Evaluation of Large Deviation Probabilities via Importance Sampling

John S. Sadowsky
Department of Electrical Engineering
Arizona State University
Tempe, AZ 85287-7206

Abstract

Error probability is the fundamental performance measure for most communications and detection systems, and many of these can be classified as probabilities of the "large deviation type." Practical examples include very diverse applications, including buffer overflow or cell loss probabilities in queuing systems, near/far bit error rates for a DS-SSMA communications system and ruin probabilities for insurance company investment policy, just to name a few. In this paper we examine the of estimation of very small large deviations probabilities via the Monte Carlo technique commonly known as importance sampling. A new theoretical result on the optimization of the importance sampling technique is presented.

1 Introduction

It is quite often the case that the "error probability" of a communications or detection system can be analyzed within the framework of *large deviations theory*, hereafter, LDT. Like central limit theory, LDT is a collection of asymptotic methods in probability. Central limit theorems deal with the nominal fluctuations of a random sum, while LDT deals with extreme fluctuations. Most well designed detection and communications systems are well equipped to handle nominal noise fluctuations. Hence, it should not be surprising many error probabilities may be cast in the large deviations domain.

There are numerous examples of applications of LDT to practical systems. References given here only lists a few, with an admitted bias towards the contributions of the author. Practical examples include buffer overflow or cell loss probabilities in queuing systems [7], bit error rates for certain optical communications systems [4], ruin probabilities for insurance company investment policy [5], error probabilities for fixed sample size and sequential detectors and decoding error probabilities for coded communications systems [6, 2].

When a probability is truly of the large deviations type, the commonly used Gaussian approximation will generally give very poor estimates. This is well illustrated by the large deviations analysis of a DS-SSMA (direct sequence spread spectrum multiple access) given in [8]. One can formulate a central limit theorem for the case that all interfering signal are of roughly equal power, and in this case Gaussian approximation do generally give good estimates of the bit error probability. However, in the so-called near/far situation, where the multiple access interference is dominated by a few very powerful interferers, central limit theory is not applicable, and Gaussian approximations break down, sometimes over optimistic by orders of magnitude. It is shown in [8] DS-SSMA near/far bit error rates are probabilities of the large deviations type, and large deviations estimates perform quite well in this situation.

2 An Abstract Large Deviations Theorem

Let $\mathbf{P} = \{P_n(\cdot)\}$ be a family of probability distributions on a topological space $\mathcal{X}$. The new result announced in this paper is presented in abstract an abstract setting: $\mathcal{X}$ is a locally convex regular Hausdorff topological vector space. $\mathcal{X}^*$ will denote the topological dual endowed with the weak* topology.

For a function g: $\mathcal{X}^* \to (-\infty, \infty]$ the *Fenchel transform* is defined as

$$g^*(x) \stackrel{\text{def}}{=} \sup_{\lambda \in \mathcal{X}^*} \{\langle \lambda, x\rangle - g(\lambda)\}.$$

$g^*(x)$ is convex and lower semicontinuous. Likewise, for f: $\mathcal{X} \to [-\infty, \infty]$, define $f^*(\lambda) \stackrel{\text{def}}{=}$

1058-6393/95 $4.00

$\sup_{x \in \mathcal{X}} \{\langle\lambda, x\rangle - f(x)\}$ for each $\lambda \in \mathcal{X}^*$. In general, $g(\cdot) \geq g^{**}(\cdot)$. If $g(\cdot)$ is convex and $g(\cdot) > -\infty$, then $g^{**}(\cdot)$ is the lower semicontinuous regularization of $g(\cdot)$.

Let $g(\cdot)$ and $f(\cdot)$ be as above. A point $x \in \mathcal{X}$ is called an *exposed point* of $f(\cdot)$ if there exists a $\lambda_x \in \mathcal{X}^*$ such that

$$f(y) \quad > \quad f(x) \;+\; \langle\lambda_x, (y-x)\rangle$$

for all $y \neq x$. λ_x is called an *exposing hyperplane* of $f(\cdot)$. We will say that x is a *regular* exposed point of $g^*(\cdot)$, with *regular* exposing hyperplane λ_x, if $g(\delta\lambda_x) < \infty$ for some $\delta > 1$. In this case, we have $g^{**}(\lambda_x) = g(\lambda_x)$ and $g^*(x) = \langle\lambda_x, x\rangle - g(\lambda_x)$ at all regular exposed points.

The family $\mathbf{P}$ is said to be exponentially tight if for every $\alpha > 0$ there exists a compact set K_α such that $\limsup_{n\to\infty} \log(P_n(K_\alpha^c))/n \leq -\alpha$.

Let X_n be an $\mathcal{X}$-valued random element having distribution $P_n(\cdot)$. Define

$$\Lambda_n(\lambda) \quad \stackrel{\text{def}}{=} \quad \frac{1}{n}\log\Big(\mathrm{E}_{P_n}\Big[e^{n\langle\lambda, X_n\rangle}\Big]\Big)$$

and $\Lambda(\lambda) \stackrel{\text{def}}{=} \lim_{n\to\infty} \Lambda_n(\lambda)$ (possibly $= +\infty$) for all $\lambda \in \mathcal{X}^*$. Let $\mathcal{F}$ denote the set of all regular exposed points of $\Lambda^*(\cdot)$. For proof of the following theorem, see Theorem 4.5.20 in Dembo and Zeitouni [3].

Theorem 1 (Baldi): Assume $\mathbf{P}$ is exponentially tight and the limit $\Lambda(\lambda)$ exists for all $\lambda \in \mathcal{X}^*$. For any Borel $B \subset \mathcal{X}$,

$$\limsup_{n\to\infty} \frac{1}{n}\log\Big(P_n(B)\Big) \quad \leq \quad -\inf_{x \in \overline{B}} \Lambda^*(x),$$

and

$$\liminf_{n\to\infty} \frac{1}{n}\log\Big(P_n(B)\Big) \quad \geq \quad -\inf_{x \in B^o \cap \mathcal{F}} \Lambda^*(x).$$

We will say that a set B is a *continuity set* if $0 < \inf_{x\in B} \Lambda^*(x) = \inf_{x \in B^o\cap\mathcal{F}} \Lambda^*(x) = \inf_{x\in\overline{B}} \Lambda^*(x) < \infty$. Of course, in this case the upper and lower bounds of Theorem 1 agree. In the case of a continuity set, we will write $\Lambda^*(B) \stackrel{\text{def}}{=} \inf_{x\in B} \Lambda^*(x)$.

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

A point $\gamma \in \overline{B}$ is called a *dominating point* if γ is a point of continuity, and there is a $\lambda^* \in \mathcal{X}^*$ such that $\Lambda^*(\gamma) = \langle\lambda^*, \gamma\rangle - \Lambda(\lambda^*)$ and $\overline{B} \subset \mathcal{H}(\gamma, \lambda^*) \stackrel{\text{def}}{=} \{x\colon \langle\lambda^*, (x-\gamma)\rangle \geq 0\}$.

3 Examples

The most elementary large deviations theorem is Cramér's theorem. Let $\{Z_k\}$ be an i.i.d. sequence of random variables and put

$$P_n(\cdot) \quad = \quad \mathcal{P}\Big(\frac{1}{n}\sum_{k=1}^{n} Z_k \in \cdot\Big).$$

The law of large numbers states that $\frac{1}{n}\sum_{k=1}^n Z_k \to \mu = \mathrm{E}[Z]$, so $P_n(\cdot) \to \delta_\mu(\cdot)$. Cramér's large deviations theorem strengthens this by identifying the rate function $\Lambda_Z^*(x)$ where $\Lambda_Z(\alpha) = \log(\mathrm{E}[\,e^{\alpha Z}\,])$. In particular, for $x \in \{\Lambda_Z^* < \infty\}^o$, we have $\Lambda_Z^*(x) = \{\theta(x)x - \Lambda_Z(\theta(x))\}$ where $\theta(x)$ is the unique solution of the equation $\Lambda_Z'(\theta) = x$. For the case $B = [\gamma, \infty)$ with $\gamma > \mu$, the reader may recognize this theorem determines the asymptotic tightness of the common *Chernoff bound*, which, in the present notation, is just $P_n([\gamma,\infty)) \leq e^{-\Lambda_Z^*(\gamma)n}$.

Cramér's theorem can be viewed as a special case of Theorem 1 in which $\mathcal{X} = \mathbf{R}$.

More interesting deal with sample paths of stochastic process. Let $\mathcal{X} = \{x(t)\colon t \in [0,1]\}$ be the Banach space of continuous functions. $\mathcal{X}^*$ consists of the regular Borel measures: $\langle\lambda, x\rangle = \int_0^1 x(t)\,\lambda(dt)$. Let $\{Z_k\}$ be an i.i.d. sequence of bounded random variables, and construct the continuous time process

$$X_n(t) \quad \stackrel{\text{def}}{=} \quad \frac{1}{n}\left\{\sum_{k=1}^{\lfloor nt\rfloor} Z_k \;+\; (t - \lfloor nt\rfloor/n)\, Z_{\lceil nt\rceil}\right\},$$

and let $P_n(\cdot)$ be its distribution on $\mathcal{X}$. Observe that for any fixed $n < \infty$, $X_n(t)$ is determined by the random variables $(Z_1, .., Z_n)$. The embedding of this problem in the infinite dimensional Banach space $\mathcal{X}$ can be viewed as a convenient way to study the asymptotics as $n \to \infty$. In this case, one finds that

$$\Lambda(\lambda) \quad = \quad \int_0^1 \Lambda_Z(\,\lambda(\,(t,1]\,)\,)\,dt.$$

where $\Lambda_Z(\alpha) = \log(\mathrm{E}[\,e^{\alpha Z}\,])$ as above, and

$$\Lambda^*(x) \quad = \quad \int_0^1 \Lambda_Z^*(\dot{x}(t))\,dt$$

for absolutely continuous $x(t)$, $\Lambda^*(x) = \infty$ otherwise. The set $\mathcal{F}$ of regular exposed points is just the set of absolutely continuous $x(t)$. Since $\mathcal{F}$ is dense in $\mathcal{X}$, the $\cap\mathcal{F}$ may be removed from the lower bound in Theorem 1. In this case, the result is known as *Mogulskii's theorem*.

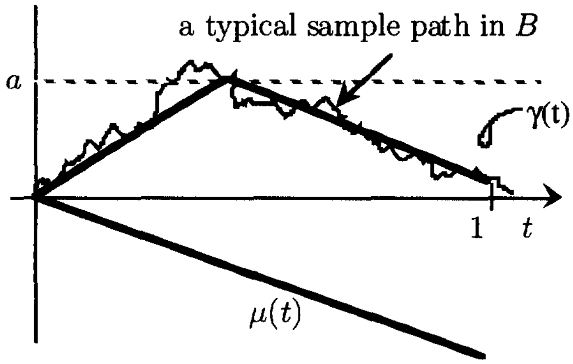

Figure 1: Illustration of Mogulskii's theorem for the level crossing problem.

For example, consider the level crossing problem that characterizes the error event for a sequential detector. In this example we have $E[Z_k] = \mu_0 < 0$, so the limiting sample path is $\mu(t) = \mu_0 t$. The level crossing set is $B_a = \{x\colon\ x(t) \geq a \text{ for some } t \in [0,1]\}$ where $a > 0$. Clearly, then, $P_n(\cdot) \to \delta_\mu(\cdot)$ and $P_n(B_a) \to 0$ because $\mu \notin B_a$. Let $\gamma(t) \in B_a$ be the sample path that minimizes $\Lambda^*(x)$ for $x \in B_a$. Then the solution of the rate function minimization is

$$\gamma(t) = \begin{cases} \Lambda'_Z(\theta)\, t & \text{for } t \leq \tau^* \\ a + \mu_0\,(t - \tau^*) & \text{for } t \geq \tau^* \end{cases}$$

where $\theta > 0$ is the solution of $\Lambda_Z(\theta) = 0$ and $\tau^* = a/\Lambda'_Z(\theta)$. This is illustrated in the Figure 1.

A striking feature of LDT is that in addition to providing estimates of the exponentially small probabilities $P_n(B)$, it also provides a good deal of insight into the nature of the conditional distribution $P_n(\cdot\,|B)$. As illustrated in Figure 1, the extremal sample path $\gamma(t)$ illustrates likely behavior of $X_n(t)$ when the level crossing occurs. More specifically, large deviations analysis provides a certain *exponentially twisted* distribution that, in a sense, approximates the conditional distribution $P_n(\cdot\,|B)$. For one random variable Z with distribution $p(\cdot)$, the exponentially twisted distribution with twisting parameter α is

$$p^{(\alpha)}(dz) = \exp(\,\alpha\, z - \Lambda_Z(\alpha)\,)\, p(dz)$$

For Cramér's theorem, the twisted distribution $P_n^\alpha(\cdot)$ is the i.i.d. sequence distribution with marginal $p^{(\alpha)}(\cdot)$.

Mogulskii's theorem has an associated twisted distribution. Recall that the distribution $P_n(\cdot)$ on the Banach space $\mathcal{X}$ is really determined by the joint distribution of the n random variables $(Z_1,..,Z_n)$ on $\mathbf{R}^n$. For $\lambda \in \mathcal{X}^*$, $P_n^\lambda(\cdot)$ is determined by the product form distribution

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

where

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

4 Importance Sampling

Let $X_n \in \mathcal{X}$ be a random element having distribution $P_n(\cdot)$. Importance Sampling is a Monte Carlo technique in which one samples independent copies $X_n^{(1)},..,X_n^{(L_n)}$ from an alternative *sampling distribution* $Q_n(\cdot)$. The importance sampling estimator for $P_n(B)$ is

$$\widehat{P}_n(B) = \frac{1}{L}\sum_{\ell=1}^{L} 1_B\Big(X_n^{(\ell)}\Big)\,\frac{dP_n}{dQ_n}\Big(X_n^{(\ell)}\Big)$$

where $1_B(x)$ is the indicator function. It is easy to check that this estimator is unbiased: $E[\widehat{P}_n(B)] = P_n(B)$. The goal is to select the sampling distribution $Q_n(\cdot)$ to maximize *sampling efficiency*, that is, minimize the number of samples, L_n, required to obtain a desired *relative precision* $\text{var}[\widehat{P}_n(B)]/P_n(B)^2$.

In addition, the sampling distribution must be implementable; that is, one must be able to implement sampling from this distribution on a digital computer, and then evaluate the likelihood ratio dP_n/dQ_n. It is well known that the conditional distribution $P_n(\cdot\,|B)$ is the most efficient sampling distribution yielding $L_n \equiv 1$ for zero variance, however, this unconstrained optimal solution is generally not implementable.

In the context of estimating large deviations probabilities, it is prudent to seek asymptotically optimal sequences of sampling distributions. Let $\mathbf{Q} = \{Q_n\}$ be a sequence of sampling distributions, and define

$$R(\mathbf{Q}) \stackrel{\text{def}}{=} \limsup_{n\to\infty} \frac{1}{n}\log(\,L_n(Q_n)\,)$$

where $L_n(Q_n)$ is set for a fixed specified relative precision. $R(\mathbf{Q})$ is the exponential growth rate of the sampling cost associated with the sequence $\mathbf{Q}$. In general $R(\mathbf{Q}) \geq 0$, and the sequence $\mathbf{Q}$ is said to be *asymptotically efficient* when $R(\mathbf{Q}) = 0$.

For example, consider the problem of estimating $P_n([\gamma, \infty))$ in the context of Cramér's theorem. In general, $Q_n(\cdot)$ may be any sampling distribution for $X_n^{(\ell)} = (Z_1^{(\ell)}, .., Z_n^{(\ell)})$. However, for the sake of obtaining an implementable solution, suppose that we impose the constraint that $Q_n(\cdot)$ an *i.i.d. distribution*, that is, the joint distribution for $(Z_1^{(\ell)}, .., Z_n^{(\ell)})$ is

$$q_n(dz_1 \times \cdots \times dz_n) \quad = \quad \prod_{k=1}^{n} q(dz_k)$$

as determined by some marginal distribution $q(\cdot)$. Then the entire sequence $\mathbf{Q} = \{Q_n\}$ is determined by the marginal $q(\cdot)$. It has been shown that such a sequence of i.i.d. sampling distribution is asymptotically efficient *if and only if* $q(\cdot) = p^{(\theta)}(\cdot)$. This result was first obtained in [1] (which actually considers the Markov chain version of Cramér's theorem), and generalized in [9] to considers higher order error moments and moments of the sample variance estimator. Related theorems are proved in [5] (for the level crossing problem) and in [7] (for the queuing problem).

The purpose of this paper is to announce the following theorem.

Theorem 2: Assume $\mathbf{P}$ satisfies the conditions of Theorem 1, and let $\mathbf{Q}$ be a candidate sequence for estimating $P_n(B)$ for some continuity set B. Let γ be any point of continuity.

(i) A necessary condition for asymptotic efficiency is

$$\limsup_{\substack{x \to \gamma \\ x \in E^\circ \cap \mathcal{F}}} \ \liminf_{n\to\infty} \ \frac{1}{n} \, \mathrm{E}_{P_n^{\lambda_x}}\left[\log\left(\frac{dP_n^{\lambda_x}}{dQ_n} \right) \ ; \ B \right] = 0$$

where λ_x is a regular exposing hyperplane at each x.

(ii) If $\gamma \in \mathcal{F}$ with regular exposing hyperplane λ_γ, then the necessary condition can be reduced to

$$\liminf_{n\to\infty} \frac{1}{n} D\Big(\widetilde{P}_n^{\lambda_\gamma} \Big\| Q_n \Big) \quad = \quad 0$$

where $\widetilde{P}_n^{\lambda_\gamma}(\cdot) = P_n^{\lambda_\gamma}(\ \cdot\ \cap B)/P_n^{\lambda_\gamma}(E)$, and $D(P\|Q) = \mathrm{E}_P[\log(dP/dQ)]$ is the differential entropy.

The significance of this new result is as follows. The previous optimization results, referenced above, have been restricted to some very specific special case problems (Cramér's theorem and level crossing), and then the admissible class of sampling distributions is restricted, say, to i.i.d. sampling distributions. In contrast, Theorem 2 only requires that $\mathbf{P}$ satisfy Theorem 1, and there are no restrictions on the candidate sequence of sampling distributions $\mathbf{Q}$. In particular, it applies to the setting of Mogulskii's theorem. Thus, this new result is much more general in its scope of applications. The theorem only gives a necessary condition. However, in the previous special case results, sufficiency has been fairly easy to establish, while necessity proofs have been quite difficult. This is especially true for the level crossing problem.

To illustrate the application of Theorem 2 consider the following generalization of the level crossing problem. Construct $X_n(t)$ as above, and consider the sample path set $B = \{x \in \mathcal{X}\colon\ x(t) \geq e(t)$ for at least one $t \in [0,1]\ \}$ where $e(t)$ is a lower semicontinuous boundary satisfying $\min_{t\in[0,1]} e(t)/t \geq \mathrm{E}_p[Z]$. Notice that this corresponds to the ordinary level crossing problem if we take $e(t) \equiv a$. For $\tau \in (0,1]$ define

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

By convexity, it is easily shown that γ_τ minimizes $\Lambda^*(\cdot)$ over $\mathcal{H}(\tau) = \{x\colon\ x(\tau) \geq e(\tau)\}$, and the value of the rate functional is $\Lambda^*(\mathcal{H}(\tau)) = \Lambda^*(\gamma_\tau) = \tau\Lambda_Z^*(e(\tau)/\tau)$. Since $E = \cup_{\tau\in(0,1]}\mathcal{H}(\tau)$, the points of continuity for E are the sample paths γ_{τ^*} where τ^* minimizes $\tau\Lambda_Z^*(e(\tau)/\tau)$. Hereafter, $\tau^* \in (0,1]$ will denote a minimizer of $\tau\Lambda_Z^*(e(\tau)/\tau)$. We will write $\gamma = \gamma_{\tau^*}$, $\lambda^* = \lambda_{\tau^*}$, and $\theta^* = \theta(e(\tau^*)/\tau^*)$. The twisted distribution for $(Z_1, .., Z_n)$ is $p_n^{\lambda^*}$ with

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

It happens that this twisted distribution $p_n^{\lambda^*}$ is *not* exponentially efficient. The reason that the set B is not a half space in $\mathcal{X}$, and even if τ^* is a minimizer of $\tau\Lambda_Z^*(e(\tau)/\tau)$, γ_{τ^*} is not a dominating point. However, we can obtain an asymptotically efficient solution by altering $p_n^{\lambda^*}$ slightly. The solution is to replace the deterministic change time τ^* with the random stopping time $T_n \stackrel{\text{def}}{=} \min\{t \in [0,1]\colon\ X_n(t) \geq e(t)\}$. In general, we consider distributions of the form

$$q_n(dz_1 \times \cdots \times dz_n) \ = \ \sum_{k=1}^{n} \left[\prod_{\kappa=1}^{k} q(dz_\kappa) \prod_{\kappa=k+1}^{n} p(dz_\kappa) \right] 1_{\{K_n = k\}}$$

where $K_n \stackrel{\text{def}}{=} \lceil nT_n \rceil$. In words, this distribution samples the Z_ks from marginal $q(\cdot)$ up to the stopping time K_n, and then reverts back to sampling from $p(\cdot)$. This is precisely the scheme of [5, 7].

Fix $(t_0, x_0) \in [0,1] \times \mathbf{R}$ such that $x_0 > e(t_0)$ and $x_0/t_0 \in \text{dom}(\Lambda_Z^*)^o$, and put

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

Then $x \in E^o \cap \mathcal{F}$, and the regular exposing hyperplane is $\lambda_x(dt) = \alpha_0\, \delta_{t_0}(dt)$ where $\alpha_0 = \theta(x_0/t_0)$. By straightforward likelihood ratio computations, we obtain

$$\frac{1}{n}\, \text{E}_{P_n^{\lambda_x}}\!\left[\log\!\left(\frac{dP_n^{\lambda_x}}{dQ_n} \right) ;\, B \right] = \text{E}\!\left[\frac{1}{n} \sum_{k=1}^{\widetilde{K}_n} \log\!\left(\frac{dp^{\alpha_0}}{dq}(Z_k^a) \right) ;\, B \right]$$

By the $P_n^{\lambda_x}(\cdot)$ law of large numbers, we have

$$\frac{1}{n} \sum_{k=1}^{\widetilde{K}_n} \log\!\left(\frac{dp^{\alpha_0}}{dq}(Z_k^a) \right) \to \tau_0\, D(p^{\alpha_0} \| q) + (t_0 - \tau_0)\, D(p^{\alpha_0} \| p)$$

where $\tau_0 = \min\{t\colon\ (x_0/t_0)\, t \ge e(t)\} < t_0$. Thus, applying Fatou's lemma we have

$$\liminf_{n \to \infty} \frac{1}{n}\, \text{E}_{P_n^{\lambda_x}}\!\left[\log\!\left(\frac{dP_n^{\lambda_x}}{dQ_n} \right) ;\, B \right] \ge \tau_0\, D(p^{\alpha_0} \| q) + (t_0 - \tau_0)\, D(p^{\alpha_0} \| p).$$

Finally, let $(t_0, x_0) \to (\tau^*, e(\tau^*))$, so $x \to \gamma$. Then $\alpha_0 = \theta(x_0/t_0) \to \theta(e(\tau^*)/\tau^*) = \theta^*$ and $\tau_0 \to \tau^*$. Then the above lower bound reduces to $\tau^*\, D(p^{\theta^*} \| q)$. As is well known, $D(p^{\theta^*} \| q) \ge 0$ and $= 0$ if and only if $q(z) \equiv p^{\theta^*}(z)$. Thus, we have shown that there is only one sampling distribution of the prescribed form (sample from $q(z)$ up to the stopping time K_n) that may be asymptotically efficient, the one determined by $q(z) \equiv p^{\theta^*}(z)$.

It can be shown that the following condition is sufficient for asymptotic efficiency of the stopped sampling distribution with $q(z) \equiv p^{\theta^*}(z)$: (τ^*, θ^*) is solution of a min-max problem

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

For lack of space, we omit the arguments.

A major advantage of the above approach is that it is easily generalized to the case that the Z_ks are multi-dimensional ($\in \mathbf{R}^d$), or the case $X_n(t) = \frac{1}{n} \int_0^{nt} Z(s)\, ds$ where $Z(s)$ is a continuous time Markov process. All this is needed is a law of large numbers for the twisted distribution.

References

[1] J. A. Bucklew, P. Ney and J. S. Sadowsky, "Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains," *Journal of Applied Probability*, **27**, pp. 44-59, 1990.

[2] J-C. Chen, D. Lu, J. S. Sadowsky and K. Yao (1993), "On Importance Sampling in Digital Communications Part I: Fundamentals, Part II: Trellis coded modulation," *IEEE Jour. Select. Areas in Commun.*, **11**, pp. 289-308.

[3] A. Dembo and O. Zeitouni (1993), *Large Deviations Techniques and Applications.* Jones and Bartlett, Boston.

[4] K. Ben Leaief and J. S. Sadowsky, "Computing bit-error probabilities by importance sampling," *IEEE Trans. on Information Theory*, **38**, pp. 1162-1169, 1992.

[5] T. Lehtonen and H. Nyrhinen, "Simulating level crossing probabilities by importance sampling," *Advances in Applied Probability*, 1992.

[6] J. S. Sadowsky and J. A. Bucklew "On large deviations theory and asymptotically efficient Monte Carlo estimation," *IEEE Trans. on Information Theory*, **36**, pp. 579-588, 1990.

[7] J. S. Sadowsky, "Large deviations theory and efficient simulation of excessive backlogs in a GI/GI/m queue," *IEEE Trans. on Automatic Control*, **36**, pp. 1383-1394, 1991.

[8] J. S. Sadowsky and R. K. Bahr, "Direct sequence spread spectrum multiple access communications with random signature sequences: a large deviations analysis," *IEEE Trans. on Information Theory*, **37**, pp. 514-527, 1991.

[9] J. S. Sadowsky, "On the optimality and stability of exponential twisting in Monte Carlo estimation," *IEEE Trans. on Information Theory*, **39**, pp. 119-128, 1993.