On asymptotically efficient simulation of large deviation probabilities.

Dieker, A.B.; Mandjes, M.R.H.

DOI
10.1239/aap/1118858638

Publication date
2005

Published in
Advances in Applied Probability

Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

PNA

On Asymptotically Efficient Simulation Of Large Deviation Probabilities

A.B. Dieker, M. Mandjes

REPORT PNA-R0311 SEPTEMBER 03, 2003

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3711

On Asymptotically Efficient Simulation Of Large Deviation Probabilities

A. B. Dieker and M. Mandjes
CWI
P.O. Box 94079
1090 GB Amsterdam, the Netherlands
and
University of Twente
Faculty of Mathematical Sciences
P.O. Box 217
7500 AE Enschede, the Netherlands

ABSTRACT

Consider a family of probabilities for which the decay is governed by a large deviation principle. To find an estimate for a fixed member of this family, one is often forced to use simulation techniques. Direct Monte Carlo simulation, however, is often impractical, particularly if the probability that should be estimated is extremely small. Importance sampling is a technique in which samples are drawn from an alternative distribution, and an unbiased estimate is found after a likelihood ratio correction. Specific exponentially twisted distributions were shown to be good sampling distributions under fairly general circumstances. In this paper, we present necessary and sufficient conditions for asymptotic efficiency of a single exponentially twisted distribution, sharpening previously established conditions. Using the insights that these conditions provide, we construct an example for which we explicitly compute the 'best' change of measure. However, simulation using the new measure faces exactly the same difficulties as direct Monte Carlo simulation. We discuss the relation between this example and other counterexamples in the liturature.

We also apply the conditions to find necessary and sufficient conditions for asymptotic efficiency of the exponential twist in a Mogul′skiĭ sample-path problem. An important special case of this problem is the probability of ruin within finite time.

2000 Mathematics Subject Classification: Primary 65C05; secondary 60F10, 60K10.
Keywords and Phrases: Rare event simulation, asymptotic efficiency, importance sampling, Monte Carlo methods, exponential twisting, large deviations, hitting probabilities.
Note: The research of A. B. Dieker was supported by the Netherlands Organization for Scientific Research (NWO) under grant 631.000.002.

1. Introduction

Given a probability distribution ν, we are interested in estimating a rare event probability $\nu(A)$. In direct Monte Carlo methods, the usual estimator is the proportion of times that A occurs in a certain number of independent samples from ν. However, an inherent problem of this approach is that many samples are needed to obtain a reliable estimate for $\nu(A)$. In fact, the required computing time for estimating $\nu(A)$ may exceed any reasonable limit.

As an important special case, direct Monte Carlo methods are inappropriate for simulating large deviation probabilities. A family of probability measures $\{\nu_\epsilon : \epsilon > 0\}$ is said to satisfy a *large deviation principle* (LDP) if $\nu_\epsilon(A)$ decays exponentially as $\epsilon \to 0$ for a wide class of sets

A. Given such a family, we refer to a probability of the form $\nu_{\epsilon_0}(A)$ for some $\epsilon_0 > 0$ and some event A as a *large deviation probability*. Probabilities of this type are encountered in many fields, e.g., statistics, operations research, information theory, and financial mathematics.

A widely used technique to estimate rare-event probabilities is *importance sampling*. In importance sampling, one samples from a probability measure λ different from ν_{ϵ_0}, such that the ν_{ϵ_0}-rare event becomes λ-likely. A practical and widely used choice for λ is a so-called *exponentially twisted* distribution; still, there is freedom to choose any distribution within this family. To evaluate the changes of measure, efficiency criteria have been developed. In this paper, we use the *asymptotic efficiency* criterion.

In some specific cases, exponentially twisting is known to be asymptotically efficient. In his seminal paper, Siegmund (1976) gives a result of this type for the estimation of the probability that a random walk exceeds a level $b > 0$ before dropping below some fixed $a \leq 0$, where he lets $b \to \infty$ (provided this probability is non-zero). A problem related to Siegmund's is the estimation of a level-crossing probability $P(X_1 + \ldots + X_n > M$, for some $n)$ for real-valued i.i.d. random variables $X_1, X_2, \ldots$. For the estimation of this probability, Lehtonen and Nyrhinen (1992b) show the asymptotic efficiency of the exponential twist in the regime $M \to \infty$; they also establish uniqueness properties of the exponential twist. Related results in a more general Markovian setting are obtained by Asmussen (1989) and Lehtonen and Nyrhinen (1992a). Collamore (2002) extends this to a multi-dimensional setting.

Another example for which an exponential twist is known to yield asymptotic efficiency relates to the 'Cramér-type' probability $P(X_1 + \ldots + X_n \geq \gamma n)$ for $\gamma > \mathbb{E}X_1$, where $n \to \infty$. In case $X_1, X_2, \ldots$ have a special Markovian structure, this was found by Bucklew et al. (1990). Sadowsky (1993) focuses on stability issues in the special case of i.i.d. random variables. To estimate $P(S_n \geq \gamma n)$ for generally distributed S_n, Sadowsky and Bucklew (1990) show that that there exists an asymptotically efficient exponential twist if $\{S_n/n\}$ satisfies the conditions of the Gärtner-Ellis theorem; this is also observed by Szechtman and Glynn (2002).

Recently, it was noted that a successful application of an importance sampling distribution based on large deviation theory critically depends on the specific problem at hand. In particular, Glasserman and Wang (1997) give variations on both the level-crossing problem and the Cramér-type problem, and show that exponential twists can be inefficient if the rare event A is not so nice. In fact, they obtain the stronger result that the so-called relative error can even become unbounded in these examples. Similar observations have been made by Glasserman and Kou (1995) in a queueing context.

Given the examples of efficient and inefficient simulation with exponentially twisted distributions, a natural question is whether there exist conditions for asymptotic efficiency. Sadowsky and Bucklew (1990) give a sufficient condition for asymptotic efficiency of the exponential twist for simulating $P(S_n/n \in A)$ in the case that $\{S_n/n\}$ satisfies the conditions of the Gärtner-Ellis theorem. Both necessary and sufficient conditions are established by Sadowsky (1996) in a general abstract setting. Sadowsky also notes that his sufficiency condition generalizes the earlier results of Sadowsky and Bucklew (1990).

The main contribution of the present paper is that it improves the conditions stated by Sadowsky (1996). The improvements are twofold. On the one hand, we relax the underlying *assumptions* that are needed for a necessary and a sufficient condition to hold. Most notably, we do not require convexity of the large deviation rate function. On the other hand, we sharpen the necessary and sufficient *conditions* themselves, i.e., we show that our necessary condition is weaker and that our sufficient condition is stronger.

To establish the results, we employ a number of standard arguments from large deviation theory. In particular, we rely on a basic large deviation result known as Varadhan's Integral Lemma or Laplace Principle. It is not new to apply this lemma to derive efficiency properties of rare event simulation; see Glasserman et al. (1999) and Dupuis and Wang (2002). However, it was not used in Sadowsky's general abstract setting to find general necessary and sufficient conditions. This is accomplished in the present paper.

The transparency of the proofs, as well as the simplicity of our conditions, enable us to derive some interesting and appealing new results. Indeed, we show that our conditions are 'tight' in the sense that the necessary condition coincides with the sufficient condition under a weak additional assumption.

Furthermore, we address the uniqueness of an asymptotically efficient exponential twist under certain convexity conditions. We show that there is only one exponentially twisted distribution that can be asymptotically efficient. This candidate twist turns out to be the twist suggested by large deviation theory. However, some problems rise if this change of measure is not asymptotically efficient, as we illustrate in a simple example. We also discuss the relation between our example and the counterexamples studied by Glasserman and Wang (1997) and Dupuis and Wang (2002).

To illustrate the use of our conditions, we apply them to a well-known problem, for which we derive new results. In fact, we obtain an equivalent characterization of asymptotic efficiency in the Mogul′skiĭ time-varying level-crossing problem studied by Sadowsky (1996).

The paper is organized as follows. After providing the necessary preliminaries in Section 2, we state and prove our necessary and sufficient conditions in Section 3. We apply these conditions to the Mogul′skiĭ sample-path problem in Section 4. Section 5 relates the use of a single exponential twist to other approaches.

2. Preliminaries

This section provides the basic background on importance sampling and asymptotic efficiency, and discusses their relationship with large deviation techniques. For a more detailed discussion on importance sampling and asymptotic efficiency, see Asmussen and Rubinstein (1995), Heidelberger (1995), and references therein. Valuable sources for large deviation techniques are the books by Dembo and Zeitouni (1998) and Deuschel and Stroock (1989). A good introduction to both large deviation theory and applications is the monograph by den Hollander (2000).

2.1 Importance sampling

Let $\mathcal{X}$ be a topological space, equipped with some σ-field $\mathcal{B}$ containing the Borel σ-field. Given a probability measure ν on $(\mathcal{X}, \mathcal{B})$, we are interested in the simulation of the ν-probability of a given event $A \in \mathcal{B}$, where $\nu(A)$ is small. The idea of importance sampling is to sample from a different distribution on $(\mathcal{X}, \mathcal{B})$, say λ, for which A occurs more frequently. This is done by specifying a measurable function $d\lambda/d\nu : \mathcal{X} \to [0, \infty]$ and by setting

$$\lambda(B) := \int_B \frac{d\lambda}{d\nu} d\nu. \tag{2.1}$$

Since λ must be a probability measure, $d\lambda/d\nu$ should integrate to unity with respect to ν.

Assuming equivalence of the measures ν and λ, set $d\nu/d\lambda := (d\lambda/d\nu)^{-1}$ and note that

$$\nu(A) = \int_A \frac{d\nu}{d\lambda} d\lambda = \int_{\mathcal{X}} \mathbf{1}_A \frac{d\nu}{d\lambda} d\lambda,$$

where $\mathbf{1}_A$ denotes the indicator function of A. The importance sampling estimator $\widehat{\nu_\lambda(A)}$ of $\nu(A)$ is found by drawing N independent samples $X^{(1)}, \ldots, X^{(N)}$ from λ:

$$\widehat{\nu_\lambda(A)} := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{X^{(i)} \in A\}} \frac{d\nu}{d\lambda}(X^{(i)}).$$

It is clear that $\widehat{\nu_\lambda(A)}$ is an unbiased estimator, i.e., $\mathbb{E}_\lambda \widehat{\nu_\lambda(A)} = \nu(A)$. However, one has the freedom to choose an efficient distribution λ in the sense that the variance of the estimator is small. In particular, it is of interest to find the change of measure that minimizes this variance. Since $\widehat{\nu_\lambda(A)}$ is by construction unbiased, it is equivalent to minimize

$$\int_A \left(\frac{d\nu}{d\lambda}\right)^2 d\lambda = \int_{\mathcal{X}} \mathbf{1}_A \left(\frac{d\nu}{d\lambda}\right)^2 d\lambda.$$

It is not difficult to see that a zero-variance estimator is found by letting λ be the conditional distribution of ν given A [see, e.g., Heidelberger (1995)]. However, the resulting estimator is infeasible for simulation purposes, since then $d\nu/d\lambda$ depends on the *unknown* quantity $\nu(A)$. This motivates the use of another optimality criterion, *asymptotic efficiency*.

2.2 Asymptotic efficiency

To formalize the concept of asymptotic efficiency, we introduce some notions that are extensively used in large deviation theory.

A function $I : \mathcal{X} \to [0, \infty]$ is said to be lower semicontinuous if the level sets $\Phi_I(\alpha) := \{x : I(x) \leq \alpha\}$ are closed subsets of $\mathcal{X}$ for all $\alpha \in [0, \infty)$. The interior and closure of a set $B \subseteq \mathcal{X}$ are denoted by B^o and $\overline{B}$ respectively.

Definition 1 *A function $I : \mathcal{X} \to [0, \infty]$ is called a* rate function *if it is lower semicontinuous. If $\Phi_I(\alpha)$ is compact for every $\alpha \geq 0$, I is called a* good rate function.

*A set $B \in \mathcal{B}$ is called an I-*continuity set *if* $\inf_{x \in B^o} I(x) = \inf_{x \in B} I(x) = \inf_{x \in \overline{B}} I(x)$.

The central notion in large deviation theory is the large deviation principle.

Definition 2 *A family of probability measures $\{\nu_\epsilon : \epsilon > 0\}$ on $(\mathcal{X}, \mathcal{B})$ satisfies a large deviation principle (LDP) with rate function I if for all $B \in \mathcal{B}$,*

$$-\inf_{x \in B^o} I(x) \leq \liminf_{\epsilon \to 0} \epsilon \log \nu_\epsilon(B) \leq \limsup_{\epsilon \to 0} \epsilon \log \nu_\epsilon(B) \leq -\inf_{x \in \overline{B}} I(x).$$

Throughout this paper, we assume that the family $\{\nu_\epsilon\}$ satisfies an LDP. We fix some *rare event* $A \in \mathcal{B}$, i.e., $\inf_{x \in A^o} I(x) > 0$, implying that $\nu_\epsilon(A)$ decays exponentially as $\epsilon \to 0$. Since $\nu(A)$ is supposed to be a large deviation probability, we have $\nu = \nu_{\epsilon_0}$ for some ϵ_0.

The definition of asymptotic efficiency is related to the so-called *relative error*. Consider an i.i.d. sample $X_{\lambda_\epsilon}^{(1)}, \ldots, X_{\lambda_\epsilon}^{(N)}$ from an importance sampling distribution λ_ϵ. The relative error of the importance sampling estimator

$$\widehat{\nu_{\lambda_\epsilon}(A)}_N := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\left\{X_{\lambda_\epsilon}^{(i)} \in A\right\}} \frac{d\nu_\epsilon}{d\lambda_\epsilon} \left(X_{\lambda_\epsilon}^{(i)}\right) \tag{2.2}$$

is defined as

$$\eta_N(\lambda_\epsilon, A) := \frac{\mathbb{V}\mathrm{ar}_{\lambda_\epsilon} \widehat{\nu_{\lambda_\epsilon}(A)}_N}{\nu_\epsilon(A)^2} = \frac{\mathbb{E}_{\lambda_\epsilon} \left(\widehat{\nu_{\lambda_\epsilon}(A)}_N\right)^2}{\nu_\epsilon(A)^2} - 1.$$

The idea behind this definition is that the square root of the relative error is proportional to the width of a confidence interval relative to the (expected) estimate itself; hence, it measures the variability of $\widehat{\nu_{\lambda_\epsilon}(A)}_N$.

For asymptotic efficiency, the number of samples required to obtain a prespecified relative error should vanish on an exponential scale. Set $N_{\lambda_\epsilon}^* := \inf\{N \in \mathbb{N} : \eta_N(\lambda_\epsilon, A) \le \eta_{\max}\}$.

Definition 3 *An importance sampling family* $\{\lambda_\epsilon\}$ *is called* asymptotically efficient *if*

$$\limsup_{\epsilon \to 0} \epsilon \log N_{\lambda_\epsilon}^* = 0, \tag{2.3}$$

for some given maximal relative error $0 < \eta_{\max} < \infty$*.*

In the literature, asymptotic efficiency is sometimes referred to as *asymptotic optimality*, *logarithmic efficiency*, or *weak efficiency*.

We briefly relate Definition 3 to other frequently used definitions for asymptotic efficiency. By definition of $\eta_N(\lambda_\epsilon, A)$, we have

$$N_{\lambda_\epsilon}^* = \inf \left\{ N \in \mathbb{N} : \frac{1}{N} \int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon}\right)^2 d\lambda_\epsilon \le (\eta_{\max} + 1)\nu_\epsilon(A)^2 \right\} = \left\lceil \frac{\int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon}\right)^2 d\lambda_\epsilon}{(\eta_{\max} + 1)\nu_\epsilon(A)^2} \right\rceil. \tag{2.4}$$

Equation (2.4) implies

$$\limsup_{\epsilon \to 0} \epsilon \log N_{\lambda_\epsilon}^* \le \limsup_{\epsilon \to 0} \epsilon \log \int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon}\right)^2 d\lambda_\epsilon - 2 \liminf_{\epsilon \to 0} \epsilon \log \nu_\epsilon(A), \tag{2.5}$$

with equality if the limit $\lim_{\epsilon \to 0} \epsilon \log \nu_\epsilon(A)$ exists [see, e.g., Section 2.4 of Royden (1968)]. Sufficient for the existence of this limit is that A be an I-continuity set; in that case, $\lim_{\epsilon \to 0} \epsilon \log \nu_\epsilon(A) = -\inf_{x \in A} I(x)$. In many applications, A is indeed an I-continuity set, in which case asymptotic efficiency is equivalent to

$$\limsup_{\epsilon \to 0} \epsilon \log \int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon}\right)^2 d\lambda_\epsilon \le -2 \inf_{x \in \overline{A}} I(x). \tag{2.6}$$

By similar arguments, one can also readily see that

$$\limsup_{\epsilon \to 0} \frac{\log \int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon}\right)^2 d\lambda_\epsilon}{\log \nu_\epsilon(A)} \ge 2$$

is equivalent to asymptotic efficiency when A is an I-continuity set.

3. The efficiency of exponential twisting

This section investigates the asymptotic efficiency of the estimators that are based on an exponential twist. After formalizing the imposed assumptions, we state our necessary and sufficient conditions for asymptotic efficiency of an exponentially twisted change of measure. The relation between these conditions and the conditions developed by Sadowsky (1996) is provided in Subsection 3.2. Under specific convexity assumptions, some intuitively appealing theorems can be proven, see Subsection 3.3. The insights gained by the theoretical results culminate in a counterintuitive example of the Cramér-type, which is the subject of Subsection 3.4.

3.1 Necessary and sufficient conditions

Let $\mathcal{X}$ be a topological space and $\mathcal{B}$ be a σ-field on $\mathcal{X}$ containing the Borel σ-field. We assume that $\mathcal{X}$ is also a vector space, but not necessarily a topological vector space. Throughout this section, we fix a rare event $A \in \mathcal{B}$ and a continuous linear functional $\xi : \mathcal{X} \to \mathbb{R}$. Having a topological vector space in mind, we write $\langle \xi, \cdot \rangle$ for $\xi(\cdot)$. We are given a family $\{\nu_\epsilon\}$ of probability measures on $(\mathcal{X}, \mathcal{B})$.

Assumption 1 *Assume that*

(i) $\mathcal{X}$ *is a vector space endowed with some regular Hausdorff topology,*

(ii) $\{\nu_\epsilon\}$ *satisfies the LDP with a good rate function* I*,*

(iii) it holds that

$$\lim_{M\to\infty} \limsup_{\epsilon\to 0} \epsilon \log \int_{\{x\in\mathcal{X}:\langle \xi,x\rangle \geq M\}} \exp[\langle \xi, x\rangle/\epsilon]\nu_\epsilon(dx) = -\infty,$$

and similarly for ξ *replaced by* $-\xi$*,*

A new family of probability measures $\{\lambda_\epsilon^\xi\}$ is defined by

$$\frac{d\lambda_\epsilon^\xi}{d\nu_\epsilon}(x) := \exp\left(\langle \xi, x\rangle/\epsilon - \log \int_{\mathcal{X}} \exp[\langle \xi, y\rangle/\epsilon]\nu_\epsilon(dy)\right) = \frac{\exp[\langle \xi, x\rangle/\epsilon]}{\int_{\mathcal{X}} \exp[\langle \xi, y\rangle/\epsilon]\nu_\epsilon(dy)}. \tag{3.1}$$

The measures $\{\lambda_\epsilon^\xi\}$ are called *exponentially twisted with twist* ξ. If the family $\{\lambda_\epsilon^\xi\}$ is asymptotically efficient, we simply call the exponential twist ξ asymptotically efficient.

The following proposition plays a key role in the proofs of this section.

Proposition 1 *Let* $d\lambda_\epsilon^\xi/d\nu_\epsilon$ *be given by (3.1), and let* $B \in \mathcal{B}$*. Under Assumption 1, we have*

$$\begin{aligned}
\liminf_{\epsilon\to 0} \epsilon \log \int_B \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}\right)^2 d\lambda_\epsilon^\xi &\geq -\inf_{x\in\mathcal{X}}[I(x) - \langle \xi, x\rangle] - \inf_{x\in B^o}[I(x) + \langle \xi, x\rangle] \\
\limsup_{\epsilon\to 0} \epsilon \log \int_B \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}\right)^2 d\lambda_\epsilon^\xi &\leq -\inf_{x\in\mathcal{X}}[I(x) - \langle \xi, x\rangle] - \inf_{x\in \overline{B}}[I(x) + \langle \xi, x\rangle].
\end{aligned}$$

Proof. Fix $B \in \mathcal{B}$ and note that

$$\begin{aligned}
\epsilon \log \int_B \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}\right)^2 d\lambda_\epsilon^\xi &= \epsilon \log \int_B \frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi} d\nu_\epsilon \\
&= \epsilon \log \int_B \exp\left(\log \int_{\mathcal{X}} \exp(\langle \xi, y\rangle/\epsilon)\nu_\epsilon(dy) - \langle \xi, x\rangle/\epsilon\right) \nu_\epsilon(dx) \\
&= \epsilon \log \int_{\mathcal{X}} \exp\left(\langle \xi, x\rangle/\epsilon\right) \nu_\epsilon(dx) + \epsilon \log \int_B \exp\left(-\langle \xi, x\rangle/\epsilon\right) \nu_\epsilon(dx). \qquad (3.2)
\end{aligned}$$

By Assumption 1 and the continuity of the functional ξ, Varadhan's Integral Lemma [Theorem 4.3.1 in Dembo and Zeitouni (1998)] applies. Thus, the limit of the first term exists and equals

$$\lim_{\epsilon \to 0} \epsilon \log \int_{\mathcal{X}} \exp\left(\langle \xi, x\rangle/\epsilon\right) \nu_\epsilon(dx) = \sup_{x \in \mathcal{X}} [\langle \xi, x\rangle - I(x)].$$

A similar argument can be applied to the second term in (3.2). The conditions of Varadhan's Integral Lemma are again satisfied to apply the lemma to the continuous functional $-\xi$. Now we use a variant of this lemma (see, e.g., Exercise 2.1.24 in Deuschel and Stroock (1989)) to see that for any open set G and any closed set F

$$\begin{aligned}
\liminf_{\epsilon \to 0} \epsilon \log \int_G \exp\left(-\langle \xi, x\rangle/\epsilon\right) \nu_\epsilon(dx) &\geq -\inf_{x \in G} [I(x) + \langle \xi, x\rangle], \\
\limsup_{\epsilon \to 0} \epsilon \log \int_F \exp\left(-\langle \xi, x\rangle/\epsilon\right) \nu_\epsilon(dx) &\leq -\inf_{x \in F} [I(x) + \langle \xi, x\rangle].
\end{aligned}$$

In particular, these inequalities hold for B^o and $\overline{B}$. The claim follows by adding the two terms in (3.2), which is allowed since the limit of the first term exists [see the reasoning following Inequality (2.5) and Section 2.4 of Royden (1968)]. □

The necessary and sufficient conditions, formulated in the next theorem, follow almost immediately from Proposition 1.

Theorem 1 *Let Assumption 1 hold. The exponential twist ξ is asymptotically efficient if*

$$\inf_{x \in \mathcal{X}} [I(x) - \langle \xi, x\rangle] + \inf_{x \in \overline{A}} [I(x) + \langle \xi, x\rangle] \geq 2 \inf_{x \in A^o} I(x). \qquad (3.3)$$

Let Assumption 1 hold and let A be an I-continuity set. If the exponential twist ξ is asymptotically efficient, then

$$\inf_{x \in \mathcal{X}} [I(x) - \langle \xi, x\rangle] + \inf_{x \in A^o} [I(x) + \langle \xi, x\rangle] \geq 2 \inf_{x \in \overline{A}} I(x). \qquad (3.4)$$

Proof. Sufficiency follows from (2.5), the upper bound of Proposition 1, and the LDP of Assumption 1(ii):

$$\begin{aligned}
\limsup_{\epsilon \to 0} \epsilon \log N^*_{\lambda_\epsilon^\xi} &\leq \limsup_{\epsilon \to 0} \epsilon \log \int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}\right)^2 d\lambda_\epsilon^\xi - 2 \liminf_{\epsilon \to 0} \epsilon \log \nu_\epsilon(A) \\
&\leq -\inf_{x \in \mathcal{X}} [I(x) - \langle \xi, x\rangle] - \inf_{x \in \overline{A}} [I(x) + \langle \xi, x\rangle] + 2 \inf_{x \in A^o} I(x).
\end{aligned}$$

For necessity the argument is similar. Note that the lower bound of Proposition 1 implies that

$$\limsup_{\epsilon\to 0} \epsilon \log \int_A \left(\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}\right)^2 d\lambda_\epsilon^\xi \geq -\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle] - \inf_{x\in A^o}[I(x)+\langle\xi,x\rangle].$$

Moreover, by the large deviation upper bound, $\liminf_{\epsilon\to 0}\epsilon\log\nu_\epsilon(A) \leq -\inf_{x\in\overline{A}} I(x)$. Combining these observations with the assumption that A is an I-continuity set, we have

$$\begin{aligned} 0 &= \limsup_{\epsilon\to 0}\epsilon\log N^*_{\lambda_\epsilon^\xi} \\ &= \limsup_{\epsilon\to 0}\epsilon\log\int_A\left(\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}\right)^2 d\lambda_\epsilon^\xi - 2\lim_{\epsilon\to 0}\epsilon\log\nu_\epsilon(A) \\ &\geq -\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle] - \inf_{x\in A^o}[I(x)+\langle\xi,x\rangle] + 2\inf_{x\in\overline{A}} I(x), \end{aligned}$$

as desired. □

As suggested by the form of Theorem 1, the sufficient condition is also necessary under a weak condition on the set A. We formalize this in the following corollary.

Corollary 1 *Let Assumption 1 hold, and assume that A is both an I-continuity set and an $(I+\xi)$-continuity set. Exponentially twisting with ξ is asymptotically efficient if and only if*

$$\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle] + \inf_{x\in\overline{A}}[I(x)+\langle\xi,x\rangle] = 2\inf_{x\in\overline{A}} I(x).$$

Proof. From Theorem 1 it is obvious that

$$\inf_{x\in\mathcal{X}}[I(x)-\langle\xi,x\rangle] + \inf_{x\in\overline{A}}[I(x)+\langle\xi,x\rangle] \geq 2\inf_{x\in\overline{A}} I(x)$$

is necessary and sufficient. The reverse inequality holds trivially. □

Remark. Sadowsky (1996) uses a more general notion than asymptotic efficiency, namely ν-efficiency. Given an I-continuity set A, the importance sampling distribution λ_ϵ^ξ is said to be ν-efficient if

$$\limsup_{\epsilon\to 0}\epsilon\log\int_A\left(\frac{d\nu_\epsilon}{d\lambda_\epsilon^\xi}\right)^\nu d\lambda_\epsilon \leq -\nu\inf_{x\in\overline{A}} I(x).$$

In this terminology, we have established conditions for 2-efficiency (see the remarks after Definition 3). To obtain conditions for ν-efficiency with general $\nu\geq 2$, the statements in the subsection are easily modified. As an example, when A is an $(I+(\nu-1)\xi)$-continuity set and when Assumption 1(iii) holds with ξ replaced by $(\nu-1)\xi$ and $-(\nu-1)\xi$, the exponential twist ξ is ν-efficient if and only if

$$\inf_{x\in\mathcal{X}}[I(x)-(\nu-1)\langle\xi,x\rangle] + \inf_{x\in\overline{A}}[I(x)+(\nu-1)\langle\xi,x\rangle] = \nu\inf_{x\in\overline{A}} I(x).$$

□

3.2 Relation with Sadowsky's conditions

General necessary and sufficient conditions for asymptotic efficiency were also developed by Sadowsky (1996). In this subsection, we compare the conditions of Theorem 1 with Sadowsky's conditions. In the course of the exposition, it becomes clear that the underlying assumptions of Subsection 3.1 are less restrictive than Sadowsky's assumptions. Moreover, we show that our sufficient condition in Theorem 1 is weaker than Sadowsky's sufficiency condition, and that the accompanying necessary condition is stronger than his necessary condition. Thus, the results of the previous subsection improve Sadowsky's conditions.

In addition to the notation of the preceding subsection, we first introduce some new notions. In this subsection, $\mathcal{X}$ denotes a topological vector space, and $\mathcal{X}^*$ denotes the space of linear continuous functionals $\xi : \mathcal{X} \to \mathbb{R}$. Let $f : \mathcal{X} \to (-\infty, \infty]$ be a convex function. A point $x \in \mathcal{X}$ is called an *exposed point* of f if there exists a $\delta \in \mathcal{X}^*$ such that $f(y) > f(x) + \langle \delta, y - x \rangle$ for all $y \neq x$. δ is then called an *exposing hyperplane* of I at x.

To compare our results to Sadowsky's, we first recapitulate Sadowsky's conditions. The following set of assumptions will be referred to as Sadowsky's assumptions.

Assumption 2 (Sadowsky) *Assume that*

(i) $\mathcal{X}$ is a locally convex Hausdorff topological vector space,

(ii) $\{\nu_\epsilon\}$ satisfies the LDP with a convex *good rate function I,*

(iii) for every $\delta \in \mathcal{X}^*$,

$$\Lambda(\delta) := \limsup_{\epsilon \to 0} \epsilon \log \int_{\mathcal{X}} \exp[\langle \delta, x \rangle / \epsilon] \nu_\epsilon(dx) < \infty,$$

(iv) A satisfies

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

where $\mathcal{F}$ denotes the set of exposed points of I.

We say that Assumption 2' is satisfied if Assumption 2(i), Assumption 2(ii), and Assumption 2(iii) hold, and if A is an I-continuity set in the sense of Definition 1 with $0 < \inf_{x \in A} I(x) < \infty$.

Although Assumption 2 looks very similar to Assumption 1, there are crucial differences.

To start with, $\mathcal{X}$ is not assumed to be a *topological* vector space in Assumption 1(i). To see the importance of this difference for applications, note that the space $D([0,1], \mathbb{R})$ of càdlàg functions on $[0,1]$ with values in $\mathbb{R}$ is a (regular, Hausdorff) vector space but no topological vector space when equipped with the Skorohod topology. Another difference is that the regularity of $\mathcal{X}$ assumed in Assumption 1(i) seems not be present in Assumption 2(i). However, this regularity is implicit: any real Hausdorff topological vector space is automatically regular.

Moreover, the convexity of the large deviation rate function is not assumed in Assumption 1(ii). Note that this convexity is granted when an LDP is derived using an (abstract)

Gärtner-Ellis type theorem, but that it can be lost by an application of the Contraction Principle.

At first sight, there is no clear relation between Assumption 1(iii) and Assumption 2(iii). However, Lemma 4.3.8 of Dembo and Zeitouni (1998) states that Assumption 2(iii) implies Assumption 1(iii), since $\gamma\xi$ and $-\gamma\xi$ are continuous linear functionals, e.g., for $\gamma = 2$. We remark that the relevant limit of Assumption 2(iii) exists by Theorem 4.5.10(a) of Dembo and Zeitouni (1998).

The fourth part of Assumption 2 is closely related to requiring that A be an I-continuity set. In fact, it is a stronger assumption than I-continuity of A. Given that Assumption 2(iv) holds for A, $\gamma \in \overline{A}$ is called a *point of continuity* if $I(\gamma) = \inf_{x\in\overline{A}} I(x)$ and there exists a sequence $\{\gamma_n\} \subset A^o \cap \mathcal{F}$ such that $\gamma_n \to \gamma$. We now show that there always exists a point of continuity for a set satisfying Assumption 2(iv) in case I is a good rate function. First note that $\inf_{x\in A^o\cap\mathcal{F}} I(x) = \inf_{x\in\overline{A}} I(x)$ implies that for any $n \in \mathbb{N}$ we can find a $\gamma_n \in A^o \cap \mathcal{F} \cap K_n$, where

$$K_n := \{x \in \mathcal{X} : I(x) \leq \inf_{y\in\overline{A}} I(y) + 1/n\}.$$

Use $\inf_{x\in\overline{A}} I(x) < \infty$ and the goodness of the rate function I to see that K_n is a compact subset of $\mathcal{X}$. Since K_n decreases in n, we obviously have $\{\gamma_n\} \subset K_1$. Hence, one can substract a subsequence that converges, say, to $\gamma^* \in K_1$. Since K_n is closed for every n and $\{\gamma_\cdot\}$ is eventually in K_n, we must also have that $\gamma^* \in K_n$ for every n. As a consequence, we have $I(\gamma^*) \leq \inf_{x\in\overline{A}} I(x)$. Moreover, since $\{\gamma_n\} \subset A^o \cap \mathcal{F}$, we also see that $\gamma^* \in \overline{A^o \cap \mathcal{F}} \subset \overline{A}$. Therefore, $I(\gamma^*) = \inf_{x\in A^o\cap\mathcal{F}} I(x) = \inf_{x\in\overline{A}} I(x)$, and γ^* is a point of continuity.

In the above comparison between Assumption 1 and Assumption 2, we have shown the following.

Proposition 2 *Assumption 2 implies that Assumption 1 holds and that A is an I-continuity set.*

In the remainder of this subsection, we compare the necessary and sufficient conditions of Theorem 1 to the conditions in Sadowsky (1996). In order to do this, we have to make sure that Sadowsky's conditions hold, i.e., we assume the stronger Assumption 2. We start by repeating Sadowsky's conditions.

Theorem 2 (Sadowsky) *Let Assumption 2 hold. The exponential twist ξ is asymptotically efficient if*

- *(a) there is a point of continuity γ such that $I(\gamma) = \langle\xi, \gamma\rangle - \Lambda(\xi)$,*
- *(b) $I(x) + \langle\xi, x\rangle \geq I(\gamma) + \langle\xi, \gamma\rangle$ for all $x \in \overline{A}$,*
- *(c) either $\langle\xi, x\rangle \geq \langle\xi, \gamma\rangle$ for all $x \in \overline{A}$, or there exists an $x \in \mathcal{F}$ such that x is an exposed point of I.*

Let Assumption 2 hold. If the twist ξ is asymptotically efficient, then

- *(a) there is a point of continuity γ such that $I(\gamma) = \langle\xi, \gamma\rangle - \Lambda(\xi)$,*

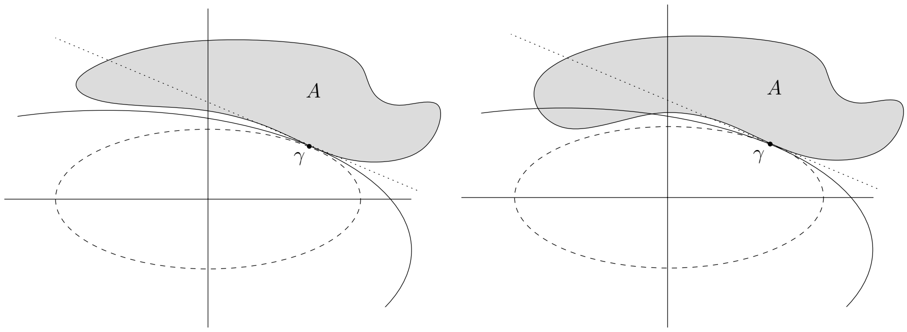

Figure 1: Efficient simulation with twist ξ_γ (left) and inefficient simulation with twist ξ_γ (right).

(b') $I(x) + \langle \xi, x\rangle \geq I(\gamma) + \langle \xi, \gamma\rangle$ for all $x \in A^o \cap \mathcal{F}$.

If $\langle \xi, x\rangle \geq \langle \xi, \gamma\rangle$ for all $x \in \overline{A}$ in part (c) of the sufficient condition, γ is called a *dominating point.*

Proposition 3 *Let Assumption 2 hold. The sufficient condition in Theorem 2 implies the sufficient condition in Theorem 1.*

Proof. By condition (a) of Theorem 2, there exists a point of continuity $\gamma \in \overline{A}$ such that $I(\gamma) = \inf_{x\in\overline{A}} I(x) = \langle \xi, \gamma\rangle - \Lambda(\xi)$. Since we assume that an LDP holds for some convex I [Assumption 2(ii)] and that Assumption 2(iii) holds, by Theorem 4.5.10(b) in Dembo and Zeitouni (1998) we have $I(x) = \sup_{\delta\in\mathcal{X}^*}[\langle \delta, x\rangle - \Lambda(\delta)]$, and hence $I(x) \geq \langle \xi, x\rangle - \Lambda(\xi)$. Combining this with $I(\gamma) = \langle \xi, \gamma\rangle - \Lambda(\xi)$, we conclude

$$\inf_{x\in\mathcal{X}}[I(x) - \langle \xi, x\rangle] \geq -\Lambda(\xi) = I(\gamma) - \langle \xi, \gamma\rangle,$$

where the inequality may obviously be replaced by an equality.

It is immediate from condition (b) of Theorem 2 that $\inf_{x\in\overline{A}}[I(x) + \langle \xi, x\rangle] = I(\gamma) + \langle \xi, \gamma\rangle$. Since $\inf_{x\in A^o} I(x) = I(\gamma)$, this implies the sufficient condition (3.3) in Theorem 1. □

Remark. It is important to notice that we did not use part (c) of Sadowsky's sufficient condition in the proof of Proposition 3. Hence this part is redundant. In particular, the notion of dominating points plays no role. Still, there is a relation between dominating points and asymptotic efficiency. We illustrate this with a two-dimensional example depicted in Figure 1.

Let ν be the distribution of a random variable X on $\mathbb{R}^d$, and denote the distribution of the sample mean of n i.i.d. copies of X by ν_n. Note that $1/n$ plays the role of ϵ in this example. Let ν be such that Cramér's theorem holds. (In Figure 1, ν_n is a zero-mean bivariate Gaussian distribution with covariance of the form Σ/n for a diagonal matrix Σ). In the left and right panel, two different sets A are drawn. We are interested in $\nu_n(A)$.

As indicated by the dashed level curve of I, the 'most likely point' in A is in both cases γ, i.e., $\arg\inf_{x\in A} I(x) = I(\gamma)$. In the next subsection, we see that there is only one exponential twist $\xi_\gamma \in \mathbb{R}^2$ that is interesting for simulation purposes. The level curve of

$I + \xi_\gamma + \inf_{x\in\mathcal{X}}[I(x) - \xi_\gamma^t x]$ that goes through γ is depicted as a solid line. Since both sets A are I- and $(I+\xi_\gamma)$-continuity sets, the twist ξ_γ is asymptotically efficient if and only if A lies entirely 'outside' the solid level curve (see Corollary 1). Hence, in the left panel the twist ξ_γ is asymptotically efficient twist and in the right panel it is not.

We now turn to dominating points. Every I-continuity set that touches γ and that is contained in the halfspace above the dotted line has dominating point γ. Obviously, such a set lies outside the solid level curve, and one can therefore estimate $\nu_n(A)$ asymptotically efficiently by an exponential twist. Notice that convexity of A implies that A has a dominating point, so that asymptotic efficiency can be achieved. However, if A is neither convex nor has a dominating point, Figure 1 indicates that it may still be possible to simulate asymptotically efficiently. $\square$

Proposition 4 *Let Assumption 2 hold. The necessary condition in Theorem 2 is implied by the necessary condition in Theorem 1.*

Proof. Since the rate function I is good [Assumption 2(ii)] and A is an I-continuity set in the sense of Assumption 2(iv), there exists a point of continuity $\gamma \in \overline{A}$ such that $\inf_{x\in\overline{A}} I(x) = I(\gamma)$, and a sequence $\{\gamma_n\}$ in $A^o \cap \mathcal{F}$ with $I(\gamma_n) \to I(\gamma)$ and $\gamma_n \to \gamma$ (cf. the reasoning above Proposition 2). By arguing along subsequences if necessary, we may assume that $\gamma_n \to \gamma$ without loss of generality. The necessary condition in Theorem 1 implies

$$\begin{aligned} 2I(\gamma) &\leq \inf_{x\in A^o}[I(x) + \langle\xi,x\rangle] - \sup_{x\in\mathcal{X}}[\langle\xi,x\rangle - I(x)] \\ &\leq \lim_{n\to\infty}[I(\gamma_n) + \xi(\gamma_n)] - [\langle\xi,\gamma\rangle - I(\gamma)] = 2I(\gamma). \end{aligned}$$

As a result, the inequalities can be replaced by equalities and we obtain

$$\sup_{x\in\mathcal{X}}[\langle\xi,x\rangle - I(x)] = \langle\xi,\gamma\rangle - I(\gamma) \text{ and } \inf_{x\in A^o}[I(x) + \langle\xi,x\rangle] = I(\gamma) + \langle\xi,\gamma\rangle.$$

By Theorem 4.5.10(a) of Dembo and Zeitouni (1998), we also have $\sup_{x\in\mathcal{X}}[\langle\xi,x\rangle - I(x)] = \Lambda(\xi)$ under Assumption 2. Hence, $I(\gamma) = \langle\xi,\gamma\rangle - \Lambda(\xi)$ and part (a) of Sadowsky's necessary condition is derived. Part (b') is immediate by noting that $\inf_{x\in A^o}[I(x)+\langle\xi,x\rangle] = I(\gamma)+\langle\xi,\gamma\rangle$ implies that $I(x) + \langle\xi,x\rangle \leq I(\gamma) + \langle\xi,\gamma\rangle$ for all $x \in A^o$. $\square$

3.3 Convexity considerations

In this subsection, we study the *uniqueness* of an asymptotically efficient exponential twist.

We restrict ourselves to the case that the large deviation rate function has specific convexity properties, motivated by the case that the LDP is established by using an (abstract) Gärtner-Ellis type theorem. Therefore, we make the same assumptions as Sadowsky (1996).

However, exposed points play no role in the analysis, so we slightly modify Sadowsky's framework to make it more easy. As a consequence, Assumption 2' is used instead of Assumption 2. Moreover, we also adapt the definition of a point of continuity that we used in the previous subsection. In this subsection, γ is called a *point' of continuity* of the I-continuity set A if $I(\gamma) = \inf_{x\in\overline{A}} I(x) < \infty$ and there exists a sequence $\{\gamma_n\} \subset A^o$ such that $\gamma_n \to \gamma$.

We already saw that Theorem 4.5.10 of Dembo and Zeitouni (1998) holds under Assumption 2', i.e., that

$$\Lambda(\xi) = \sup_{x\in\mathcal{X}}[\langle\xi,x\rangle - I(x)] \text{ and } I(x) = \sup_{\xi\in\mathcal{X}^*}[\langle\xi,x\rangle - \Lambda(\xi)]. \tag{3.5}$$

Theorem 3 *Let Assumption 2' hold, and let both* Λ *and* I *be* strictly *convex on their domains. If there exist two (different) points' of continuity, then there is no asymptotically efficient exponential twist.*

Proof. Let γ be a point' of continuity, and let $\{\gamma_n\} \subset A^o$ satisfy $\gamma_n \to \gamma$ and $I(\gamma_n) \to I(\gamma)$. Due to the strict convexity of Λ and by the second identity in (3.5), there exists *at most* one $\xi_\gamma \in \mathcal{X}^*$ such that $I(\gamma) = \langle \xi_\gamma, \gamma\rangle - \Lambda(\xi_\gamma)$. For all other twists $\xi \in \mathcal{X}^*$, we apparently have $I(\gamma) > \langle \xi, \gamma\rangle - \Lambda(\xi)$, and therefore

$$\inf_{x\in\mathcal{X}}[I(x) - \langle \xi, x\rangle] = -\sup_{x\in\mathcal{X}}[\langle \xi, x\rangle - I(x)] = -\Lambda(\xi) < I(\gamma) - \langle \xi, \gamma\rangle.$$

Consequently, $\xi \neq \xi_\gamma$ cannot be asymptotically efficient, as

$$\begin{aligned}\inf_{x\in\mathcal{X}}[I(x) - \langle \xi, x\rangle] + \inf_{x\in A^o}[I(x) + \langle \xi, x\rangle] &< I(\gamma) - \langle \xi, \gamma\rangle + \lim_{n\to\infty}[I(\gamma_n) + \langle \xi, \gamma_n\rangle] \\ &= I(\gamma) - \langle \xi, \gamma\rangle + I(\gamma) + \langle \xi, \gamma\rangle = 2I(\gamma),\end{aligned}$$

contradicting the necessary condition in Theorem 1. If ξ_γ does not exist, then the claim is proven. Hence, we assume its existence; i.e., the supremum over $\mathcal{X}^*$ in (3.5) is attained for $x = \gamma$. This leaves ξ_γ as the only candidate for asymptotic efficiency.

The same argument applies to a different point' of continuity, say, $\eta \neq \gamma$. This point also gives us a candidate twist (existence may again be assumed). The two twists are denoted by ξ_γ and ξ_η. The claim follows after showing that $\xi_\gamma \neq \xi_\eta$.

Suppose $\xi_\gamma = \xi_\eta$. Since $\langle \xi_\gamma, \gamma\rangle - \Lambda(\xi_\gamma) = I(\gamma) = I(\eta) = \langle \xi_\eta, \eta\rangle - \Lambda(\xi_\eta)$, we have for $0 \le \alpha \le 1$:

$$I(\gamma) = \alpha[\langle \xi_\gamma, \gamma\rangle - \Lambda(\xi_\gamma)] + (1-\alpha)[\langle \xi_\eta, \eta\rangle - \Lambda(\xi_\eta)] = \langle \xi_\gamma, \alpha\gamma + (1-\alpha)\eta\rangle - \Lambda(\xi_\gamma).$$

From this it follows that $\Lambda(\xi_\gamma) = \langle \xi_\gamma, \alpha\gamma + (1-\alpha)\eta\rangle - I(\gamma)$. This equality obviously holds for $\alpha = 1$, so we obtain a contradiction with the first identity in (3.5) and the strict convexity of I. □

The arguments in the above proof yield the following corollary.

Corollary 2 *Let Assumption 2' hold, and let* Λ *be* strictly *convex on its domain. There exists at most one asymptotically efficient twist.*

Remark. Under the assumptions of Theorem 3 and in the notation of its proof, the exponential twist ξ_γ is asymptotically efficient if $\inf_{x\in\overline{A}}[I(x) + \langle \xi_\gamma, x\rangle] = I(\gamma) + \langle \xi_\gamma, \gamma\rangle$; condition (3.3) is thus neatly rewritten. The necessary condition (3.4) can be rephrased in a similar way. □

Evidently, the twist of Corollary 2 is the best possible choice for simulation *if it is asymptotically efficient.* However, this this is not *always* the best choice. Using a simple example, we discuss more counterintuitive observations in the next subsection.

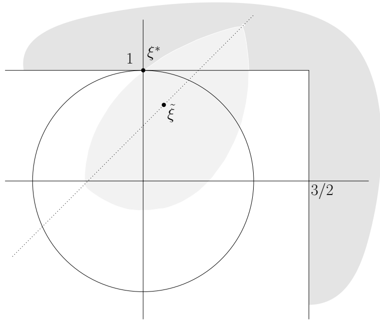

Figure 2: Illustration of the example in Subsection 3.4

3.4 An example

In this subsection, we discuss the difficulties that arise when the candidate twist of Corollary 2, hereafter also called the large deviation twist, is not asymptotically efficient. We construct a simple example to illustrate these issues, and we address the relation with other examples given by Glasserman and Wang (1997) and Dupuis and Wang (2002).

Before introducing the example, we first describe a criterion for discriminating between exponential twists. Building upon the idea behind the definition of asymptotic efficiency, it is sensible to use $C_\xi := \limsup_{\epsilon \to 0} \epsilon \log N^*_{\lambda^\xi_\epsilon}$, where $N^*_{\lambda^\xi_\epsilon}$ is given in (2.4). Obviously, an exponential twist ξ_1 outperforms ξ_2 if $C_{\xi_1} < C_{\xi_2}$. In other words, the number of samples required to obtain a fixed relative error should have the lowest possible exponential rate. Note that Equation (2.5) shows that, under specific assumptions on A, it is equivalent to minimize the decay of the second moment of the estimator.

We construct an example in the well-studied Cramér framework to illustrate some problems with this criterion when no asymptotically efficient twist exists. Consider an i.i.d. sequence $X_1, X_2, \ldots$ of two-dimensional zero-mean Gaussian vectors with covariance I, i.e., X_i^j are standard normal and X_i^1 is independent of X_i^2. The distribution of $\frac{1}{n}\sum_{i=1}^n X_i$ is denoted by ν_n. By Cramér's theorem (e.g., Theorem 2.2.30 of Dembo and Zeitouni (1998)), $\{\nu_n\}$ satisfies the LDP with good rate function

$$I(x) = \frac{1}{2}\|x\|^2,$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^2$. Let $A := \{(x,y) : x \geq 3/2, y \geq 1\}$, and consider the simulation of $\nu_n(A)$. This example is illustrated in Figure 2.

It is readily checked that A is both an I-continuity set and an $(I+\xi)$-continuity set for any $\xi \in \mathbb{R}^2$. One can also easily see that

$$\begin{aligned} C_\xi &= -\inf_{x\in\mathbb{R}^2}\left[\frac{1}{2}\|x\|^2 - \xi^t x\right] - \inf_{x\in A}\left[\frac{1}{2}\|x\|^2 + \xi^t x\right] + \inf_{x\in A}\|x\|^2 \\ &= \frac{1}{2}\|\xi\|^2 - \inf_{x\in A}\left[\frac{1}{2}\|x\|^2 + \xi^t x\right] + \inf_{x\in A}\|x\|^2 \\ &= -\frac{1}{2}\inf_{x\in A}\|x+\xi\|^2 + \inf_{x\in A}\|x\|^2 + \|\xi\|^2. \end{aligned}$$

The most probable point in A is $\arg\inf_{x\in A}\|x\|^2 = (0,1)$, and one can check that the large deviation twist reads $\xi^* = (0, 1)$. However, this twist is not asymptotically efficient: $C_{\xi^*} = 7/8$. Even direct Monte Carlo, corresponding to the zero-twist, shows better performance than the large deviation twist: $C_{(0,0)} = 1/2 < 7/8 = C_{(1,0)}$! This possibility was first observed by Glasserman and Wang (1997).

In case the large deviation twist is asymptotically inefficient, Dupuis and Wang (2002) do some simulation experiments to show that simulation with this twist may give unstable estimates. Since the large deviation twist can apparently theoretically be outperformed (in terms of C_ξ), it is natural to study the stability of other (better) twists. In order to compute the C_ξ-minimizing twist in our example, one can readily see that for $-\xi \notin A$,

$$\inf_{x \in A} \|x + \xi\|^2 = \begin{cases} (\xi_1 + 3/2)^2 & \xi_2 > 1/2 + \xi_1; \\ (\xi_2 + 1)^2 & \text{otherwise.} \end{cases}$$

Straightforward calculations show that the best twist is $\tilde{\xi} = (1/6, 2/3)$, with $C_{\tilde{\xi}} = 1/12$.

It is also interesting to compute the twists that perform better than the large deviation twist ξ^*, i.e., $\{\xi : C_\xi < 7/8\}$. Equivalently, we compute for which ξ it holds that

$$\frac{1}{2} \inf_{x \in A} \|x + \xi\|^2 - \|\xi\|^2 > 1/8.$$

This set is also drawn in Figure 2. Notice that $\tilde{\xi}$ has a 'central' position in the set.

Although the twist $\tilde{\xi}$ is theoretically optimal, simulation from the twisted measure faces exactly the same difficulties as direct Monte Carlo. Of course, this is caused by the fact that the mean of the $\tilde{\xi}$-twisted measure is no element of A. To illustrate the problem, we perform a small simulation experiment. We let $n = 50$ and perform $1\,000\,000\,000$ simulation runs under the measure induced by the (theoretically best) twist $\tilde{\xi}$. (The period of the random number generator is large enough.) Note that the probability of interest can be bounded by $\nu_n(A) \le \overline{\Phi}(\sqrt{n}) + \overline{\Phi}(3\sqrt{n}/2) \approx 7.69\, 10^{-13}$, where $\overline{\Phi}$ denotes the complement of the standard normal distribution function. Not surprisingly, we do not hit A in any of the simulation runs, leaving 0 as the resulting estimate.

This example shows that it should *always* be checked first whether there exists an asymptotically efficient twist, otherwise the resulting estimate may be highly unreliable. If there is no such twist, alternative simulation methods should be considered. In Section 5, we discuss some of these methods.

4. Application: Mogul′skiĭ sample-path probabilities

In this section, we apply the results of Section 3 to the simulation problem of a time-varying level-crossing probability. The underlying large deviations are governed by an LDP that is, following Dembo and Zeitouni (1998), usually referred to as Mogul′skiĭ's theorem.

Sadowsky (1996) was the first to consider this problem in the form that we do. In fact, he showed that Assumption 2 holds for this problem, and used Theorem 2 to study asymptotically efficient simulation. However, the transparency of our conditions in Theorem 1 and the ideas used in the proofs make it possible to gain new insight into this example. To be more specific, we obtain necessary and sufficient conditions for asymptotic efficiency of the large deviation exponential twist in Subsection 4.2. By giving a counterexample, we correct Sadowsky's claim that this exponential twist is *never* asymptotically optimal. Moreover, in

Subsection 4.3, we formulate necessary and sufficient conditions for asymptotic efficiency of an appealing estimator based on the best exponential twist, to which Sadowsky refers as the *sequential estimator*. An important application of the results of this section is the simulation of ruin probabilities with finite time horizon.

We start with some preliminaries.

4.1 Preliminaries

Notation Let $Z_1, Z_2, \ldots$ be a sequence of i.i.d. zero-mean random variables taking values in $\mathbb{R}$, with distribution P_Z. The logarithmic moment generating function of P_Z is given by $\Lambda_Z(\theta) := \log \mathbb{E}(e^{\theta Z_1})$ for $\theta \in \mathbb{R}$. Its Fenchel-Legendre transform is defined as $\Lambda_Z^*(z) := \sup_{\xi\in\mathbb{R}}[\xi z - \Lambda_Z(\xi)]$. We set $\operatorname{dom} \Lambda_Z^* := \{z \in \mathbb{R} : \Lambda_Z^*(z) < \infty\}$.

For $0 \leq t \leq 1$, let the scaled polygonal approximation for the partial sums of Z_i be given by

$$S_n(t) := \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor} Z_i + \left(t - \frac{\lfloor nt\rfloor}{n}\right) Z_{\lfloor nt\rfloor+1}, \tag{4.1}$$

where $\lfloor\cdot\rfloor$ denotes the largest integer that is smaller than $\cdot$. Take $\mathcal{X}$ to be the Banach space of continuous functions $x : [0,1] \to \mathbb{R}$ with $x(0) = 0$, endowed with the sup norm topology. We use the interval $[0,1]$ only for simplicity; the reasoning in this section applies to any interval of the form $[0,T]$, $T > 0$. The distribution of $S_n(\cdot)$ in $\mathcal{X}$ is denoted by ν_n. Note that we replace ϵ by $1/n$ throughout this section. As in Section 3.2, we define for $\xi \in \mathcal{X}^*$,

$$\Lambda(\xi) := \lim_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} \exp(n\langle\xi, x\rangle)\nu_n(dx).$$

The space of absolutely continuous functions $\mathcal{AC}$ on $[0,1]$ plays an important role in the sequel. It is defined as

$$\mathcal{AC} := \left\{ x : \sum_{\ell=1}^{k} |t_\ell - s_\ell| \to 0, s_\ell < t_\ell \leq s_{\ell+1} < t_{\ell+1} \Longrightarrow \sum_{\ell=1}^{k} |x(t_\ell) - x(s_\ell)| \to 0 \right\}.$$

As in Sadowsky (1996), we consider the estimation of a time-varying level-crossing probability. That is, we are interested in estimating $\nu_n(A)$ efficiently, where

$$A := \{x \in \mathcal{X} : x(t) \geq e(t) \text{ for some } t \in [0,1]\},$$

for some lower semicontinuous (with respect to the relative Euclidean topology on $[0,1]$) $e : [0,1] \to (0,\infty]$.

It is possible to consider the (more general) problem with non-centered random variables Z and $e : [0,1] \to (-\infty,\infty]$ such that $e(t)/t > \mathbb{E}Z_1$. However, since a simple transformation converts this problem to our framework, we restrict ourselves without loss of generality to the simpler setup.

In the special case that e has the form $e(t) = a + bt$ for some $a, b \geq 0$, $\nu(A)$ corresponds to a ruin probability for the finite horizon case. Simulation of this probability is studied by Lehtonen and Nyrhinen (1992b), whereas Asmussen (2000, X.4) and Asmussen et al. (2002) consider the analog in continuous time; then, a Lévy process replaces the random walk.

Proposition 7 below generalizes the findings in Lehtonen and Nyrhinen (1992b) to the case of a time-varying level crossing probability.

Assumptions To be able to apply the uniqueness results of Subsection 3.3, we have to make sure that Assumption 2' and some additional strict convexity assumptions hold. To this end, we need the following problem-specific assumptions:

Assumption 3 *We assume that*

(i) $\Lambda_Z(\theta) < \infty$ *for all* $\theta \in \mathbb{R}$,

(ii) P_Z *is non-degenerate,*

(iii) Λ_Z^* *is strictly convex,*

(iv) $0 < \inf_{t\in[0,1]} e(t) < \infty$.

For Assumption 2(ii) to hold, we need an LDP for the family $\{\nu_n\}$. In Section 5.1 of Dembo and Zeitouni (1998), it is shown that $\{\nu_n\}$ satisfies the LDP under Assumption 3(i) with the convex good rate function

$$I(x) := \begin{cases} \int_0^1 \Lambda_Z^*(\dot{x}(t))dt & \text{if } x \in \mathcal{AC}; \\ \infty & \text{otherwise.} \end{cases}$$

This result is referred to as *Mogul′skiĭ's theorem*.

Note that Assumption 3(i) can be considerably relaxed for a Mogul′skiĭ-type LDP to hold. One then uses different spaces equipped with different topologies, which are less convenient than $\mathcal{X}$ with the topology of uniform convergence. Mogul′skiĭ (1976) allows the logarithmic moment generating function to be finite only in a neighborhood of zero and uses the space of càdlàg functions D endowed with the (completed) Skorohod topology; see also Mogul′skiĭ (1993). Dembo and Zajic (1995) and de Acosta (1994) work under the hypothesis of a finite logarithmic generating function of $|Z|$.

In the present paper, however, we use Assumption 3 to ease the exposition. In fact, Assumption 3 was introduced to apply the results of Section 3. This is highlighted by following proposition.

Proposition 5 *If* $e(\tilde{\tau})/\tilde{\tau} \in (\mathrm{dom}\,\Lambda_Z^*)^o$ *for some* $\tilde{\tau} \in \arg\inf_{\tau\in[0,1]} \tau\Lambda_Z^*(e(\tau)/\tau)$, *then Assumption 3 implies Assumption 2'. Moreover, both the rate function* I *and the logarithmic generating function* Λ *are strictly convex on their domains.*

Before proving this proposition, we need some auxiliary lemmas. We start by showing that A is closed. To this end, consider

$$A_m^M := \{x \in \mathcal{X} : x(t) \geq M(t) \text{ for some } t \in [0,1] \text{ or } x(t) \leq m(t) \text{ for some } t \in [0,1]\}, \quad (4.2)$$

where $M : [0,1] \to (-\infty,\infty]$ is lower semicontinuous and $m : [0,1] \to [-\infty,\infty)$ is upper semicontinuous with $m \leq M$ on $[0,1]$. Lemma 1 states that A_m^M is closed, which implies that A is closed by choosing $m = e$ and $M \equiv -\infty$.

Lemma 1 A_m^M *is closed in* $\mathcal{X}$.

Proof. Let $\{x_n\}$ be a sequence in A_m^M converging in sup norm to some $x \in \mathcal{X}$. Suppose that $x \notin A_m^M$, and set $\epsilon := \min(\inf_{t\in[0,1]}[M(t) - x(t)], \inf_{t\in[0,1]}[x(t) - m(t)])/2$. Since $[0,1]$ is compact, the infima in this expression are attained, so that $\epsilon > 0$. From the convergence in sup norm it follows that $|x_n(t)-x(t)| \leq \epsilon$ for all $t \in [0,1]$ and n large enough. By construction of ϵ, a contradiction is obtained by noting that this would imply $x_n \notin A_m^M$. □

Because A is a non-empty [cf. Assumption 3(iv)] closed set and I is a good rate function, there must exist an $\tilde{x} \in A$ with $I(\tilde{x}) = \inf_{x\in A} I(x)$. Set $A_\tau := \{x \in \mathcal{X} : x(\tau) \geq e(\tau)\}$ for $\tau \in [0,1]$. It is standard that by Jensen's inequality,

$$\inf_{x\in A_\tau} \int_0^1 \Lambda_Z^*(\dot{x}(t))dt \geq \inf_{x\in A_\tau} \tau\frac{1}{\tau}\int_0^\tau \Lambda_Z^*(\dot{x}(t))dt \geq \inf_{x\in A_\tau} \tau\Lambda_Z^*\left(\int_0^\tau \dot{x}(s)ds/\tau\right) = \tau\Lambda_Z^*(e(\tau)/\tau).$$

Consequently, a minimizing argument $\tilde{x}_\tau$ of $\inf_{x\in A_\tau} I(x)$ is a piecewise straight line that reaches e at τ and is constant on $[\tau,1]$. If $\tilde{\tau}$ minimizes $\tau\Lambda_Z^*(e(\tau)/\tau)$, then $\arg\inf_{x\in A} I(x) = \tilde{x}_{\tilde{\tau}}$. Since we know that $\tilde{x}$ exists, we also know that $\tilde{\tau}$ exists (which can also be seen directly). Note that $\tilde{x}$ and therefore $\tilde{\tau}$ may not be unique.

Lemma 2 *If $e(\tilde{\tau})/\tilde{\tau} \in (\mathrm{dom}\,\Lambda_Z^*)^o$ for some $\tilde{\tau}$ with $\tilde{\tau}\Lambda_Z^*(e(\tilde{\tau})/\tilde{\tau}) = \inf_{x\in A} I(x)$, then A is an I-continuity set.*

Proof. Similar arguments as in the proof of Lemma 1 show that $A^o = \{x \in \mathcal{X} : x(t) > e(t) \text{ for some } t \in [0,1]\}$. As A is closed, it suffices to prove that $\inf_{x\in A} I(x) = \inf_{x\in A^o} I(x)$. For $\tau \in [0,1]$ and $\epsilon > 0$, define γ_τ^ϵ by

$$\gamma_\tau^\epsilon(t) := \begin{cases} (e(\tau)+\epsilon)t/\tau & \text{if } 0 \leq t \leq \tau; \\ e(\tau)+\epsilon & \text{if } \tau < t \leq 1. \end{cases} \tag{4.3}$$

Let $\tilde{\tau}$ be such that $\tilde{\tau}\Lambda_Z^*(e(\tilde{\tau})/\tilde{\tau}) = \inf_{x\in A} I(x)$. Note that $\gamma_{\tilde{\tau}}^\epsilon \in A^o$ and that $I(\gamma_{\tilde{\tau}}^\epsilon) = \tilde{\tau}\Lambda_Z^*([e(\tilde{\tau})+\epsilon]/\tilde{\tau})$. By convexity of Λ_Z^* and the fact that there is a neighborhood of $e(\tilde{\tau})/\tilde{\tau}$ on which Λ_Z^* is finite, Λ_Z^* is continuous on this neighborhood, and therefore $\Lambda_Z^*([e(\tilde{\tau})+\epsilon]/\tilde{\tau}) \downarrow \Lambda_Z^*(e(\tilde{\tau})/\tilde{\tau})$ as $\epsilon \downarrow 0$ [note that $\inf_{t\in[0,1]} e(t)/t > 0$ as a consequence of Assumption 3(iv)]. By monotone convergence, $I(\gamma_{\tilde{\tau}}^\epsilon)$ converges to $I(\gamma_{\tilde{\tau}})$. □

Proof of Proposition 5. First note that Assumption 3(iv) guarantees that A is a rare event and avoids trivialities. As a (separable) Banach space, $\mathcal{X}$ is a locally convex topological vector space, as assumed in Assumption 2(i). We already noted that Assumption 3(i) implies the required LDP of Assumption 2(ii), as in Section 5.1 of Dembo and Zeitouni (1998).

To check Assumption 2(iii), Sadowsky (1996, p. 407) shows that there exists a bounded function $f_\xi : [0,1] \to \mathbb{R}$ with the property that $\xi \mapsto f_\xi$ is linear and

$$\Lambda(\xi) = \lim_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} \exp(n\langle\xi, x\rangle)\nu_n(dx) = \int_0^1 \Lambda_Z(f_\xi(t))dt.$$

Since Λ_Z is finite and convex, Λ_Z is continuous. Moreover, the boundedness of f_ξ implies that $\Lambda(\xi)$ is finite for any $\xi \in \mathcal{X}^*$.

The claim that A is an I-continuity set follows from Lemma 2.

We now turn to the convexity. It follows from the proof of the Hölder inequality [see, e.g., Royden (1968)] that Λ_Z is strictly convex when P_Z is non-degenerate [Assumption 3(ii)]. Obviously, Λ inherits strict convexity from Λ_Z, and I is also strictly convex on $\mathcal{AC}$ by Assumption 3(iii). $\square$

We are now ready to apply the results of Section 3.

4.2 Exponentially twisted simulation

To be able to state the necessary and sufficient conditions for the time-varying level-crossing problem, we have to specialize the conditions in Theorem 1 to this case. We have already seen that the paths

$$\gamma_\tau(t) := \begin{cases} t(e(\tau)/\tau) & \text{if } 0 \le t \le \tau; \\ e(\tau) & \text{if } \tau < t \le 1, \end{cases} \tag{4.4}$$

minimize I over A_τ, and that $I(\gamma_\tau) = \tau\Lambda_Z^*(e(\tau)/\tau)$. As earlier, we set

$$\tilde\tau := \arg\inf_{\tau\in[0,1]} \tau\Lambda_Z^*(e(\tau)/\tau).$$

The existence of $\tilde\tau$ was already established from the goodness of the Mogul′skiĭ rate function. Its uniqueness turns out to be essential for efficient simulation, as the following lemma shows.

Lemma 3 *Let Assumption 3 hold, and let, for some $\tilde\tau \in \arg\inf_{\tau\in[0,1]} \tau\Lambda_Z^*(e(\tau)/\tau)$, $e(\tilde\tau)/\tilde\tau \in (\operatorname{dom}\Lambda_Z^*)^o$. There is at most one asymptotically efficient twist.*

If $\tilde\tau$ is unique, the only exponential twist that can achieve asymptotic efficiency is $\langle \xi_{\tilde\tau}, x\rangle = \alpha x(\tilde\tau)$, where $\alpha = \arg\sup_{\theta\in\mathbb{R}}[\theta(e(\tilde\tau)/\tilde\tau) - \Lambda_Z(\theta)]$.

If there are two points $\tilde\tau_1, \tilde\tau_2 \in \arg\inf_{\tau\in[0,1]} \tau\Lambda_Z^*(e(\tau)/\tau)$ satisfying $e(\tilde\tau_1)/\tilde\tau_1, e(\tilde\tau_2)/\tilde\tau_2 \in (\operatorname{dom}\Lambda_Z^*)^o$ and $\tilde\tau_1 \neq \tilde\tau_2$, there is no asymptotically efficient exponential twist.

Proof. The first claim follows from Corollary 2 and Proposition 5.

We now focus on the case that $\tilde\tau$ is unique. It is readily seen that $\xi_{\tilde\tau} \in \mathcal{X}^*$. As is clear from the proof of Theorem 3, the first part of the claim follows once we have shown that $I(\gamma_{\tilde\tau}) = \langle \xi_{\tilde\tau}, \gamma_{\tilde\tau}\rangle - \Lambda(\xi_{\tilde\tau})$. Let $\underline{\tilde\tau}_n := \lfloor n\tilde\tau\rfloor/n$, so that $\underline{\tilde\tau}_n \to \tilde\tau$ as $n\to\infty$. Then we have by independence,

$$\begin{aligned}
\int_{\mathcal{X}} \exp(n\alpha\langle\xi_{\tilde\tau}, x\rangle)\nu_n(dx) &= \int_{\mathcal{X}} \exp(n\alpha x(\tilde\tau))\nu_n(dx) \\
&= \int_{\mathbb{R}^n} \exp\left(\alpha\sum_{i=1}^{n\underline{\tilde\tau}_n} z_i + \alpha(\tilde\tau - \underline{\tilde\tau}_n) z_{n\underline{\tilde\tau}_n+1}\right) P_Z(dz_1)\cdots P_Z(dz_n) \\
&= \int_{\mathbb{R}} \exp(\alpha(\tilde\tau - \underline{\tilde\tau}_n)z) P_Z(dz) \left(\int_{\mathbb{R}} \exp(\alpha z) P_Z(dz)\right)^{n\underline{\tilde\tau}_n}.
\end{aligned}$$

Observe that $\Lambda_Z(\theta) < \infty$ for all $\theta\in\mathbb{R}$ by Assumption 3(i), and that Λ_Z is continuous due to its convexity. Consequently, the first integral of the last expression converges to 1. We conclude that

$$\Lambda(\xi_{\tilde\tau}) = \lim_{n\to\infty} \frac{1}{n}\log\int_{\mathcal{X}} \exp(n\alpha x(\tilde\tau))\nu_n(dx) = \tilde\tau\Lambda_Z(\alpha),$$

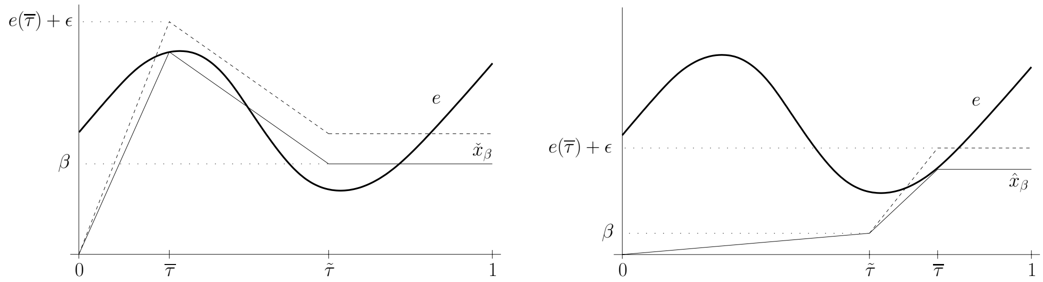

Figure 3: Two possibilities for the minimizing argument of $\inf_{x\in A}[I(x)+\alpha x(\tilde{\tau})]$.

implying $\xi_{\tilde{\tau}}(\gamma_{\tilde{\tau}})-\Lambda(\xi_{\tilde{\tau}})=\alpha e(\tilde{\tau})-\tilde{\tau}\Lambda_Z(\alpha)$. By definition of α, this equals $\tilde{\tau}\Lambda_Z^*(e(\tilde{\tau})/\tilde{\tau})=I(\gamma_{\tilde{\tau}})$.

The last claim is a consequence of Theorem 3. □

Motivated by Lemma 3, we assume the uniqueness of the minimizer $\tilde{\tau}$ of $\tau\Lambda_Z^*(e(\tau)/\tau)$ in the remainder of this section. Before the necessary and sufficient condition of Section 3 can be applied to this problem, we need some additional assumptions to ensure that A is an $(I+\xi_{\tilde{\tau}})$-continuity set.

A minimizing argument of $\inf_{x\in A}[I(x)+\alpha x(\tilde{\tau})]$ must hit e in $[0,1]$ and has some value $\beta\in\mathbb{R}$ at time $\tilde{\tau}$. Moreover, it is a piecewise straight line by Jensen's inequality, so that there are two possibilities. In the first case, x reaches $e(\tau)$ at some $\tau\le\tilde{\tau}$, then assumes some value $\beta\in\mathbb{R}$ at $\tilde{\tau}$, and is constant on $[\tilde{\tau},1]$. This path is denoted by $\check{x}_{\beta,\tau}$. Another possibility is that x has some value β at $\tilde{\tau}$, reaches $e(\tau)$ for some $\tau>\tilde{\tau}$, and then becomes constant. This path is denoted by $\hat{x}_{\beta,\tau}$. The two possible cases are illustrated by the solid lines in Figure 3.

Lemma 4 *Assume that $\tilde{\tau}$ is unique. If one of the following two conditions holds, then A is an $(I+\xi_{\tilde{\tau}})$-continuity set:*

1. *there exists an $\overline{x}\in\arg\inf_{x\in A}[I(x)+\alpha x(\tilde{\tau})]$ of the form $\check{x}_{\beta,\overline{\tau}}$ for some $\beta\in\mathbb{R}$ and $\overline{\tau}\le\tilde{\tau}$, for which $e(\overline{\tau})/\overline{\tau}\in(\mathrm{dom}\,\Lambda_Z^*)^o$,*

2. *there exists an $\overline{x}\in\arg\inf_{x\in A}[I(x)+\alpha x(\tilde{\tau})]$ of the form $\hat{x}_{\beta,\overline{\tau}}$ for some $\beta\in\mathbb{R}$ and $\overline{\tau}>\tilde{\tau}$, for which $(e(\overline{\tau})-\overline{x}(\tilde{\tau})/(\overline{\tau}-\tilde{\tau})\in(\mathrm{dom}\,\Lambda_Z^*)^o$.*

Proof. The claims can be proven using an ϵ-argument similar to the one used the proof of Lemma 2. The perturbed paths are drawn as the dashed lines in Figure 3; the details are left to the reader. □

Lemma 2 and Lemma 4 are useful for applying the following proposition.

Proposition 6 *Let Assumption 3 hold, and suppose that $\tau\Lambda_Z^*(e(\tau)/\tau)$ has a unique minimizer $\tilde{\tau}$. Moreover, let A be both an I-continuity set and an $(I+\xi_{\tilde{\tau}})$-continuity set.*

The exponential twist $\xi_{\tilde{\tau}}(x)=\alpha x(\tilde{\tau})$ is asymptotically efficient if and only if

$$\begin{aligned}\tilde{\tau}\Lambda_Z^*\left(\tfrac{e(\tilde{\tau})}{\tilde{\tau}}\right)+\alpha e(\tilde{\tau})=\min\Big\{&\inf_{\tau\in[0,\tilde{\tau}]}\left(\tau\Lambda_Z^*\left(\tfrac{e(\tau)}{\tau}\right)+\inf_{\beta\in\mathbb{R}}\left[(\tilde{\tau}-\tau)\Lambda_Z^*\left(\tfrac{\beta-e(\tau)}{\tilde{\tau}-\tau}\right)+\alpha\beta\right]\right),\\ &\inf_{\tau\in(\tilde{\tau},1]}\inf_\beta\left[\tilde{\tau}\Lambda_Z^*\left(\tfrac{\beta}{\tilde{\tau}}\right)+(\tau-\tilde{\tau})\Lambda_Z^*\left(\tfrac{e(\tau)-\beta}{\tau-\tilde{\tau}}\right)+\alpha\beta\right]\Big\}.\end{aligned}\tag{4.5}$$

Proof. The claim follows from Corollary 1. Indeed, note that

$$-\inf_{x\in\mathcal{X}}[I(x)-\langle\xi_{\tilde\tau},x\rangle]=\Lambda(\xi_{\tilde\tau})=\tilde\tau\Lambda_Z(\alpha)=\alpha e(\tilde\tau)-\tilde\tau\Lambda_Z^*(e(\tilde\tau)/\tilde\tau),$$

so the left hand side of (4.5) is just $2\inf_{x\in\overline{A}}I(x)-\inf_{x\in\mathcal{X}}[I(x)-\langle\xi_{\tilde\tau},x\rangle]$. As discussed before Lemma 4, there are two possible cases for the minimizing argument of $\inf_{x\in A}[I(x)+\alpha x(\tilde\tau)]$. It is immediate that $\check{x}_{\beta,\tau}$ satisfies for $\tau\le\tilde\tau$

$$I(\check{x}_{\beta,\tau})+\alpha\check{x}_{\beta,\tau}(\tilde\tau)=\tau\Lambda_Z^*\left(\frac{e(\tau)}{\tau}\right)+(\tilde\tau-\tau)\Lambda_Z^*\left(\frac{\beta-e(\tau)}{\tilde\tau-\tau}\right)+\alpha\beta,$$

where the second term should be interpreted as zero for $\tau=\tilde\tau$. This corresponds to the left panel in Figure 3. By the same reasoning, one obtains for the other possible paths $\hat{x}_{\beta,\tau}$,

$$I(\hat{x}_{\beta,\tau})+\alpha\hat{x}_{\beta,\tau}(\tilde\tau)=\tilde\tau\Lambda_Z^*\left(\frac{\beta}{\tilde\tau}\right)+(\tau-\tilde\tau)\Lambda_Z^*\left(\frac{e(\tau)-\beta}{\tau-\tilde\tau}\right)+\alpha\beta.$$

Hence, the right hand side of (4.5) equals $\inf_{x\in A}[I(x)+\alpha x(\tilde\tau)]$. □

Remark. Equation (4.5) can be slightly simplified using Λ_Z. Note that

$$\begin{aligned}\inf_{\beta\in\mathbb{R}}\left[(\tilde\tau-\tau)\Lambda_Z^*\left(\tfrac{\beta-e(\tau)}{\tilde\tau-\tau}\right)+\alpha\beta\right]&=\\ -(\tilde\tau-\tau)\sup_{\beta\in\mathbb{R}}\left[-\alpha\tfrac{\beta-e(\tau)}{\tilde\tau-\tau}-\Lambda_Z^*\left(\tfrac{\beta-e(\tau)}{\tilde\tau-\tau}\right)\right]&+\alpha e(\tau),\end{aligned}$$

and that the sup-term in this expression equals $\Lambda_Z(-\alpha)$ by the duality lemma [Lemma 4.5.8 of Dembo and Zeitouni (1998)]. Thus, (4.5) is equivalent to

$$\begin{aligned}\tilde\tau\Lambda_Z^*\left(\tfrac{e(\tilde\tau)}{\tilde\tau}\right)+\alpha e(\tilde\tau)=\min\Big\{&\inf_{\tau\in[0,\tilde\tau]}\left(\tau\Lambda_Z^*\left(\tfrac{e(\tau)}{\tau}\right)-(\tilde\tau-\tau)\Lambda_Z(-\alpha)+\alpha e(\tau)\right),\\ &\inf_{\tau\in(\tilde\tau,1]}\inf_\beta\left[\tilde\tau\Lambda_Z^*\left(\tfrac{\beta}{\tilde\tau}\right)+(\tau-\tilde\tau)\Lambda_Z^*\left(\tfrac{e(\tau)-\beta}{\tau-\tilde\tau}\right)+\alpha\beta\right]\Big\}.\end{aligned}$$

□

Example. Let the Z_i have a standard normal distribution, i.e., $\Lambda_Z(\xi)=\Lambda_Z^*(\xi)=\frac{1}{2}\xi^2$. Set $e(\tau)=1+|2\tau-1|$. It is easy to check that $\tau\Lambda_Z^*(e(\tau)/\tau)=e(\tau)^2/(2\tau)$ is minimized for $\tilde\tau=1/2$, and thus $\alpha=2$. It is also immediate that $e(\tau)^2/(2\tau)+2\tau-1+2e(\tau)$ attains its minimal value over $[0,1/2]$ in $\tau=1/2$. The minimizing β in (4.5) turns out to be $1/(2\tau)$, and the infimum over $[1/2,1]$ of the resulting function is $\tau=1/2$. Consequently, we can estimate the desired probability efficiently by exponential twisting. Therefore, this example corrects the unproven claim of Sadowsky (1996) that no exponential twist is asymptotically efficient.

Different behavior is observed if $e(\tau)=1+|\tau-1/2|$. Again, $\tilde\tau=1/2$ and $\alpha=2$, but now it turns out that the infimum in (4.5) is attained for $\tau=1$. Therefore, the *same* twist as earlier is now asymptotically *in*efficient. □

Obviously, to implement the simulation procedure, the exponential twist in the abstract Banach setting should be translated into an importance sampling distribution for $(Z_1,\dots,Z_n)$. Sadowsky (1996) shows that the exponentially θ-twisted distribution of Z,

$$P_Z^\theta(dz):=\exp(\theta z-\Lambda_Z(\theta))P_Z(dz),$$

are the 'building blocks' for the required exponential twist. Namely, $Z_1, \ldots, Z_{\lfloor n\tilde{\tau} \rfloor}$ should be sampled from P_Z^α, $Z_{\lceil n\tilde{\tau} \rceil}$ from $P_Z^{\alpha(n\tilde{\tau} - \lfloor n\tilde{\tau} \rfloor)}$, and $Z_{\lceil n\tilde{\tau} \rceil + 1}, \ldots, Z_n$ from P_Z; the Z_i should also be mutually independent. Using the realizations of the Z_i, one can construct a sample path with (4.1). The resulting paths are samples from the exponentially $\xi_{\tilde{\tau}}$-twisted distribution $\lambda_n^{\xi_{\tilde{\tau}}}$.

4.3 Sequential simulation

The remainder of this section is devoted to a simplification of the simulation scheme (i.e., the measure $\lambda_n^{\xi_{\tilde{\tau}}}$) studied in Section 4.2. The new scheme overcomes an intuitive difficulty with an exponentially twisted change of measure. Suppose a path sampled from $\lambda_n^{\xi_{\tilde{\tau}}}$ remains below e on $[0, \lceil n\tilde{\tau} \rceil / n]$, it has little chance of hitting e after $\lceil n\tilde{\tau} \rceil / n$. Indeed, sampling from the original measure P_Z typically avoids hitting e. By the form of the estimator (2.2), such a sample path does not contribute to the resulting estimate.

The idea of the simplified simulation scheme is to sample every random variable Z_i from P_Z^α, *until e has been hit*. The simulation is then stopped. This setup was studied in Sadowsky (1996), where it was called *sequential sampling*. Note that this contrasts with exponential twisting as described in the preceding subsection, since there we twist up to a *fixed* twist-horizon $n\tilde{\tau}$. In the simplified setting of this subsection, this horizon is sample-dependent.

Since both exponential twisting and sequential sampling are algorithms for estimating the same probability, it is legitimate to ask which procedure is better. To answer this question to some extent, our aim is to develop conditions for asymptotic efficiency of sequential sampling. These conditions are then the 'sequential analog' of Proposition 6.

Intuitively, it depends on the specific form of e if the probability of hitting e on $[\lceil n\tilde{\tau} \rceil / n, 1]$ is small enough for the simplification to work. In Proposition 2 of Sadowsky (1996), a sufficient condition is found in terms of a saddle point inequality. The sufficient condition of Proposition 7 improves this result significantly. Moreover, our necessary condition is also extremely 'close' to the sufficiency condition.

Throughout this subsection, we adopt the setup and notation that was introduced in Subsection 4.1. Also recall the definition of α in Lemma 3. Since the assumptions used here differ slightly from Assumption 3, it is worth to state them as Assumption 4.

Assumption 4 *We assume that*

(i) $\Lambda_Z(\theta) < \infty$ *for all* $\theta \in \mathbb{R}$,

(ii) P_Z *is non-degererate,*

(iii) $\tilde{\tau} \in \arg\inf \tau \Lambda_Z^*(e(\tau)/\tau)$ *is unique,*

(iv) $e(\tilde{\tau})/\tilde{\tau} \in (\mathrm{dom}\, \Lambda_Z^*)^o$,

(v) $0 < \inf_{t \in [0,1]} e(t) < \infty$.

In the previous subsection, we have seen that this set of assumptions guarantees that an LDP holds, and that A is an I-continuity set [Assumption 4(iii) and Assumption 4(iv)], see Lemma 2. The uniqueness of $\tilde{\tau}$ of Assumption 4(iii) is required to have a unique twist α for the distribution of the Z_i.

Proposition 7 *Let Assumption 4 hold, and let e be lower semicontinuous. Sequential sampling is asymptotically efficient if*

$$\inf_{t\in[0,1]} \left[t\Lambda_Z^*\left(\frac{e(t)}{t}\right) + \alpha e(t) - t\Lambda_Z(\alpha) \right] = 2\tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right). \tag{4.6}$$

Let Assumption 4 hold, and let e be upper semicontinuous. If sequential sampling is asymptotically efficient, then

$$\inf_{\{t\in(0,1]:e(t)/t\in(\mathrm{dom}\,\Lambda_Z^*)^o\}} \left[t\Lambda_Z^*\left(\frac{e(t)}{t}\right) + \alpha e(t) - t\Lambda_Z(\alpha) \right] = 2\tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right). \tag{4.7}$$

Proof. It is important to note that (2.6) is equivalent to asymptotic efficiency, since A is an I-continuity set.

We introduce some notation used throughout the proof.

Notation. Let $g : [0,1] \to [0,\infty]$ be given by $g(t) := t\Lambda_Z^*(e(t)/t)$ for $t > 0$ and $g(0) := 0$, and define $f : [0,1] \to [-\infty,\infty)$ by

$$f(t) := -\alpha e(t) + t\Lambda_Z(\alpha).$$

Define for $\tau \in (0,1]$

$$\tilde{A}_\tau := \{x \in \mathcal{X} : x(t) < e(t) \text{ for } t \in [0,\tau), x(\tau) \geq e(\tau)\},$$

i.e., $\tilde{A}_\tau$ are the paths that hit e for the first time at τ. Note that the $\tilde{A}_\tau$ are disjoint and that $\bigcup_{\tau\in[0,1]} \tilde{A}_\tau = A$.

Paths generated by the sequential sampling procedure are in general no elements of $\mathcal{X}$, since the simulation is stopped at some random time. To overcome this, note that stopping a simulation run is the same as continuing the simulation by drawing from P_Z. In other words, importance sampling is 'turned off' in the sense that the sampling distribution becomes P_Z after hitting e. Therefore, the distribution in $\mathcal{X}$ of sample paths generated by the sequential estimation procedure is well-defined and is denoted by μ_n. One can readily check that on $\tilde{A}_\tau$, we have (for x in the support of ν_n)

$$\frac{d\nu_n}{d\mu_n}(x) = \exp(-n\alpha x(\tau) + n\tau\Lambda_Z(\alpha)). \tag{4.8}$$

In the proof of the sufficient condition, we use the function $\zeta : \mathcal{X} \to [-\infty,\infty)$ given by

$$\zeta(x) := \begin{cases} f(\tau) & \text{if } x \in \tilde{A}_\tau; \\ -\infty & \text{otherwise.} \end{cases}$$

Since ζ is in general not upper semicontinuous, we cannot apply Varadhan's Integral Lemma to prove the sufficient condition. However, we it is quite fruitful to use some ideas of its proof [Theorem 4.3.1 in Dembo and Zeitouni (1998)].

The sufficient condition. In the proof it is essential that the functions involved have specific continuity properties. Obviously, f is upper semicontinuous under the assumption that e is lower semicontinuous. In order to see that g is lower semicontinuous, let $\{t_n\}$ be a sequence in $[0,1]$ converging to some $t \in [0,1]$. For $t=0$, it certainly holds that $\liminf_{n\to\infty} g(t_n) \geq 0 = g(0)$. Therefore, we assume $t>0$. Since $\inf_{t\in[0,1]} e(t)/t > 0$ and Λ_Z^* is non-decreasing on $[0,\infty)$ (Z_1 is centered), we observe that

$$\begin{aligned}
\liminf_n t_n \Lambda_Z^*(e(t_n)/t_n) &= t \liminf_n \Lambda_Z^*(e(t_n)/t_n) = t\Lambda_Z^*(\liminf_n e(t_n)/t_n) \\
&= t\Lambda_Z^*(\liminf_n e(t_n)/t) \geq t\Lambda_Z^*(e(t)/t),
\end{aligned}$$

where the last inequality uses the lower semicontinuity of e. Hence, g is lower semicontinuous.

Let $\epsilon > 0$. For any $t \in [0,1]$, by semicontinuity we know that there exists an open neighborhood T_t of t with

$$\inf_{\tau\in\overline{T_t}} g(\tau) \geq g(t) - \epsilon \text{ and } \sup_{\tau\in\overline{T_t}} f(\tau) \leq f(t) + \epsilon. \tag{4.9}$$

Since $\bigcup_{t\in[0,1]} T_t$ is an open cover of the compact space $[0,1]$, one can find N and $t_1,\ldots,t_N \in [0,1]$ such that $\bigcup_{i=1}^N T_{t_i} = [0,1]$.

As $d\nu_n/d\mu_n \leq \exp(n\zeta)$ on each of the sets $\tilde{A}_\tau$, the cover-argument implies that [see Lemma 1.2.15 of Dembo and Zeitouni (1998)]

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \int_A \frac{d\nu_n}{d\mu_n} d\nu_n &\leq \limsup_{n\to\infty} \frac{1}{n} \log \int_A \exp(n\zeta(x))\nu_n(dx) \\
&= \max_{i=1}^N \limsup_{n\to\infty} \frac{1}{n} \log \int_{\bigcup_{\tau\in T_{t_i}} \tilde{A}_\tau} \exp(n\zeta(x))\nu_n(dx).
\end{aligned}$$

The integral in this expression can be bounded by noting that ζ is majorized on $\bigcup_{\tau\in T_{t_i}} \tilde{A}_\tau$ using (4.9):

$$\int_{\bigcup_{\tau\in T_{t_i}} \tilde{A}_\tau} \exp(n\zeta(x))\nu_n(dx) \leq \exp[f(t_i)+\epsilon]\nu_n\left(\bigcup_{\tau\in T_{t_i}} \tilde{A}_\tau\right)$$

Although $\bigcup_{\tau\in T_{t_i}} \tilde{A}_\tau$ is in general not closed, it is contained in the closed set $\{x : x(t) \geq e(t) \text{ for some } t \in \overline{T_{t_i}}\}$ [to see that this set is closed, use Lemma 1 for $M = e$ on $\overline{T_{t_i}}$ and $M = \infty$ on $[0,1]\backslash\overline{T_{t_i}}$]. Therefore, by the large deviation upper bound, Jensen's inequality, and (4.9),

$$\limsup_{n\to\infty} \frac{1}{n} \log \nu_n\left(\bigcup_{\tau\in T_{t_i}} \tilde{A}_\tau\right) \leq - \inf_{\{x:x(t)\geq e(t) \text{ for some } t\in\overline{T_{t_i}}\}} I(x) = -\inf_{t\in\overline{T_{t_i}}} g(t) \leq -g(t_i)+\epsilon.$$

Combining the preceding three displays, we obtain

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \int_A \frac{d\nu_n}{d\mu_n} d\nu_n &\leq \max_{i=1}^N [f(t_i) - g(t_i)] + 2\epsilon \\
&\leq \sup_{t\in[0,1]} [f(t) - g(t)] + 2\epsilon.
\end{aligned}$$

The sufficiency condition follows by letting $\epsilon \to 0$.

The necessary condition. We now turn to the necessary condition. Since A is an I-continuity set and we suppose that sequential sampling is asymptotically efficient, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_A \frac{d\nu_n}{d\mu_n} d\nu_n \leq -2\tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right). \tag{4.10}$$

Let $\epsilon > 0$. The upper semicontinuity of e implies that for all $t \in (0,1]$ there exists a $\delta \in (0,t)$ such that

$$\sup_{\tau\in(t-\delta,t]} e(\tau) \leq e(t) + \epsilon. \tag{4.11}$$

Fix $t \in (0,1]$, and define

$$A_t^{\delta,\epsilon} := \left\{ \begin{array}{ll} x: & x(\tau) < e(\tau) \text{ for } \tau \in [0, t-\delta]; x(t) > e(t); \\ & x(\tau) < \sup_{s\in(t-\delta,t]} e(s) + \epsilon \text{ for } \tau \in (t-\delta, t] \end{array} \right\}.$$

Note that $A_t^{\delta,\epsilon} \subset A$ and that it is open by the fact that A_m^M in Lemma 1 is closed [set $m(t) = e(t)$ and $m = -\infty$ on $[0,1]\backslash\{t\}$; $M = e$ on $[0, t-\delta]$ and $M = \sup_{s\in(t-\delta,t]} e(s) + \epsilon$ on $(t-\delta, t]$].

We deduce that by definition of $A_t^{\delta,\epsilon}$,

$$\begin{aligned} \frac{1}{n} \log \int_A \frac{d\nu_n}{d\mu_n} d\nu_n &\geq \frac{1}{n} \log \int_{A_t^{\delta,\epsilon}} \frac{d\nu_n}{d\mu_n} d\nu_n \\ &\geq \frac{1}{n} \log \int_{A_t^{\delta,\epsilon}} \exp\left(-n\alpha \left[\sup_{\tau\in(t-\delta,t]} e(\tau) + \epsilon\right] + nt\Lambda_Z(\alpha)\right) \nu_n(dx) \\ &\geq -\alpha[e(t) + 2\epsilon] + t\Lambda_Z(\alpha) + \frac{1}{n} \log \nu_n(A_t^{\delta,\epsilon}), \end{aligned}$$

where we used (4.11) for the last inequality.

Recall the definition of γ_τ and γ_τ^ϵ in (4.4) and (4.3). Now two cases are distinguised.

Case 1: γ_t and e do not intersect before t.

Let t be such that γ_t and e do not intersect before t. Choose δ such that (4.11) is met, and set

$$\eta := \frac{1}{2} \min\left(\inf_{\tau\in[0,t-\delta]} [e(\tau) - \gamma_t(\tau)], \epsilon\right).$$

By the usual arguments, it is readily seen that $\eta > 0$ and $\gamma_t^\eta \in A_t^{\delta,\epsilon}$. Since $I(\gamma_t^\eta) = t\Lambda_Z^*([e(t)+\eta]/t)$, we have by monotonicity of Λ_Z^* on $[0,\infty)$ and the large deviation lower bound,

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log \int_A \frac{d\nu_n}{d\mu_n} d\nu_n &\geq f(t) - 2\alpha\epsilon - \inf_{x\in A_t^{\delta,\epsilon}} I(x) \\ &\geq f(t) - 2\alpha\epsilon - t\Lambda_Z^*([e(t)+\eta]/t) \\ &\geq f(t) - 2\alpha\epsilon - t\Lambda_Z^*([e(t)+\epsilon/2]/t). \end{aligned}$$

Since ϵ was arbitrary, we obtain a nontrivial lower bound if $e(t)/t \in (\mathrm{dom}\,\Lambda_Z^*)^o$.

An auxiliary result. Before proceeding with the complementary case, we first prove an auxiliary result: asymptotic efficiency implies that for any $t \in (0,1]$ with $e(t)/t \in (\mathrm{dom}\,\Lambda_Z^*)^o$,

$$\alpha\frac{e(t)}{t} - \Lambda_Z(\alpha) + \Lambda_Z^*\left(\frac{e(t)}{t}\right) \geq 0. \tag{4.12}$$

We work towards a contradiction by supposing that (4.12) is not satisfied for some $\hat{t}$ with $e(\hat{t})/\hat{t} \in (\mathrm{dom}\,\Lambda_Z^*)^o$. Without loss of generality, we may suppose that $\gamma_{\hat{t}}$ does not intersect with e before $\hat{t}$. By the above derived lower bound for 'Case 1',

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\int_A\frac{d\nu_n}{d\mu_n}d\nu_n &\geq \liminf_{n\to\infty}\frac{1}{n}\log\int_A\frac{d\nu_n}{d\mu_n}d\nu_n\\ &\geq f(\hat{t}) - \hat{t}\Lambda_Z^*\left(\frac{e(\hat{t})}{\hat{t}}\right)\\ &> 0.\end{aligned}$$

Since $-2\tilde{\tau}\Lambda_Z^*(e(\tilde{\tau})/\tilde{\tau}) \leq 0$, this contradicts the assumption that sequential sampling is asymptotically efficient.

Case 2: γ_t intersects e before t. We now suppose that γ_t intersects e before t, and the first time that this occurs is denoted by $\bar{t} < t$. Use $e(t)/t = e(\bar{t})/\bar{t}$ and the 'auxiliary result' to see that

$$\begin{aligned}-f(t) + t\Lambda_Z^*\left(\frac{e(t)}{t}\right) &= t\left[\alpha\frac{e(t)}{t} - \Lambda_Z(\alpha) + \Lambda_Z^*\left(\frac{e(t)}{t}\right)\right]\\ &\geq \bar{t}\left[\alpha\frac{e(\bar{t})}{\bar{t}} - \Lambda_Z(\alpha) + \Lambda_Z^*\left(\frac{e(\bar{t})}{\bar{t}}\right)\right]\\ &= -f(\bar{t}) + \bar{t}\Lambda_Z^*\left(\frac{e(\bar{t})}{\bar{t}}\right).\end{aligned}$$

We conclude that the infimum in (4.7) will not be attained by a t for which γ_t intersects with e before t.

Therefore, if sequential sampling is asymptotically efficient, we must have by (4.10)

$$\begin{aligned}\inf_{\{t\in(0,1]:e(t)/t\in(\mathrm{dom}\,\Lambda_Z^*)^o\}}\left[t\Lambda_Z^*\left(\frac{e(t)}{t}\right) - f(t)\right] &\geq -\liminf_{n\to\infty}\frac{1}{n}\log\int_A\frac{d\nu_n}{d\mu_n}d\nu_n\\ &\geq -\limsup_{n\to\infty}\frac{1}{n}\log\int_A\frac{d\nu_n}{d\mu_n}d\nu_n\\ &\geq 2\tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right).\end{aligned}$$

The reverse inequality follows from the assumption that $e(\tilde{\tau})/\tilde{\tau} \in (\mathrm{dom}\,\Lambda_Z^*)^o$. □

As a result of the sufficient condition in Proposition 7, sequential sampling is asymptotically efficient if the saddle point inequality

$$\alpha e(t) - t\Lambda_Z(\alpha) \geq \tilde{\tau}\Lambda_Z^*(e(\tilde{\tau})/\tilde{\tau})$$

holds for all $t \in [0,1]$. This sufficient condition was given by Sadowsky (1996).

Example. Consider the example given on page 21, in which $e(\tau) = 1 + |\tau - 1/2|$. We saw already that $\tilde{\tau} = 1/2$ and $\alpha = 2$. The infimum on the left hand side of (4.6) is attained at $\tau = 1/2$, which implies that sequential sampling is asymptotically efficient. Note that exponential twisting was not asymptotically efficient. □

5. Discussion

In case any exponential twist for estimating $\nu(A)$ is asymptotically inefficient, there are a number of alternatives. First of all, it may be possible to write the rare event A as a union of $m < \infty$ disjoint rare events $A_1, \ldots, A_m$, for which the probabilities can be estimated efficiently by an exponential twist. The sum of these probabilities is then an asymptotically efficient estimator for $\nu(A)$. In many applications, however, A cannot be written in that form. To overcome this, one can approximate $\nu(A)$ by $\nu(\bigcup_{i=1}^m A_i)$ for suitably chosen $A_1, \ldots, A_m$ and bound the error in some sense, as in Boots and Mandjes (2002). A variant of this approach is based on mixing relevant exponential twists; details can be found in Sadowsky and Bucklew (1990). In a hitting probability framework, Collamore (2002) uses related ideas to find an estimator that is arbitrarily 'close' to asymptotic efficiency.

Another possibility to deal with asymptotically inefficient exponential twists is the recent adaptive approach to importance sampling described by Dupuis and Wang (2002). Although the authors illustrate this approach in an setting based on Cramér's Theorem, they claim it is useful in a more general setting. This dynamic exponential twisting contrasts with the approach taken in this paper, as we consider a fixed exponential twist.

Although our definition of asymptotic efficiency is mathematically convenient, several other criteria for discriminating between estimators have been proposed. Notably, the amount of time (or work) required to generate one simulation replication was not taken into account in our definition of asymptotic efficiency. Glynn and Whitt (1992) elaborate on a definition in which this is incorporated.

As the application in Section 4 indicates, it requires some work to apply the rather abstract conditions in a practical problem. In particular, the conditions can be used to find an efficient exponential twist in an abstract setting, but this twist should be translated in an implementable change of measure. This translation can only be done on a case by case basis, if possible at all.

Acknowledgment

The authors thank Zbigniew Palmowski for stimulating discussions resulting in the counterexample of Section 3.4.

References

Asmussen, S. (1989). Risk theory in a Markovian environment. *Scand. Actuar. J.* 69–100.

Asmussen, S. (2000). *Ruin probabilities.* World Scientific Publishing Co. Inc.

Asmussen, S., Fuckerieder, P., Jobmann, M., and Schwefel, H.-P. (2002). Large deviations and fast simulation in the presence of boundaries. *Stochastic Process. Appl.* **102**, 1–23.

Asmussen, S. and Rubinstein, R. Y. (1995). Steady state rare events simulation in queueing models and its complexity properties. In Dshalalow, J., editor, *Advances in queueing*, 429–461. CRC, Boca Raton, FL.

Boots, N. K. and Mandjes, M. (2002). Fast simulation of a queue fed by a superposition of many (heavy-tailed) sources. *Probab. Engrg. Inform. Sci.* **16**, 205–232.

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.

Collamore, J. F. (2002). Importance sampling techniques for the multidimensional ruin problem for general Markov additive sequences of random vectors. *Ann. Appl. Probab.* **12**, 382–421.

de Acosta, A. (1994). Large deviations for vector-valued Lévy processes. *Stochastic Process. Appl.* **51**, 75–115.

Dembo, A. and Zajic, T. (1995). Large deviations: from empirical mean and measure to partial sums process. *Stochastic Process. Appl.* **57**, 191–224.

Dembo, A. and Zeitouni, O. (1998). *Large deviations techniques and applications.* Springer-Verlag, New York, second edition.

den Hollander, F. (2000). *Large deviations.* American Mathematical Society, Providence, RI.

Deuschel, J.-D. and Stroock, D. W. (1989). *Large deviations.* Academic Press Inc., Boston, MA.

Dupuis, P. and Wang, H. (2002). Importance sampling, large deviations, and differential games. Technical report, Lefschetz Center for Dynamical Systems, Brown University.

Glasserman, P., Heidelberger, P., and Shahabuddin, P. (1999). Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance* **9**, 117–152.

Glasserman, P. and Kou, S. (1995). Analysis of an importance sampling estimator for tandem queues. *ACM Trans. Modeling Comp. Simulation* **4**, 22–42.

Glasserman, P. and Wang, Y. (1997). Counterexamples in importance sampling for large deviations probabilities. *Ann. Appl. Probab.* **7**, 731–746.

Glynn, P. W. and Whitt, W. (1992). The asymptotic efficiency of simulation estimators. *Oper. Res.* **40**, 505–520.

Heidelberger, P. (1995). Fast simulation of rare events in queueing and reliability models. *ACM Trans. Modeling Comp. Simulation* **5**, 43–85.

Lehtonen, T. and Nyrhinen, H. (1992a). On asymptotically efficient simulation of ruin probabilities in a Markovian environment. *Scand. Actuar. J.* 60–75.

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

Mogul′skiĭ, A. A. (1976). Large deviations for the trajectories of multidimensional random walks. *Teor. Verojatnost. i Primenen.* **21**, 309–323.

Mogul′skiĭ, A. A. (1993). Large deviations for processes with independent increments. *Ann. Probab.* **21**, 202–215.

Royden, H. L. (1968). *Real analysis.* The Macmillan Co., New York.

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. (1996). On Monte Carlo estimation of large deviations probabilities. *Ann. Appl. Probab.* **6**, 399–422.

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.

Szechtman, R. and Glynn, P. W. (2002). Rare-event simulation for infinite server queues. In *Proceedings of the 2002 Winter Simulation Conference.*