Efficient simulation of random walks exceeding a nonlinear boundary

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

Let $S_n : [0,1] \to \mathbb{R}$ denote the polygonal approximation of a random walk with zero-mean increments, where both time and space are scaled by n. We consider the estimation of the probability that, for fixed $n \in \mathbb{N}$, S_n exceeds some positive function e.

As a result of the scaling, this probability decays exponentially in n, and importance sampling can be used to achieve variance reduction. Two simulation methods are considered: path-level twisting and step-level twisting. We give necessary and sufficient conditions for both methods to be asymptotically efficient as $n \to \infty$. Our conditions improve upon those in earlier work of Sadowsky [17].

1 Introduction

Let $Z_1, Z_2, \dots$ be a sequence of i.i.d. zero-mean random variables taking values in $\mathbb{R}$, with distribution P_Z. For $0 \le t \le 1$, let the scaled polygonal approximation for the partial sums of Z_i be given by

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

where $\lfloor t \rfloor$ denotes the largest integer smaller than or equal to t.

We consider the estimation of a 'time-varying level-crossing probability'. That is, we are interested in estimating $P(S_n(\cdot) \in A)$ efficiently, where

$$A := \{x \in C([0,1]) : x(t) \ge e(t) \text{ for some } t \in [0,1]\}, \tag{2}$$

for some lower semicontinuous function $e : [0,1] \to (0,\infty]$. As $\mathbb{E}S_n(t) = 0$ for any $n \in \mathbb{N}$ and $t \in [0,1]$, the probability $P(S_n(\cdot) \in A)$ clearly corresponds to a *rare event*, i.e., it vanishes as $n \to \infty$.

We remark that it is possible to consider the (more general) problem with noncentered random variables Z and with S_n defined on $[0,T]$ for some $T > 0$. If one imposes that $e : [0,T] \to (-\infty,\infty]$ satisfies $e(t)/t > \mathbb{E}Z_1$ for all $t \in [0,T]$, it is readily seen that we may restrict ourselves without loss of generality to the above simpler setup.

*The research was supported by the Netherlands Organization for Scientific Research (NWO) under grant 631.000.002.

Only in special cases, explicit expressions are available for $P(S_n(\cdot) \in A)$. When these are not known, one may resort to simulation. As the probability of interest is small, direct simulation can be extremely time-consuming. Unfortunately, the development of efficient simulation methods is usually nontrivial. For the special case that e is affine, i.e., $e(t) = a + bt$ for some $a, b \geq 0$, $P(S_n(\cdot) \in A)$ corresponds to a ruin probability for the finite-horizon case. Simulation of this probability is studied by Lehtonen and Nyrhinen [13], while Asmussen [1, Sec. X.4] and Asmussen *et al.* [2] consider the analogue in continuous time; then, a Lévy process replaces the random walk.

In the present paper, two approaches for simulating $P(S_n(\cdot) \in A)$ are considered, both based on importance sampling. That is, samples are drawn from a probability measure under which A is not rare anymore, and the simulation output is then corrected with likelihood ratios to retrieve an unbiased estimate. For both methods, the importance sampling is based on sampling steps from an exponentially *twisted* distribution $P_Z^\theta(dz) = e^{\theta z}\mathbb{E}(e^{\theta Z_1})^{-1}P_Z(dz)$ for some $\theta > 0$.

In the first approach, which we call *path-level twisting*, typical sample paths under the importance-sampling measure have the form of the 'most likely' path to exceedance under the original measure; let $\tilde{\tau}$ be the epoch where this most likely path exceeds e. This results in the following procedure: (i) Sample the Z_i from P_Z^θ for $i = 1, \ldots, n\tilde{\tau}$ (with θ chosen such that these Z_i get a positive mean). Then (ii) sample Z_i from the original distribution P_Z for $i = n\tilde{\tau} + 1, \ldots, n$. It is important to note that in this approach the importance sampling is 'turned off' at an *a priori* determined epoch $n\tilde{\tau}$.

In the second approach, which we call *step-level twisting*, we always draw the Z_i from P_Z^θ, and each simulation run is stopped at the *random* moment that S_n exceeds e for the first time (or, if this event does not occur, the end of the simulation horizon). This, quite natural, method has been considered in [1, 13]. However, it does *not* correspond to exponentially twisting in a path space, which makes it somewhat more difficult to handle from a theoretical point of view.

The two methods that we study in the present paper are not the only possible approaches for simulating $P(S_n(\cdot) \in A)$. We mention in particular a 'sequential' method, which has been recently proposed by Dupuis and Wang [10]. Translated into the setting of our paper, their method would suggest to apply an 'adaptive exponential twist': calculate, for any i, the 'best' exponential twist for Z_i *given* the realizations of $Z_1, \ldots, Z_{i-1}$. It is known that in various situations non-adaptive exponential twisting may have a malign impact on the variance properties of the estimator (and consequently on the simulation speed), see the examples in [11]. The sequential method neutralizes these effects by twisting adaptively. A disadvantage of the sequential method, however, is that the best twist for Z_i (for each i) needs to be calculated through solving an optimization program, which has a considerable impact on the simulation speed [8]. It is therefore of practical interest to know when a simpler, faster method suffices; the main objective of this work is to study this in detail.

As for the technical part of this paper, we rely on elementary large-deviation theory, in particular Mogul′skiĭ's theorem and Varadhan's integral lemma. Mogul′skiĭ's theorem is the basis of the present paper, as it describes the large deviations of $S_n(\cdot)$. In some cases, Varadhan's lemma yields necessary and sufficient conditions for asymptotic efficiency, see [9] for details. Asymptotic efficiency is the standard optimality notion in rare event simulation, and entails that the so-called relative error vanishes on an exponential scale.

Sadowsky [17] was the first to consider the above two simulation methods for estimating $P(S_n(\cdot) \in A)$. The ideas used in the proofs of [9] make it possible to improve upon the results of [17] for both simulation methods. Specifically, for the first method (path-level twisting), we correct Sadowsky's claim that it is *never* asymptotically optimal. For the second method (step-level twisting), we give a sufficient condition for asymptotic efficiency that is sharper than Sadowsky's. We exemplify this by establishing a closely related necessary condition.

The paper is organized as follows. We start with some preliminaries in Section 2. Section 3 discusses path-level twisting, and finds necessary and sufficient conditions for its asymptotic efficiency. Step-level twisting is studied in Section 4; also here conditions for asymptotic efficiency are derived. We compare the efficiency conditions of the two methods in Section 5.

2 Preliminaries

This section provides the basic background on sample-path large deviations, importance sampling and asymptotic efficiency. General references on large-deviation theory are [6, 7], and the reader is referred to [3, 12] and references therein for a detailed discussion on importance sampling and asymptotic efficiency.

2.1 Sample-path large deviations

In this subsection, we describe the large deviations of S_n. We define the cumulant-generating function of Z_1 as $\Lambda_Z(\theta) := \log \mathbb{E}(e^{\theta Z_1})$ for $\theta \in \mathbb{R}$, and we suppose that this function is everywhere finite for simplicity; we refer to Section 3 for a detailed discussion on this assumption.

The Fenchel-Legendre transform of Λ_Z is given by $\Lambda_Z^*(z) := \sup_{\xi\in\mathbb{R}}[\xi z - \Lambda_Z(\xi)]$. Cramér's theorem says that for every closed set F and open set G in $\mathbb{R}$,

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log P(S_n(1) \in G) &\geq -\inf_{z\in G} \Lambda_Z^*(z) \\
\limsup_{n\to\infty} \frac{1}{n} \log P(S_n(1) \in F) &\leq -\inf_{z\in F} \Lambda_Z^*(z).
\end{aligned}$$

This theorem is one of the most well-known examples of a so-called large-deviation principle. The function Λ_Z^* is called the rate function associated to the large-deviation principle. In the present paper, we need a sample-path version of this theorem.

To state such a sample-path version, it is important to first introduce a topology on a path space, so that open and closed sets are defined. The path space that we consider is the space $C([0,1])$ of continuous functions on $[0,1]$. It is a metric space with the metric

$$d(x,y) = \sup_{t\in[0,1]} |x(t) - y(t)|.$$

The space of absolutely continuous functions $\mathcal{AC}$ plays an important role in the sample-path large-deviation principle that we consider here. It is defined as

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

In particular, for any $x \in \mathcal{AC}$, there exists some measurable function $\dot{x}$ such that $\int_0^t \dot{x}(s)ds = x(t)$; if x is differentiable, $\dot{x}$ is the derivative of x. We can now describe the large deviations of the paths S_n; more details can be found in Section 5.1 of Dembo and Zeitouni [6].

Proposition 1 (Mogul′skiĭ) *Suppose that $\Lambda_Z(\theta) < \infty$ for every $\theta \in \mathbb{R}$. For every closed set F and open set G in $C([0,1])$,*

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log P(S_n \in G) &\geq -\inf_{x\in G} I(x) \\
\limsup_{n\to\infty} \frac{1}{n} \log P(S_n \in F) &\leq -\inf_{x\in F} I(x),
\end{aligned}$$

where

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

We remark that the level sets of I are compact, which is used in the sequel to apply results from [9].

Since we are interested in exceedance probabilities, it is of interest to know how to compute $\inf_{x\in G} I(x)$ for 'exceedance sets' G. Fix some $\tau \in [0,1]$ and set $A_\tau := \{x \in C([0,1]) : x(\tau) \geq e(\tau)\}$. It is standard that by Jensen's inequality,

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

Consequently, the unique minimizing argument $\tilde{x}_\tau$ of $\inf_{x\in A_\tau} I(x)$ is a piecewise straight line given by

$$\gamma_\tau(t) := \begin{cases} t(e(\tau)/\tau) & \text{if } 0 \leq t \leq \tau; \\ e(\tau) & \text{if } \tau < t \leq 1. \end{cases} \tag{3}$$

Let us now consider the minimizer of I on A as defined in (2). Note that $A = \bigcup_{\tau\in[0,1]} A_\tau$, and define

$$\tilde{\tau} := \arg\inf_{\tau\in(0,1]} \tau \Lambda_Z^*(e(\tau)/\tau), \tag{4}$$

which exists by lower semicontinuity of $t \mapsto e(t)/t$, but is not necessarily unique.

A minimizer over the set A is then given by $\arg\inf_{x\in A} I(x) = \arg\inf_{x\in A_{\tilde{\tau}}} I(x) = \gamma_{\tilde{\tau}}$. Note that it need not be unique.

It is also useful to define an 'ϵ-perturbed' version of γ_τ. For $\tau \in [0,1]$ and $\epsilon > 0$, define γ_τ^ϵ as

$$\gamma_\tau^\epsilon(t) := \begin{cases} (e(\tau)+\epsilon)t/\tau & \text{if } 0 \leq t \leq \tau; \\ e(\tau)+\epsilon & \text{if } \tau < t \leq 1. \end{cases} \tag{5}$$

2.2 Importance sampling

Given a probability measure ν on $C([0,1])$, we are interested in the simulation of the ν-probability of a given event A, where $\nu(A)$ is small. The idea of importance sampling is to sample from a different distribution on $C([0,1])$, say λ, for which A occurs more frequently. This is done by specifying a measurable function $d\lambda/d\nu : C([0,1]) \to [0,\infty]$ and by setting

$$\lambda(B) := \int_B \frac{d\lambda}{d\nu} d\nu. \tag{6}$$

Since λ must be a probability measure, $d\lambda/d\nu$ should integrate to one 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

$$\nu(A) = \int_A \frac{d\nu}{d\lambda} d\lambda = \int_{C([0,1])} \mathbf{1}_A \frac{d\nu}{d\lambda} d\lambda,$$

where $\mathbf{1}_A$ denotes the indicator function of A. A widely used choice for λ is a so-called *exponentially twisted* measure. That is, for some linear continuous functional $\xi : C([0,1]) \to \mathbb{R}$, λ^ξ is defined by

$$\begin{aligned} \frac{d\lambda^\xi}{d\nu}(x) &:= \exp\left(\xi(x) - \log \int_{C([0,1])} \exp[\xi(y)]\nu(dy) \right) \\ &= \frac{\exp[\xi(x)]}{\int_{C([0,1])} \exp[\xi(y)]\nu(dy)}. \end{aligned}$$

The importance sampling estimator $\widehat{\nu_\lambda(A)}$ of $\nu(A)$ is found by drawing N independent samples $X_\lambda^{(1)}, \ldots, X_\lambda^{(N)}$ from λ:

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

It is clear that $\widehat{\nu_\lambda(A)}$ is an unbiased estimator, i.e., $\mathbb{E}_\lambda \widehat{\nu_\lambda(A)} = \nu(A)$.

Since it may seem unnatural to use an exponential twist on a path space, it is worthwhile to cast these notions into the framework of the present paper. Let ν be the distribution of $S_n(\cdot)$. Any twist ξ directly translates into a *vector* of exponential twists for the distributions of the Z_i in (1). Indeed, let us write $s_n^z(\cdot)$ for the polygonal approximation for the partial sums of the vector $z \in \mathbb{R}^n$. It can be seen that there exists some $\theta^{\xi,n} \in \mathbb{R}^n$ such that for any measurable $B \subset C([0,1])$,

$$\lambda^\xi(B) = \int_{\{z \in \mathbb{R}^n : s_n^z(\cdot) \in B\}} \exp\left(\sum_{i=1}^{n} \theta_i^{\xi,n} z_i - \sum_{i=1}^{n} \Lambda_Z(\theta_i^{\xi,n}) \right) P_Z(dz_1) \cdots P_Z(dz_n). \tag{7}$$

This important representation indicates how each of the step-size distributions of a random walk should be changed when sampling from λ^ξ; only the $\theta^{\xi,n}$ need to be determined.

Since any distribution λ (or twist ξ) gives an unbiased importance-sampling estimator, one has the freedom to select an 'efficient' distribution, i.e., a distribution for which the variance of the estimator is small. In particular, it is of interest to find the change of measure that minimizes this variance, or, equivalently,

$$\int_A \left(\frac{d\nu}{d\lambda} \right)^2 d\lambda = \int_{C([0,1])} \mathbf{1}_A \left(\frac{d\nu}{d\lambda} \right)^2 d\lambda = \int_A \frac{d\nu}{d\lambda} d\nu.$$

A zero-variance estimator is found by letting λ be the conditional distribution of ν given A (see, e.g., [12]), but this change of measure is infeasible for simulation since $d\nu/d\lambda$ then depends on the *unknown* quantity $\nu(A)$. This motivates the use of another optimality criterion, *asymptotic efficiency.*

2.3 Asymptotic efficiency and the relative error

Suppose that a sequence of measures $\{\nu_n\}$ is given, along with a set A. We are interested in $\nu_n(A)$ for fixed n. To estimate this probability, we draw an i.i.d. sample $X_{\lambda_n}^{(1)}, \ldots, X_{\lambda_n}^{(N)}$ from some importance sampling distribution λ_n, as described in the preceding subsection.

The definition of asymptotic efficiency is related to the so-called *relative error.* We define the relative error $\eta_N(\lambda_n, A)$ of the importance sampling estimator

$$\widehat{\nu_{\lambda_n}(A)}_N := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\left\{X_{\lambda_n}^{(i)} \in A\right\}} \frac{d\nu_n}{d\lambda_n} \left(X_{\lambda_n}^{(i)} \right) \tag{8}$$

as

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

The relative error is proportional to the width of a confidence interval relative to the (expected) estimate itself; hence, it measures the variability of $\widehat{\nu_{\lambda_n}(A)}_N$.

In view of this, it is desirable that the relative error associated to the importance-sampling family $\{\lambda_n\}$ be bounded in n, i.e., that $\limsup_{n\to\infty} \eta_N(\lambda_n, A) < \infty$. Since this property rarely holds in practice, a second (much weaker) form of 'efficiency' has been introduced, *asymptotic efficiency.* This allows the relative error to grow, but not on an exponential scale.

Definition 1 *An importance sampling family* $\{\lambda_n\}$ *is called* asymptotically efficient *if*

$$\lim_{n\to\infty} \frac{1}{n} \log \eta_N(\lambda_n, A) = 0. \tag{9}$$

It can be seen (as in, e.g., [9]) that asymptotic efficiency is equivalent to $\frac{1}{n}\log N^*_{\lambda_n} \to 0$, where $N^*_{\lambda_n} := \inf\{N \in \mathbb{N} : \eta_N(\lambda_n, A) \leq \eta_{\max}\}$ for some given maximal relative error $0 < \eta_{\max} < \infty$. In other words, the number of samples required to achieve a fixed relative error $\eta_{\max}$ increases more slowly than any exponential. We note that the specific value of $\eta_{\max}$ is irrelevant.

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

3 Path-level twisting

In this section, we study the simulation of $P(S_n(\cdot) \in A)$ by path-level twisting, where A is defined in (2):

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

This analysis culminates in a necessary and sufficient condition for efficiency of this simulation method.

We start with the formulation of the underlying assumptions:

Assumption 1 *We assume that*

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

(ii) P_Z *is nondegenerate, and*

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

Note that Assumption 1(i) implies that Proposition 1 applies. This assumption can be considerably relaxed for a Mogul′skiĭ-type large-deviation principle to hold; one then uses other spaces, other topologies, and slightly modified rate functions. For instance, Mogul′skiĭ [14] allows the cumulant-generating function to be finite only in a neighborhood of zero and uses the space of càdlàg functions D endowed with the (completed) Skorokhod topology; see also Mogul′skiĭ [15]. Although Mogul′skiĭ's rate function is slightly different from I, the infima over exceedance sets are attained by straight lines as in Section 2.1, which is its only essential property for this paper. In a more general context, Dembo and Zajic [5] and de Acosta [4] work under the hypothesis of a finite cumulant-generating function of $|Z|$; this is equivalent to Assumption 1(i).

It is explained in Section 2.2 how to define exponentially twisted distributions for a random variable with values in $C([0,1])$, but some care needs to be taken due to the presence of the parameter n. For a twist ξ, we introduce the measure λ_n^ξ by setting for any Borel set B in $C([0,1])$,

$$\lambda_n^\xi(B) = \int_B \exp\left(n\xi(x) - \log \int_{C([0,1])} \exp[n\xi(y)] P(S_n \in dy) \right) P(S_n \in dx).$$

Since λ_n^ξ is a measure on the path-space $C([0,1])$, ξ is a *path-level twist*.

Before formulating a condition that is equivalent to asymptotic efficiency, we show in the next lemma that there is at most one asymptotically efficient twist. Furthermore, the uniqueness of $\tilde{\tau}$, as defined in (4), is necessary for a path-level twist to be asymptotically efficient.

For these assertions to hold, A has to be an I-continuity set, that is, it has to satisfy $\inf_{x\in\overline{A}} I(x) = \inf_{x\in A^o} I(x)$, where $\overline{A}$ and A^o denote the closure and interior of A respectively. At the end of this section we give sharp conditions for A to satisfy this and related conditions. We write $\mathrm{dom}\,\Lambda_Z^* := \{z \in \mathbb{R} : \Lambda_Z^*(z) < \infty\}$.

Lemma 1 *Let Assumption 1 hold, and let A be an I-continuity set.*

(i) *There is at most one asymptotically efficient path-level twist.*

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

(iii) *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 $e(\tilde{\tau}_i)/\tilde{\tau}_i \in (\mathrm{dom}\,\Lambda_Z^*)^o$ and $\tilde{\tau}_1 \neq \tilde{\tau}_2$, there is no asymptotically efficient path-level twist.*

Proof. We start with some elementary observations. Define

$$\Lambda(\xi) := \lim_{n\to\infty} \frac{1}{n} \log \int_{C([0,1])} \exp[n\xi(x)] P(S_n \in dx),$$

and note that this function is finite (in particular, the limit exists) for every functional ξ on $C([0,1])$ as shown in Sadowsky [17, p. 407]. Moreover, Λ is strictly convex whenever Λ_Z is strictly convex; this follows from Assumption 1(ii) as one easily deduces from the proof of Hölder's inequality (see, e.g., Royden [16]).

Therefore, Theorem 4.5.10 of Dembo and Zeitouni [6] shows that for all ξ and x,

$$\Lambda(\xi) = \sup_{x\in C([0,1])} [\xi(x) - I(x)] \text{ and } I(x) = \sup_{\xi\in C([0,1])^*} [\xi(x) - \Lambda(\xi)]. \tag{10}$$

Choose a sequence $\{\gamma_n\} \subset A^o$ and a point $\gamma \in \overline{A}$ such that $\gamma_n \to \gamma$ and $I(\gamma_n) \to I(\gamma)$; this is possible since A is an I-continuity set; see the proof of Proposition 4 of [9]. Due to the strict convexity of Λ and by the second identity in (10), there exists *at most* one ξ_γ such that $I(\gamma) = \xi_\gamma(\gamma) - \Lambda(\xi_\gamma)$. For all other twists ξ, we apparently have $I(\gamma) > \xi(\gamma) - \Lambda(\xi)$, and therefore

$$\inf_{x\in C([0,1])} [I(x) - \xi(x)] = -\sup_{x\in C([0,1])} [\xi(x) - I(x)] = -\Lambda(\xi) < I(\gamma) - \xi(\gamma).$$

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

$$\begin{aligned} \inf_{x\in C([0,1])} [I(x) - \xi(x)] + \inf_{x\in A^o} [I(x) + \xi(x)] &< I(\gamma) - \xi(\gamma) + \lim_{n\to\infty} [I(\gamma_n) + \xi(\gamma_n)] \\ &= I(\gamma) - \xi(\gamma) + I(\gamma) + \xi(\gamma) = 2I(\gamma), \end{aligned}$$

contradicting the necessary condition for efficient simulation in Theorem 1 of [9] (it is readily checked that the underlying assumptions are satisfied as a result of Assumption 1). This proves the first claim.

For claim (ii) we have to show that $I(\gamma_{\tilde{\tau}}) = \xi_{\tilde{\tau}}(\gamma_{\tilde{\tau}}) - \Lambda(\xi_{\tilde{\tau}})$, where $\tilde{\tau}$ is now unique. Observe that the minimizer in $\inf_{x\in A} I(x)$ is $\gamma_{\tilde{\tau}}$, and that $\xi_{\tilde{\tau}}$ is a continuous linear functional on $C([0,1])$. We first calculate $\Lambda(\xi_{\tilde{\tau}})$.

Let $\underline{\tilde{\tau}}_n := \lfloor n\tilde{\tau} \rfloor / n$, so that $\underline{\tilde{\tau}}_n \to \tilde{\tau}$ as $n \to \infty$. We then have by independence,

$$\begin{aligned}
&\int_{C([0,1])} \exp(n\alpha x(\tilde{\tau})) P(S_n \in dx) \\
&= \int_{\mathbb{R}^n} \exp\left(\alpha \sum_{i=1}^{n\underline{\tilde{\tau}}_n} z_i + \alpha(\tilde{\tau} - \underline{\tilde{\tau}}_n) z_{n\underline{\tilde{\tau}}_n + 1} \right) P_Z(dz_1) \cdots P_Z(dz_n) \\
&= \int_{\mathbb{R}} \exp(\alpha(\tilde{\tau} - \underline{\tilde{\tau}}_n)z) P_Z(dz) \left(\int_{\mathbb{R}} \exp(\alpha z) P_Z(dz) \right)^{n\underline{\tilde{\tau}}_n},
\end{aligned}$$

which should be compared with (7). Observe that $\Lambda_Z(\theta) < \infty$ for all $\theta \in \mathbb{R}$ by Assumption 1(i), and that Λ_Z is continuous due to its convexity. Consequently, the first integral of the last expression converges to one. Using the definition of $\xi_{\tilde{\tau}}$, we conclude that

$$\Lambda(\xi_{\tilde{\tau}}) = \lim_{n\to\infty} \frac{1}{n} \log \int_{C([0,1])} \exp(n\alpha x(\tilde{\tau})) P(S_n \in dx) = \tilde{\tau}\Lambda_Z(\alpha),$$

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 $\tilde{\tau}\Lambda_Z^*(e(\tilde{\tau})/\tilde{\tau}) = I(\gamma_{\tilde{\tau}})$.

We now proceed with the proof of (iii). Recall the definition of γ_τ^ϵ in (5). Observe that $\{\gamma_{\tilde{\tau}_i}^{1/n}\} \subset A^o$, and that both $\gamma_{\tilde{\tau}_i}^{1/n} \to \gamma_{\tilde{\tau}_i}$ and, as a consequence of the assumption imposed on the $\tilde{\tau}_i$, $I(\gamma_{\tilde{\tau}_i}^{1/n}) \to I(\gamma_{\tilde{\tau}_i})$. Therefore, the reasoning that established the first claim shows that the only candidate path-level twists are $\xi_{\tilde{\tau}_1}$ and $\xi_{\tilde{\tau}_2}$. It suffices to observe that these twists are unequal since $\tilde{\tau}_1 \neq \tilde{\tau}_2$. $\square$

Motivated by Lemma 1, we often assume the uniqueness of the minimizer $\tilde{\tau}$ of $\tau \mapsto \tau\Lambda_Z^*(e(\tau)/\tau)$ in the remainder of this paper. We now state the main theorem of this section.

Theorem 1 *Let Assumption 1 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 path-level twist $\xi_{\tilde{\tau}}$ defined as $\xi_{\tilde{\tau}}(x) = \alpha x(\tilde{\tau})$ is asymptotically efficient if and only if

$$\begin{aligned}
&\tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right) + \alpha e(\tilde{\tau}) \qquad (11) \\
&\leq \min\left\{ \inf_{\tau\in(0,\tilde{\tau})} \left(\tau\Lambda_Z^*\left(\frac{e(\tau)}{\tau}\right) + \inf_{\beta\in\mathbb{R}} \left[(\tilde{\tau} - \tau)\Lambda_Z^*\left(\frac{\beta - e(\tau)}{\tilde{\tau} - \tau}\right) + \alpha\beta \right] \right), \right. \\
&\qquad \left. \inf_{\tau\in(\tilde{\tau},1]} \inf_{\beta\in\mathbb{R}} \left[\tilde{\tau}\Lambda_Z^*\left(\frac{\beta}{\tilde{\tau}}\right) + (\tau - \tilde{\tau})\Lambda_Z^*\left(\frac{e(\tau) - \beta}{\tau - \tilde{\tau}}\right) + \alpha\beta \right] \right\}.
\end{aligned}$$

Proof. We prove the claim by invoking Corollary 1 of [9]. Note that the underlying assumptions hold. Hence, the path-level twist $\xi_{\tilde{\tau}}$ is asymptotically efficient if and only if

$$\inf_{x\in C([0,1])} [I(x) - \xi_{\tilde{\tau}}(x)] + \inf_{x\in A} [I(x) + \xi_{\tilde{\tau}}(x)] = 2 \inf_{x\in A} I(x). \qquad (12)$$

The rest of the proof consists of rewriting condition (12).

The first term on the left-hand side of (12) is

$$\inf_{x\in C([0,1])} [I(x) - \xi_{\tilde{\tau}}(x)] = -\Lambda(\xi_{\tilde{\tau}}) = -\tilde{\tau}\Lambda_Z(\alpha) = \tilde{\tau}\Lambda_Z^*(e(\tilde{\tau})/\tilde{\tau}) - \alpha e(\tilde{\tau}),$$

so the left-hand side of (11) equals $2\inf_{x\in A} I(x) - \inf_{x\in C([0,1])}[I(x) - \xi_{\tilde{\tau}}(x)]$. Since clearly $\tilde{\tau}\Lambda_Z^*(e(\tilde{\tau})/\tilde{\tau}) + \alpha e(\tilde{\tau}) \geq \inf_{x\in A}[I(x) + \alpha x(\tilde{\tau})]$, the condition

$$\tilde{\tau}\Lambda_Z^*(e(\tilde{\tau})/\tilde{\tau}) + \alpha e(\tilde{\tau}) \leq \inf_{x\in A} [I(x) + \alpha x(\tilde{\tau})] \qquad (13)$$

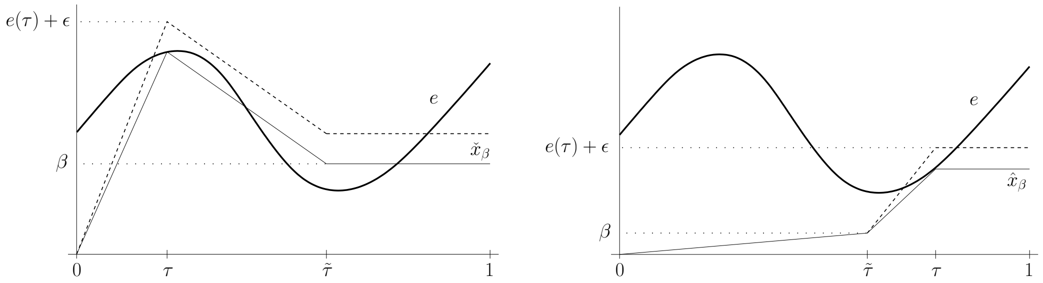

Figure 1: Two possibilities for the minimizing argument of $\inf_{x\in A}[I(x)+\alpha x(\tilde{\tau})]$.

is necessary and sufficient for asymptotic efficiency. It remains to investigate the right-hand side of this inequality.

Jensen's inequality shows that a minimizing argument of $\inf_{x\in A}[I(x)+\alpha x(\tilde{\tau})]$ is a piecewise straight line, which must exceed e in $[0,1]$, say at time τ, and has some value $\beta\in\mathbb{R}$ at time $\tilde{\tau}$. The right-hand side of (13) is the infimum over β and τ when these paths are substituted in the expression $I(x)+\alpha x(\tilde{\tau})$. Since $\tilde{\tau}\Lambda_Z^*(e(\tilde{\tau})/\tilde{\tau})+\alpha e(\tilde{\tau}) = I(\gamma_{\tilde{\tau}})+\alpha e(\tilde{\tau})$, we may assume that $\tau\neq\tilde{\tau}$ in order to derive a condition that is equivalent with (13).

Two possibilities arise. First, x can exceed e for the first time at some $\tau<\tilde{\tau}$, 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 $\tilde{\tau}$, exceeds e for some $\tau>\tilde{\tau}$, and then becomes constant. This path is denoted by $\hat{x}_{\beta,\tau}$. These two possible cases are illustrated by the solid lines in Figure 1.

It is immediate that $\check{x}_{\beta,\tau}$ satisfies for $\tau<\tilde{\tau}$

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

This corresponds to the left-hand diagram in Figure 1. The same argument shows that for $\tau>\tilde{\tau}$,

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

which finishes the proof. □

We remark that (11) can be slightly simplified using Λ_Z. Note that

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

and that the sup-term in this expression equals $\Lambda_Z(-\alpha)$ by the duality lemma (Lemma 4.5.8 of [6]). Thus, (11) is equivalent to

$$\begin{aligned}
&\tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right)+\alpha e(\tilde{\tau})\\
&\quad\leq \min\left\{\inf_{\tau\in(0,\tilde{\tau})}\left(\tau\Lambda_Z^*\left(\frac{e(\tau)}{\tau}\right)-(\tilde{\tau}-\tau)\Lambda_Z(-\alpha)+\alpha e(\tau)\right),\right.\\
&\qquad\qquad\left.\inf_{\tau\in(\tilde{\tau},1]}\inf_{\beta\in\mathbb{R}}\left[\tilde{\tau}\Lambda_Z^*\left(\frac{\beta}{\tilde{\tau}}\right)+(\tau-\tilde{\tau})\Lambda_Z^*\left(\frac{e(\tau)-\beta}{\tau-\tilde{\tau}}\right)+\alpha\beta\right]\right\}.
\end{aligned}$$

To illustrate Theorem 1, we now work out an 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 can be seen (for instance with Lemmas 3 and 4 below) that A has the required continuity properties for application of Theorem 1. Moreover, it is readily checked that $\tau\Lambda_Z^*(e(\tau)/\tau) = e(\tau)^2/(2\tau)$ is minimized for $\tilde{\tau} = 1/2$, showing that $\alpha = 2$. It is also immediate that $e(\tau)^2/(2\tau) + 2\tau - 1 + 2e(\tau)$ 'attains' its minimum value over $(0, 1/2)$ as $\tau \uparrow 1/2$. The second minimizing β in (11) is then $1/(2\tau)$, and the minimum value over $(1/2, 1]$ of the resulting function is attained for $\tau \downarrow 1/2$. Consequently, we can estimate the desired probability efficiently by path-level twisting. Therefore, this example corrects the unproven claim of Sadowsky [17, p. 408] that no path-level 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 (11) is attained for $\tau = 1$. Therefore, the *same* twist as before is now asymptotically *in*efficient.

To implement the simulation procedure, the path-level twist $\xi_{\tilde{\tau}}$ should be translated into an importance sampling distribution for $(Z_1, \ldots, Z_n)$. In view of (7), this amounts to finding the vector $\theta^{\xi,n}$; the exponentially step-level θ-twisted distribution of Z,

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

is the 'building block' for the required path-level exponential twist. Sadowsky [17] shows that the step sizes $Z_1, \ldots, Z_{\lfloor n\tilde{\tau}\rfloor}$ 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 (1). The resulting paths are samples from the path-level twisted distribution $\lambda_n^{\xi_{\tilde{\tau}}}$.

Both Lemma 1 and Theorem 1 require certain continuity properties of A. The remainder of this section is devoted to sharp conditions for these to hold.

Continuity properties of A

We start by showing that A is closed. In fact, for later use, we prove this in slightly more generality. Consider the set

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

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 $[0,1]$. We prove that A_m^M is closed, which implies that A is closed by choosing $m = e$ and $M \equiv -\infty$.

Lemma 2 *A_m^M is closed in $C([0,1])$.*

Proof. Let $\{x_n\}$ be a sequence in A_m^M converging in sup-norm to some $x \in C([0,1])$. 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 $x_n \notin A_m^M$. □

We now give a sharp condition for A to be an I-continuity set, for which we do not require uniqueness of $\tilde{\tau}$.

Lemma 3 *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 2 show that $A^o = \{x \in C([0,1]) : x(t) > e(t) \text{ for some } t \in [0,1]\}$. As A is closed, it suffices to prove that $\inf_{x\in A} I(x) = \inf_{x\in A^o} I(x)$. Let $\tilde\tau$ be such that $\tilde\tau\Lambda_Z^*(e(\tilde\tau)/\tilde\tau) = \inf_{x\in A} I(x)$. With $\gamma_{\tilde\tau}^\epsilon$ as in (5), we have $\gamma_{\tilde\tau}^\epsilon \in A^o$ and $I(\gamma_{\tilde\tau}^\epsilon) = \tilde\tau\Lambda_Z^*([e(\tilde\tau)+\epsilon]/\tilde\tau)$. 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 1(iii). By the monotone convergence theorem, $I(\gamma_{\tilde\tau}^\epsilon)$ converges to $I(\gamma_{\tilde\tau})$. □

It is also of interest to give a condition for A to be an $(I+\xi_{\tilde\tau})$-continuity set. This is the content of the next lemma, but we omit the proof since it is a variation on the ϵ-argument given in the proof of Lemma 3. The perturbed paths are drawn as the dashed lines in Figure 1. Recall the definitions of $\check x_{\beta,\tau}$ and $\hat x_{\beta,\tau}$ in the proof of Theorem 1.

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

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

(ii) *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 $(e(\overline\tau)-\overline{x}(\tilde\tau))/(\overline\tau-\tilde\tau) \in (\mathrm{dom}\,\Lambda_Z^*)^o$.*

4 Step-level twisting

This section is devoted to a simplification of the simulation scheme (i.e., the measure $\lambda_n^{\xi_{\tilde\tau}}$) studied in Section 3. The new scheme overcomes an intuitive difficulty with a path-level twisted change of measure. If 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 exceeding e after $\lceil n\tilde\tau\rceil/n$. Indeed, since the original measure P_Z is then used for sampling, e is rarely exceeded after $\lceil n\tilde\tau\rceil/n$. By the form of the estimator (8), a sample path that does not exceed e does not contribute to the resulting estimate, so that it is undesirable to have too many of such paths in the simulation.

The idea of the simplified simulation scheme is to sample every random variable Z_i from P_Z^α, *until e has been exceeded.* The simulation is then stopped and the likelihood is calculated. We refer to this setup, which has first been studied in [17], as *step-level twisting.* Note that this contrasts with path-level twisting as described in the preceding section, since there the step-size distribution is twisted up to a *fixed* twist-horizon $n\tilde\tau$. In the setting of this section, this horizon is sample-dependent.

Since both path-level twisting and step-level twisting are algorithms for estimating the same probability, it is legitimate to ask which procedure is better. To answer this question rigorously, it is our aim to develop necessary and sufficient conditions for asymptotic efficiency of step-level twisting. These conditions are the 'step analogue' of Theorem 1. A comparison of the two sets of conditions is the subject of Section 5.

Intuitively, it depends on the specific form of e if the probability of exceeding e on $[\lceil n\tilde\tau\rceil/n, 1]$ is small enough for the simplification to work. Sadowsky [17, Prop. 2] finds a sufficient condition in terms of a saddle-point inequality. The sufficient condition of Theorem 2 below improves upon this result significantly: our necessary condition is extremely 'close' to the sufficiency condition.

Throughout this section, we adopt the setup and notation of the previous section. It is worthwhile to specify the exact assumptions that we impose.

Assumption 2 *We assume that*

(i) *Assumption 1 holds,*

(ii) $\tilde{\tau} = \arg\inf \tau\Lambda_Z^*(e(\tau)/\tau)$ *is unique, and*

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

In the previous section, we have seen that this set of assumptions guarantees that Mogul′skiĭ's large-deviation principle holds, and that A is an I-continuity set, see Lemma 3. Lemma 1 shows that the uniqueness of $\tilde{\tau}$ in Assumption 2(ii) is required to have a unique twist α for the distribution of the Z_i.

The next theorem generalizes the findings of Lehtonen and Nyrhinen [13] to nonlinear boundaries e.

Theorem 2 *Let Assumption 2 hold.*

If e is lower semicontinuous, then step-level twisting is asymptotically efficient if

$$\inf_{\tau\in(0,1]} \left[\tau\Lambda_Z^*\left(\frac{e(\tau)}{\tau}\right) + \alpha e(\tau) - \tau\Lambda_Z(\alpha)\right] \geq 2\tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right). \tag{14}$$

Conversely, suppose that e is upper semicontinuous. If step-level twisting is asymptotically efficient, then

$$\inf_{\{\tau\in(0,1]:\, e(\tau)/\tau\in(\mathrm{dom}\,\Lambda_Z^*)^o\}} \left[\tau\Lambda_Z^*\left(\frac{e(\tau)}{\tau}\right) + \alpha e(\tau) - \tau\Lambda_Z(\alpha)\right] \geq 2\tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right). \tag{15}$$

Proof. As seen in [9], since A is an I-continuity set, asymptotic efficiency is equivalent to

$$\limsup_{n\to\infty} \frac{1}{n} \log E_n^{(2)} \leq -2 \inf_{x\in A} I(x),$$

where $E_n^{(2)}$ denotes the second moment of the estimator.

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 $f : [0,1] \to [-\infty,\infty)$ as

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

We also set for $\tau \in (0,1]$,

$$\tilde{A}_\tau := \{x \in C([0,1]) : x(t) < e(t) \text{ for } t \in [0,\tau), x(\tau) \geq e(\tau)\},$$

i.e., $\tilde{A}_\tau$ are the paths that exceed 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 step-level-twisting procedure are in general no elements of $C([0,1])$, since the simulation is stopped at some random time, not necessarily at time 1. To overcome this, note that stopping a simulation run amounts to 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 exceeding e. Therefore, the distribution in $C([0,1])$ of sample paths generated by step-level twisting is well-defined; we denote it by μ_n. The 'original' distribution of S_n in $C([0,1])$ is denoted by ν_n. One can readily check that on $\tilde{A}_\tau$, we have (for x in the support of ν_n)

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

In the proof of the sufficient condition, we use the function $\zeta : C([0,1]) \to [-\infty, \infty)$ 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, it is quite fruitful to use some ideas of its proof (see Theorem 4.3.1 in Dembo and Zeitouni [6]).

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. We now prove that g is lower semicontinuous. For this, 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 nondecreasing on $[0,\infty)$ (Z_1 is centered), we observe that

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

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

$$\inf_{\tau\in\overline{T_t}} g(\tau) \geq g(t) - \epsilon \quad \text{and} \quad \sup_{\tau\in\overline{T_t}} f(\tau) \leq f(t) + \epsilon. \tag{16}$$

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 [6])

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

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 (16):

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

Although $\bigcup_{\tau\in T_{t_i}} \tilde{A}_\tau$ is in general not closed, it is a subset of $\{x : x(t) \geq e(t) \text{ for some } t \in \overline{T_{t_i}}\}$. This set is closed by Lemma 2 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 (16),

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

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 sufficient 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 step-level twisting is asymptotically efficient, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_A \frac{d\nu_n}{d\mu_n} d\nu_n \leq -2\tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right). \tag{17}$$

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

$$\sup_{\tau\in(t-\delta,t]} e(\tau) \leq e(t) + \epsilon. \tag{18}$$

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 2 is closed. Indeed, 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 (18) for the last inequality.

Recall the definition of γ_τ and γ_τ^ϵ in (3) and (5). Now two cases are distinguished.

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 (18) is met, and set

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

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 $e(t)/t \in (\text{dom}\,\Lambda_Z^*)^o$.

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 $e(t)/t \in (\text{dom}\,\Lambda_Z^*)^o$,

$$\alpha\frac{e(t)}{t} - \Lambda_Z(\alpha) + \Lambda_Z^*\left(\frac{e(t)}{t}\right) \geq 0. \tag{19}$$

We work towards a contradiction by supposing that (19) 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 step-level twisting 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

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

Hence, the infimum in (15) is not attained by t for which γ_t intersects with e before t.

Therefore, if step-level twisting is asymptotically efficient, we must have by (17)

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

which proves the claim. □

As a result of the sufficient condition in Theorem 2, step-level twisting 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 is Sadowsky's sufficient condition [17].

5 A comparison

In Theorems 1 and 2, we have provided necessary and sufficient conditions for asymptotic efficiency of path-level twisting and step-level twisting respectively. It is our present aim to compare these conditions, and we start by showing that the conditions must be different.

Consider the example given on page 9, 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 (14) is attained at $\tau = 1/2$, implying that step-level twisting is asymptotically efficient. Note that path-level twisting was not asymptotically efficient.

This raises the question how the conditions for the two methods are related. The following corollary is of practical interest. Informally, it entails that path-level efficiency implies step-level efficiency. In other words, comparing the conditions for path-level efficiency and those for step-level efficiency, the conditions for step-level efficiency are the weaker.

Corollary 1 *Condition (11) for path-level efficiency implies both the sufficient condition (14) and the necessary condition (15) for step-level efficiency.*

Proof. Since the sufficient condition (14) implies the necessary condition (15), it suffices to compare (11) and (14). The first step is to note that $\tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right) > 0$ and that

$$\alpha e(\tilde{\tau}) - \tilde{\tau}\Lambda_Z(\alpha) = \tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right),$$

so that (14) is equivalent to

$$\begin{aligned}
&\tilde{\tau}\Lambda_Z^*\left(\frac{e(\tilde{\tau})}{\tilde{\tau}}\right) + \alpha e(\tilde{\tau}) \\
&\quad \leq \min\left\{ \inf_{\tau\in(0,\tilde{\tau})}\left(\tau\Lambda_Z^*\left(\frac{e(\tau)}{\tau}\right) + \alpha e(\tau) + (\tilde{\tau}-\tau)\Lambda_Z(\alpha)\right), \right. \\
&\qquad\qquad \left. \inf_{\{\tau\in(\tilde{\tau},1]:\, \alpha e(\tau)-\tau\Lambda_Z(\alpha)\geq 0\}}\left(\tau\Lambda_Z^*\left(\frac{e(\tau)}{\tau}\right) + \alpha e(\tau) - (\tau-\tilde{\tau})\Lambda_Z(\alpha)\right)\right\}.
\end{aligned} \tag{20}$$

We now prove that the right-hand side of (11) cannot exceed the right-hand side of (20).

Clearly, the last term in the minimum of (20) cannot be smaller than

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

Since $\Lambda_Z(\alpha) \geq 0$ as a result of the fact that Z_1 has zero mean, this immediately yields that the right-hand side of (20) cannot be smaller than

$$\min\left\{\inf_{\tau\in(0,\tilde{\tau}]}\left(\tau\Lambda_Z^*\left(\frac{e(\tau)}{\tau}\right) + \alpha e(\tau)\right), \inf_{\tau\in(\tilde{\tau},1]}\left(\tau\Lambda_Z^*\left(\frac{e(\tau)}{\tau}\right) + \alpha\tilde{\tau}\frac{e(\tau)}{\tau}\right)\right\}.$$

To see that the right-hand side of (11) does not exceed this quantity, choose $e(\tau)$ and $\tilde{\tau}e(\tau)/\tau$ for the first and second β in (11) respectively. □

References

[1] S. Asmussen, *Ruin probabilities*, World Scientific Publishing Co. Inc., 2000.

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

[3] S. Asmussen and R. Y. Rubinstein, *Steady state rare events simulation in queueing models and its complexity properties*, in Advances in queueing, J. Dshalalow, ed., CRC, Boca Raton, FL, 1995, pp. 429–461.

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

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

[6] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer, New York, second ed., 1998.

[7] J.-D. Deuschel and D. W. Stroock, *Large deviations*, Academic Press Inc., Boston, MA, 1989.

[8] A. B. Dieker and M. Mandjes, *Fast simulation of overflow probabilities in a queue with Gaussian input*, Tech. Rep. PNA-E0404, CWI, the Netherlands, 2004.

[9] ———, *On asymptotically efficient simulation of large deviation probabilities*, Adv. in Appl. Probab., 37 (2005), pp. 539–552.

[10] P. Dupuis and H. Wang, *Importance sampling, large deviations, and differential games*, Stoch. Stoch. Rep., 76 (2004), pp. 481–508.

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

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

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

[14] A. A. Mogul′skiĭ, *Large deviations for the trajectories of multidimensional random walks*, Theory Probab. Appl., 21 (1976), pp. 309–323.

[15] ———, *Large deviations for processes with independent increments*, Ann. Probab., 21 (1993), pp. 202–215.

[16] H. L. Royden, *Real analysis*, The Macmillan Co., New York, 1968.

[17] J. S. Sadowsky, *On Monte Carlo estimation of large deviations probabilities*, Ann. Appl. Probab., 6 (1996), pp. 399–422.