle] - \nu D(\mu\|P_\varepsilon) + (\mu-1)D(\mu\|Q_\varepsilon)\right\}.$$

If the right-hand side above is $+\infty$, we are done. So, suppose $\Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon) < \infty$. Putting $\mu = Q^\lambda_{\nu,\varepsilon}$ and then using definition (25) to evaluate the $dP_\varepsilon/dQ^\lambda_{\nu,\varepsilon}$, we get

$$\begin{aligned}&\sup_{\mu\in\mathscr{M}_\varepsilon(E)}\left\{\mathrm{E}_\mu[\langle\lambda, X_\varepsilon\rangle] - \nu\varepsilon D(\mu\|P_\varepsilon) + (\nu-1)\varepsilon D(\mu\|Q_\varepsilon)\right\}\\ &\quad\geq \mathrm{E}_{Q^\lambda_{\nu,\varepsilon}}[\langle\lambda, X_\varepsilon\rangle] + \nu\varepsilon\mathrm{E}_{Q^\lambda_{\nu,\varepsilon}}\left[\log\left(\frac{dP_\varepsilon}{dQ^\lambda_{\nu,\varepsilon}}\right)\right] + (\nu-1)\varepsilon\mathrm{E}_{Q^\lambda_{\nu,\varepsilon}}\left[\log\left(\frac{dQ^\lambda_{\nu,\varepsilon}}{dQ_\varepsilon}\right)\right]\\ &\quad= \mathrm{E}_{Q^\lambda_{\nu,\varepsilon}}\left[\varepsilon\log\left(\exp\left(\frac{\langle\lambda, X_\varepsilon\rangle}{\varepsilon}\right)\left(\frac{dP_\varepsilon}{dQ_\varepsilon}\right)^{(\nu-1)}\frac{dP_\varepsilon}{dQ^\lambda_{\nu,\varepsilon}}\right)\right]\\ &\quad= \mathrm{E}_{Q^\lambda_{\nu,\varepsilon}}[\Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon)]\\ &\quad= \Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon).\end{aligned}$$

Combining the last two displays yields the result. □

PROOF OF THEOREM 2. Fix $x \in E^o \cap \mathscr{F}$. Applying Lemma 6 to $\underline{\Lambda}_\nu(\lambda; E, \mathbf{Q}) \overset{\text{def}}{=} \liminf_{\varepsilon\to 0}\Lambda_{\nu,\varepsilon}(\lambda; E, Q_\varepsilon)$ and putting $\mu = \tilde{P}^{\lambda_x}_\varepsilon \in \mathscr{M}_\varepsilon(E)$, we obtain

$$\begin{aligned}\underline{\Lambda}_\nu(\lambda; E, \mathbf{Q}) \geq\ & \liminf_{\varepsilon\to 0}\mathrm{E}_{\tilde{P}^{\lambda_x}_\varepsilon}[\langle\lambda, X_\varepsilon\rangle]\\ &- \nu\limsup_{\varepsilon\to 0}\varepsilon D\big(\tilde{P}^{\lambda_x}_\varepsilon\|P_\varepsilon\big) + (\nu-1)\liminf_{\varepsilon\to 0}\varepsilon D\big(\tilde{P}^{\lambda_x}_\varepsilon\|Q_\varepsilon\big).\end{aligned}$$

Plugging this into the Fenchel transform yields

$$\underline{\Lambda}_\nu^*(x;E,\mathbf{Q}) \stackrel{\text{def}}{=} \sup_{\lambda\in\mathscr{X}^*}\{\langle\lambda,x\rangle - \underline{\Lambda}_\nu(\lambda;E,\mathbf{Q})\}$$
$$\le -\inf_{\lambda\in\mathscr{X}^*}\liminf_{\varepsilon\to 0}\mathrm{E}_{\tilde{P}_\varepsilon^{\lambda_x}}[\langle\lambda,(X_\varepsilon - x)\rangle]$$
$$+\ \nu\limsup_{\varepsilon\to 0}\varepsilon D\big(\tilde{P}_\varepsilon^{\lambda_x}\|P_\varepsilon\big) - (\nu-1)\liminf_{\varepsilon\to 0}\varepsilon D\big(\tilde{P}_\varepsilon^{\lambda_x}\|Q_\varepsilon\big).$$

Now consider each term in this bound separately. By Lemma 3 we have the exponential convergence $\tilde{P}_\varepsilon^{\lambda_x}\to\delta_x$. Thus, $\lim_{\varepsilon\to 0}\mathrm{E}_{\tilde{P}_\varepsilon^{\lambda_x}}[\langle\lambda, X_\varepsilon - x\rangle] = 0$ for all $\lambda\in\mathscr{X}^*$ and, hence, the first term in the above bound vanishes. Next, from definition (26) we have

$$\varepsilon D\big(\tilde{P}_\varepsilon^{\lambda_x}\|P_\varepsilon\big) = \mathrm{E}_{\tilde{P}_\varepsilon^{\lambda_x}}\big[\langle\lambda_x, X_\varepsilon\rangle - \tilde{\Lambda}_\varepsilon(\lambda_x;E)\big].$$

By the exponential convergence $\tilde{P}_\varepsilon^{\lambda_x}\to\delta_x$ from Lemma 3 and the limit $\tilde{\Lambda}_\varepsilon(\lambda_x;E)\to\Lambda(\lambda_x)$ from Lemma 4, we have

$$\lim_{\varepsilon\to 0}\varepsilon D\big(\tilde{P}_\varepsilon^{\lambda_x}\|P_\varepsilon\big) = \langle\lambda_x,x\rangle - \Lambda(\lambda_x) = \Lambda^*(x) = I(x)$$

by Theorem 1. Thus, the previous bound becomes

$$\underline{\Lambda}_\nu^*(x;E,\mathbf{Q}) \le \nu I(x) - (\nu-1)\liminf_{\varepsilon\to 0}\varepsilon D\big(\tilde{P}_\varepsilon^{\lambda_x}\|Q_\varepsilon\big).$$

Applying this to the bound (23) of Lemma 2, we have

$$I_\nu(E,\mathbf{Q}) \le \inf_{y\in\mathscr{X}}\underline{\Lambda}_\nu^*(y;E,\mathbf{Q}) \le \underline{\Lambda}_\nu^*(x;E,\mathbf{Q})$$
$$\le \nu I(x) - (\nu-1)\liminf_{\varepsilon\to 0}\varepsilon D\big(\tilde{P}_\varepsilon^{\lambda_x}\|Q_\varepsilon\big).$$

By the lower semicontinuity of $I(\cdot)$, as $x\to\gamma$ (with $x\in E^o\cap\mathscr{F}$), we have $I(x)\to I(\gamma) = I(E)$ and, hence,

$$I_\nu(E,\mathbf{Q}) \le \nu I(E) - (\nu-1)\limsup_{\substack{x\to\gamma\\ x\in E^o\cap\mathscr{F}}}\liminf_{\varepsilon\to 0}\varepsilon D\big(\tilde{P}_\varepsilon^{\lambda_x}\|Q_\varepsilon\big).$$

From Lemma 2, recall that ν-efficiency is equivalent to $I_\nu(E,\mathbf{Q}) = \nu I(E)$. Application of Lemma 5 completes the proof. □

PROOF OF THEOREM 3. First we show that condition (c) is sufficient for the bound (24) of Lemma 2 to hold. That $\bar{E}\subset\mathscr{H}(\gamma,\lambda^*)$ is sufficient is stated in Lemma 2. If $(1-\nu)\lambda^* = \lambda_z$ for some $z\in\mathscr{F}$, using (22) and (6) we get

$$\xi_{\nu,\varepsilon}(dx) \le \exp\bigg(\frac{1}{\varepsilon}\big[(\nu-1)\Lambda_\varepsilon(\lambda^*) + \Lambda_\varepsilon((1-\nu)\lambda^*)\big]\bigg)P_\varepsilon^{\lambda_x}(dx).$$

The family $\mathbf{P}_\varepsilon^{\lambda_z}$ is exponentially tight by Lemma 2. By Remark 2 following Lemma 2, it follows that the measures $\{\xi_{\nu,\varepsilon}(\cdot)\}$ are exponentially tight, and by Lemma 2 this is a sufficient condition for the bound (24).

Now assume conditions (a), (b) and (c) hold. Plugging (6) into (21), one easily evaluates

$$\Lambda_{\nu,\varepsilon}\big(\lambda; E, P_\varepsilon^{\lambda^*}\big) = \tilde{\Lambda}_\varepsilon(\lambda - (\nu - 1)\lambda^*; E) + (\nu - 1)\Lambda_\varepsilon(\lambda^*).$$

Since $\tilde{\Lambda}_\varepsilon(\,\cdot\,; E) \le \Lambda_\varepsilon(\cdot)$ and taking the limit superior as $\varepsilon \to 0$, we have

$$\begin{aligned}\overline{\Lambda}_\nu\big(\lambda + (\nu - 1)\lambda^*; E, \mathbf{P}^{\lambda^*}\big) &\le \Lambda(\lambda) + (\nu - 1)\Lambda(\lambda^*)\\ &= \Lambda(\lambda) + (\nu - 1)\{\langle \lambda^*, \gamma\rangle - I(E)\},\end{aligned}$$

where the last equality follows because $\Lambda(\lambda^*) = \langle \lambda^*, \gamma\rangle - \Lambda^*(\gamma) = \langle \lambda^*, \gamma\rangle - I(E)$ from (a). Applying the Fenchel transform to both sides of the above inequality yields

$$\begin{aligned}\overline{\Lambda}^*_\nu(x; E, \mathbf{P}^{\lambda^*}) &= \sup_{\lambda \in \mathscr{X}^*} \big\{\langle \lambda + (\nu - 1)\lambda^*, x\rangle - \overline{\Lambda}_\nu\big(\lambda + (\nu - 1)\lambda^*; E, \mathbf{P}^{\lambda^*}\big)\big\}\\ &\ge \sup_{\lambda \in \mathscr{X}^*} \{\langle \lambda, x\rangle - \Lambda(\lambda)\} + (\nu - 1)\{\langle \lambda^*, x - \gamma\rangle + I(E)\}\\ &= I(x) + (\nu - 1)\{\langle \lambda^*, x - \gamma\rangle + I(E)\}.\end{aligned}$$

Since we assume (c), we may apply the bound (24) to get

$$\begin{aligned}\nu I(E) \ge I_\nu(E; \mathbf{Q}) &\ge \inf_{x \in \overline{E}} \overline{\Lambda}^*_\nu(x; E, \mathbf{P}^{\lambda^*})\\ &\ge \inf_{x \in \overline{E}} \{I(x) + (\nu - 1)\langle \lambda^*, x - \gamma\rangle\} + I(E).\end{aligned}$$

Finally, condition (b) [that $\overline{E} \subset E_\nu(\gamma, \lambda^*)$] reduces this to

$$\nu I(E) \ge I_\nu(E; \mathbf{Q}) \ge (\nu - 1)I(\gamma) + I(E) = \nu I(E),$$

where the last equality follows because γ is a point of continuity. Thus, we have ν-efficiency.

Next, we prove part (ii). Let λ_x be an exposing hyperplane for $\Lambda^*(\cdot)$ at any point $x \in E^o \cap \mathscr{F}$. Applying the minorization (30) of Lemma 4, we have

$$\begin{aligned}\underline{\Lambda}_\nu\big(\lambda + (\lambda - 1)\lambda^*; E, \mathbf{P}^{\lambda^*}\big) &= \liminf_{\varepsilon \to 0} \big\{\overline{\Lambda}_\varepsilon(\lambda; E) + (\nu - 1)\Lambda_\varepsilon(\lambda^*)\big\}\\ &\ge \langle \lambda - \lambda_x, x\rangle + \Lambda(\lambda_x) + (\nu - 1)\Lambda(\lambda^*)\end{aligned}$$

for all $\lambda \in \mathscr{X}^*$. Applying this bound to the Fenchel transform yields

$$\begin{aligned}\underline{\Lambda}^*_\nu(x; E, \mathbf{P}^{\lambda^*}) &\le \sup_{\lambda \in \mathscr{X}^*} \big\{\langle \lambda + (\nu - 1)\lambda^*, x\rangle\\ &\qquad - [\langle \lambda - \lambda_x, x\rangle + \Lambda(\lambda_x) + (\nu - 1)\Lambda(\lambda^*)]\big\}\\ &= I(x) + (\nu - 1)\{\langle \lambda^*, x\rangle - \Lambda(\lambda^*)\}\end{aligned}$$

because $\langle \lambda_x, x\rangle - \Lambda(\lambda_x) = \Lambda^*(x) = I(x)$. Applying this to the bound (23) of Lemma 2, we have

$$\begin{aligned}I_\nu(E, \mathbf{Q}) &\le \inf_{y \in \mathscr{X}} \underline{\Lambda}^*_\nu(y; E, \mathbf{P}^{\lambda^*}) \le \underline{\Lambda}^*_\nu(x; E, \mathbf{P}^{\lambda^*})\\ &\le I(x) + (\nu - 1)\{\langle \lambda^*, x\rangle - \Lambda(\lambda^*)\}\end{aligned}$$

for any $x \in E^o \cap \mathscr{F}$. Let $\gamma \in \bar{E}$ be any point of continuity. Letting $x \to \gamma$, we have

$$I_\nu(E, \mathbf{Q}) \le I(\gamma) + (\nu - 1)\{\langle \lambda^*, \gamma \rangle - \Lambda(\lambda^*)\} \le \nu I(\gamma) = \nu I(E).$$

By Lemma 1, $\mathbf{Q}$ is ν-efficient if and only if equality holds in both inequalities of the last display. Equality in the second implies $I(\gamma) = \langle \lambda^*, \gamma \rangle - \Lambda(\lambda^*)$. Thus, condition (a) is necessary. Now return to the previous bound assuming (a) holds. Using $\Lambda(\lambda^*) = \langle \lambda^*, \gamma \rangle - I(\gamma)$, we have

$$I_\nu(E, \mathbf{Q}) \le I(x) + (\nu - 1)[\langle \lambda^*, x - \gamma \rangle + I(\gamma)]$$

for all $x \in E^o \cap \mathscr{F}$. This last upper bound must not take on a value less than $\nu I(E) = \nu I(\gamma)$. Thus, we must have $E^o \cap \mathscr{F} \subset E_\nu(\gamma, \lambda^*)$. That is, condition (b′) is also necessary. □

Proof of Corollary 1. The necessary condition of Theorem 2 provides a point of continuity γ such that conditions (a) and (b′) hold. As noted in Section 2, $E_\nu(\gamma, \lambda^*) \downarrow \mathscr{H}(\gamma, \lambda^*)$ as $\nu \uparrow \infty$; hence, (b′) for all $\nu < \infty$ implies $E^o \cap \mathscr{F} \subset \mathscr{H}(\gamma, \lambda^*)$. Since we assume $\bar{E} = \overline{E^o \cap \mathscr{F}}$ and $\mathscr{H}(\gamma, \lambda^*)$ is closed, we also have $\bar{E} \subset \mathscr{H}(\gamma, \lambda^*)$. Thus, we have conditions (b) and (c), in particular, γ is a dominating point. □

5. Discussion and conclusion. We have adopted the hypothesis of Theorem 1 as the foundation of our analysis; however, these conditions might be relaxed. In particular, the condition that $\Lambda(\lambda) < \infty$ for all $\lambda \in \mathscr{X}^*$ is excessive. The key element in our analysis that must be preserved in any generalization is the exponential convergence $P_\varepsilon^{\lambda_x} \to \delta_x$, which hinges on the assumption that x is an exposed point. This convergence was established in the proof of Theorem 4.5.20 in Dembo and Zeitouni (1993) under the condition that $\Lambda(\delta\lambda_x) < \infty$ for some $\delta > 1$. Thus, our proofs would go through if we take $\mathscr{F}$ to be the set of exposed points that also satisfy this extra condition.

Moreover, it appears unnecessary that $\mathbf{P}$ satisfy a full large deviations principle. $\mathbf{P}$ is said to satisfy a *weak* large deviations principle if the lower bound (2) holds for all open sets, the upper bound (3) holds for compacts. Our analysis really requires that the upper bound (3) hold only for the fixed set E. For example, when the Z_k are not bounded, the proof of exponential tightness fails and Mogulskii's rate function is known to yield only a weak large deviations theorem. However, other methods (i.e., Wald's identity) establish the upper bound (3) for the level crossing set E. In this example it may be difficult to establish the exponential convergence $P_\varepsilon^{\lambda_x} \to \delta_x$ for all exposed points (satisfying the extra condition above). However, in Section 3 we considered only exposed points of the type illustrated in Figure 1, and for these we may directly establish the exponential convergence $P_\varepsilon^{\lambda_x} \to \delta_x$ via Cramér's theorem following the law of large numbers arguments used in the proof of Proposition 1.

Acknowledgment. The author thanks Ofer Zeitouni for stimulating discussion, and for pointing out the simple one line proof of part (ii) of Lemma 2.

REFERENCES

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

BEN LETAIEF, K. and SADOWSKY, J. S. (1992). Computing bit error probabilities for avalanche diode optical receivers by large deviations theory. *IEEE Trans. Inform. Theory* **38** 1162–1169.

BEN LETAIEF, K. and SADOWSKY, J. S. (1994). New importance sampling methods for simulating sequential decoders. *IEEE Trans. Inform. Theory* **39** 1716–1723.

BUCKLEW, J. A. (1990). *Large Deviations Techniques in Decision, Simulation, and Estimation.* Wiley, New York.

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.

CHANG, C-S., HEIDELBERGER, P., JUNEJA, S. and SHAHABUDDIN, P. (1992). Effective bandwidth and fast simulation of ATM intree networks. Research Report RC 18586 (81346), IBM.

CHEN, J-C., LU, D., SADOWSKY, J. S. and YAO, K. (1993). On importance sampling in digital communications. Part I: Fundamentals; Part II: Trellis coded modulation. *IEEE Journal on Selected Areas in Communications* **11** 289–308.

DEMBO, A. and ZEITOUNI, O. (1993). *Large Deviations Techniques and Applications*. Jones and Barlett, Boston.

LEHTONEN, T. and NYHRINEN, H. (1992a). Simulating level-crossing probabilities by importance sampling. *Adv. in Appl. Probab.* **24** 858–874.

LEHTONEN, T. and NYHRINEN, H. (1992b). On asymptotically efficient simulation of ruin probabilities in a Markovian environment. *Scand. Actuarial J.* **1** 60–75.

NEY, P. and NUMMELIN, E. (1987). Markov additive process I: Eigenvalue properties and limit theorems. *Ann. Probab.* **15** 561–592.

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

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

SADOWSKY, J. S. and BAHR, R. K. (1991). Direct sequence spread spectrum multiple access communications with random signature sequences: a large deviations analysis. *IEEE Trans. Inform. Theory* **37** 514–527.

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.

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

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

DEPARTMENT OF ELECTRICAL ENGINEERING
ARIZONA STATE UNIVERSITY
TEMPE, ARIZONA 85287-7206
E-MAIL: sadowski@asu.edu