Surveys in Operations Research and Management Science 17 (2012) 38–59

Contents lists available at SciVerse ScienceDirect

Surveys in Operations Research and Management Science

journal homepage: www.elsevier.com/locate/sorms

Review

State-dependent importance sampling for rare-event simulation: An overview and recent advances

Jose Blanchet [a,*], Henry Lam [b]

[a] *Columbia University, United States*

[b] *Boston University, United States*

ARTICLE INFO

Article history:
Received 7 February 2011
Received in revised form
19 September 2011
Accepted 20 September 2011

ABSTRACT

This paper surveys recent techniques that have been developed for rare-event analysis of stochastic systems via simulation. We review standard (state-independent) techniques that take advantage of large deviations results for the design of efficient importance sampling estimators. Classical examples and counter-examples are discussed to illustrate the reach and limitations of the state-independent approach. Then we move to state-dependent techniques. These techniques can be applied to both light and heavy-tailed systems and are based on subsolutions (see e.g. Dupuis and Wang (2004) [5], Dupuis and Wang (2007) [6], Dupuis and Wang (2009) [80], Dupuis et al. (2007) [7]) and Lyapunov bounds (Blanchet and Glynn (2008) [9], Blanchet et al. (2007) [11], Blanchet (2009) [12]). We briefly review the ideas behind these techniques, and provide several examples in which they are applicable.

Contents

1. Introduction

In recent years, there has been a substantial amount of research related to the efficient design of state-dependent importance sampling estimators. This research has been motivated by a series of examples and counter-examples (see [1,2]) related to the use of large deviations principles in the design of efficient importance sampling estimators for light-tailed systems. The development of efficient importance sampling estimators for heavy-tailed systems, which as explained in [3] must often be state-dependent, has also driven the research developments that are the focus of this survey (the paper [4] discusses additional challenges that arise

* Corresponding author.
E-mail address: jose.blanchet@gmail.com (J. Blanchet).

1876-7354/$ – see front matter © 2011 Published by Elsevier B.V.
doi:10.1016/j.sorms.2011.09.002

in the context of efficient rare-event simulation for heavy-tailed systems).

In order to systematically address the construction of state-dependent importance sampling estimators that can be rigorously shown to be efficient in large deviations environments for light-tailed systems, Dupuis, Wang, and their students have developed a method based on control theory and the use of subsolutions of an associated Isaacs equation for constructing and testing asymptotically optimal importance sampling estimators [5–8]. A related approach, based on the construction of Lyapunov inequalities, has been used in [9–11] for the construction and analysis of state-dependent importance sampling estimators for heavy-tailed systems. The Lyapunov method has also been used for counting bipartite graphs with a given degree sequence in [12].

In this paper we summarize the ideas behind the development of state-dependent importance sampling estimators and illustrate their applications in some simple examples. However, in order to put these recent ideas in context we will review some of the basic notions and techniques that motivated the development of state-dependent samplers and their efficiency analysis.

We shall start by discussing standard notions of statistical efficiency that are used in order to rigorously test the optimality of a simulation estimator. This will be done in Section 2. As we shall see, when the cost per replication of the estimator is appropriately controlled, these notions guarantee a substantial improvement, in terms of running time, relative to crude Monte Carlo evaluation—which typically offers a natural general benchmark. The introduction of these notions of efficiency will naturally lead to the question of how to actually construct importance sampling estimators that can be rigorously shown to be efficient. The zero-variance change-of-measure, while virtually impossible to implement in most situations of interest, provides a good guidance in order to construct such estimators. It is very important to emphasize, as we shall see, that importance sampling allows to efficiently estimate conditional expectations given the event of interest.

The zero-variance change-of-measure turns out to be the conditional distribution of the underlying process given the rare event of interest. So, it is natural to leverage-off the available asymptotic theory for rare events in order to understand such a conditional distribution. This is basically the departing point in the design of efficient importance sampling estimators for rare events, especially in the setting of light-tailed systems, which will be our main focus in the first part of the paper and which encompasses most classical models such as Jackson networks. Regardless of whether one is studying systems with light or heavy tails, it is extremely important, however, to note that once an efficient importance sampling estimator is in place, then one obtains a computational mechanism (via Monte Carlo) that allows to enhance a given asymptotic result by providing an estimate (obtained by averaging iid replications) whose accuracy can be improved to any desired level of relative precision by increasing the sample size. In turn, the guaranteed efficiency will typically ensure that the number of replications will remain under control as the event of interest becomes more and more rare according to the underlying large deviations parameter.

In Section 3 we will review some of the earlier ideas in the literature showing the basic connection between large deviations theory and the design of efficient importance sampling estimators. As an illustrative example, we will study the sample-path large deviations of random walk, how it suggests a conditional path and how it leads to an efficient state-independent exponential tilting scheme for a first passage problem.

The successful application of large deviations theory in the design of efficient importance sampling estimators is not straightforward and this feature turned out to be the norm rather than the exception, especially in Operations Research applications such as the analysis of queueing networks. This might be somewhat surprising given the fact that large deviations theory allows to characterize the asymptotic conditional distribution (i.e. the zero-variance change-of-measure) leading to the occurrence of the rare event. The problem, as we shall see, arises from the fact that such characterization is done in fluid scale (i.e. in the spatial and temporal scales corresponding to the Law of Large Numbers) and there are many equivalent changes-of-measure that possess the same Law of Large Numbers limit and that coincide with that of the zero-variance change-of-measure. In Section 3 we will first illustrate with an example the successful application of direct large deviations results and then study in detail counter-examples that flesh-out precisely the issues that can go wrong in the variance control of importance sampling estimators.

Our discussion in Section 3 then will lead us to introduce the approach proposed by Dupuis and Wang, which formulates the design of state-dependent importance sampling estimators in terms of a stochastic control problem in which the value function is the second moment of the estimator. Taking an asymptotic limit, as we shall see in Section 4, the stochastic control problem becomes a deterministic control problem. The simplification comes from the fact that only a subsolution is needed for the design of an efficient sampling scheme. A trivial supersolution is given, thanks to Jensen's inequality, by twice the underlying large deviations rate function. The whole point of Dupuis and Wang's approach is to construct a subsolution that matches the trivial supersolution at one point, namely, the initial condition of the system under consideration. In many problems of interest it turns out that such a subsolution is piecewise affine and therefore relatively easy to manipulate.

We will then move to the analysis of heavy-tailed systems, which will be introduced in Section 5. These types of systems arise in the analysis of finance and insurance models. There are fundamental qualitative differences between light and heavy-tailed systems when it comes to the analysis of rare events. This is why it is incorrect to use a model with light-tailed components to study large deviations properties of systems that are built from heavy-tailed building blocks. Light-tailed large deviations occur gradually, while heavy-tailed large deviations occur suddenly and this feature manifests itself in the design of efficient importance sampling estimators for heavy-tailed systems. This difference is illustrated by, for example, the problem of estimating the first passage probability for a random walk i.e. $P(\tau_n < \infty)$, where $\tau_n = \inf\{k \geq 0 : S_k > n\}$, $S_k = X_1 + \cdots + X_k$ and X_i's are iid increments with negative mean. If X_i has exponential moments i.e. $E\exp(\theta X_i) < \infty$ for any $\theta \in \mathbb{R}$, then the most likely way for the first passage to occur is by small gradual contribution from each increment to drift upwards. This translates into an exponential tilting for each increment until τ_n is hit. On the other hand, if X_i is regularly varying (i.e. basically having a tail distribution function with power-law decay) then the large deviations event occurs typically by a big jump in one of the increments, whereas all the other increments behave like the nominal (unconditional) distribution (i.e. with the same unconditional negative mean under fluid scale). One efficient importance sampling scheme in this case will be a mixture algorithm that recognizes a suitably chosen probability for a sudden jump to occur. This first passage problem will be visited in Example 1 (for the light-tailed case) and Example 5 (for the regularly varying case and beyond).

Throughout the paper we use Landau's notation for the asymptotic behavior of functions. In particular, given two non-negative functions $(f(N) : N \geq 1)$ and $(g(N) : N \geq 1)$ we say that $f(N) = O(g(N))$ if there exists a constant $c \in (0, \infty)$ independent of N such that $f(N) \leq cg(N)$ for $N \geq 1$. We write $f(N) = \Omega(g(N))$ if $f(N) \geq c'g(N)$ for some constant $c' \in (0, \infty)$ independent

of N. We write $f(N) = \Theta(g(N))$ if both $f(N) = O(g(N))$ and $f(N) = \Omega(g(N))$. Finally we say that $f(N) = o(g(N))$ as $N \to \infty$ if $f(N)/g(N) \to 0$ as $N \to \infty$.

We emphasize that this paper aims to introduce readers the key concepts and ideas in rare-event problems. Throughout the exposition we may drop some mathematical details and highlight only the central arguments, pointing readers to other references. The rest of the paper is organized as follows. In Section 2 we introduce the basic notions, criteria and guidelines for rare-event simulation estimators. In Section 3, we discuss the classical large deviations theory for light-tailed systems from a sample path perspective, and how it leads to the idea of exponential tilting in importance sampling. There we will also point out some issues of this classical idea through examples. Section 4 will then be devoted to the recent developments of state-dependent techniques that aim to fix these issues, using a subsolution approach as discussed briefly above. Lastly, Section 5 focuses on heavy-tailed systems, including large deviations results, classical estimators using conditional Monte Carlo and hazard rate twisting, and the recent work on state-dependent importance samplers via weak convergence and Lyapunov technique.

2. Notions on efficiency and importance sampling techniques

The central question that we study consists in estimating via simulation $P(A)$, where A is some rare event of interest, which implies that $P(A) \approx 0$. For instance, the event A might correspond to the probability of bankruptcy within a given time horizon of a financial or insurance company. A related question consists in computing conditional expectations of certain quantities of interest given that A occurs. For instance, what is the mean value of the deficit at the time of ruin or, in the case of companies that pay dividends to their shareholders, what is the distribution of the net present value of the dividends paid up to the time of ruin. In this section we discuss notions that allow to quantify the efficiency of rare event simulation estimators and review basic properties of importance sampling estimators for rare events.

2.1. Notions of efficiency

In order to design algorithms that achieve high accuracy in relative terms and that are designed to perform well in rare event settings, a typical approach is to embed the question of interest in a suitable asymptotic regime, often motivated by an applied standpoint. For example, in the setting of bankruptcy of an insurance company, it is often the case that the capital or reserve of the company is large relative to individual claim sizes, and this is basically why one might expect the probability of bankruptcy, $P(A)$, to be small. So, one might introduce a parameter, say n, for instance as the initial capital in the insurance example, and consider $\alpha_n = P(A_n)$ as a function of the parameter n which is often referred to as the "rarity parameter". From now on we will assume that underlying the rare event there is a rarity parameter n such that $\alpha_n \longrightarrow 0$ as $n \longrightarrow \infty$.

The idea is then to design a simulation estimator whose error is controlled in relative terms as $n \nearrow \infty$ compared to crude (or naive) Monte Carlo, which provides an obvious alternative and, therefore, a general benchmark. Note that crude Monte Carlo involves simulating N iid Bernoulli's, $I_n(1), \ldots, I_n(N)$, with success parameter α_n, thereby producing an estimator of the form

$$\widehat{\alpha}_n^c(N) = \frac{1}{N}\sum_{j=1}^{N} I_n(j). \tag{1}$$

The super-index "c" that appears in $\widehat{\alpha}_n^c(N)$ corresponds to crude Monte Carlo. Since $\widehat{\alpha}_n^c(N)$ is an unbiased estimator of α_n, its mean squared error equals $\mathrm{Var}(\widehat{\alpha}_n^c(N))^{1/2} = \alpha_n^{1/2}(1-\alpha_n)^{1/2}/N^{1/2}$. In turn, we obtain a relative mean squared error equal to the coefficient of variation of $\widehat{\alpha}_n^c(N)$, namely,

$$CV(\widehat{\alpha}_n^c(N)) := \frac{(1-\alpha_n)^{1/2}}{\alpha_n^{1/2}N^{1/2}}.$$

In order to control the relative error of $\widehat{\alpha}_n^c(N)$ (i.e. $CV(\widehat{\alpha}_n^c(N))$ remains bounded as $n \nearrow \infty$), one needs $N = \Omega(1/\alpha_n)$ Bernoulli's.

The problem with the estimator $\widehat{\alpha}_n^c(N)$ is that each of the underlying replications, the $I_n(j)$'s, has a very big variance in relative terms. The number of replications N must grow at the same rate as $CV(I_n(j))^2 = (1-\alpha_n)/\alpha_n$. The same phenomenon occurs in any estimator obtained out of averaging iid replications, as in (1). The objective is then to design an estimator R_n with a controlled mean squared error. We concentrate only on unbiased estimators, so controlling the behavior of the mean squared error boils down to controlling the behavior of the second moment. The overall estimate is then obtained by averaging N iid copies of R_n. Motivated by these considerations we have the following definitions.

Definition 1. An estimator R_n is said to be *strongly efficient* if

$$ER_n^2 = O\left(\alpha_n^2\right)$$

as $n \nearrow \infty$.

Definition 2. R_n is said to be *weakly efficient* or *asymptotically optimal* if for each $\varepsilon > 0$ we have that

$$ER_n^2 = O\left(\alpha_n^{2-\varepsilon}\right)$$

as $n \nearrow \infty$.

Definition 3. We say that R_n has *polynomial complexity* of order at most $l \geq 0$ if

$$ER_n^2 = O\left[\alpha_n^2(\log(1/\alpha_n))^{2l}\right], \tag{2}$$

or in other words, if its relative mean squared error (or coefficient of variation) grows at most at a polynomial rate with degree l in $\log(1/\alpha_n)$. The estimator is *strongly efficient* if $l = 0$ in (2).

In most cases the analysis of importance sampling estimators (especially state-dependent samplers in light-tailed cases) concludes only weak efficiency of the estimators. Another class of importance sampling changes-of-measure that have been proposed recently exhibits asymptotically negligible relative mean squared error, defined next (see [13]):

Definition 4. The estimator R_n exhibits *asymptotically negligible*relative error if given any $\varepsilon > 0$ one has that

$$\overline{\lim_{n\to\infty}} \frac{ER_n^2}{\alpha_n^2} \leq 1 + \varepsilon,$$

or in other words, if its relative mean squared error (or coefficient of variation) is less than ε.

In order to see the gain in efficiency, suppose that an estimator R_n has polynomial complexity of order at most l and consider $\widetilde{\alpha}_n(N) = \sum_{i=1}^{N} R_n(i)/N$ where the $R_n(i)$'s are independent replications of R_n. It follows from Chebyshev's inequality that at most $N = O[\varepsilon^{-2}\delta^{-1}\log(1/\alpha_n)^{2l}]$ replications are required to conclude that $\widetilde{\alpha}_n(N)$ is ε-close to α_n in relative terms with at least $1-\delta$ confidence (i.e. $P(|\widetilde{\alpha}_n(N) - \alpha_n| > \varepsilon\alpha_n) \leq \delta$). Observe that in some sense, assuming that the cost per replication of $R_n(j)$ is comparable to that of $I_n(j)$ (in the setting of crude Monte Carlo), *polynomially efficient estimators provide an exponential gain in efficiency relative to crude Monte Carlo.*The paper of [14] provides

further study and examples related to each of the aforementioned classes of estimators.

Besides variance control, it is worth pointing out other measures for importance sampling estimators, and we refer interested readers to the following references. One criterion is the robustness of higher moment estimates as the rarity parameter goes to zero, as discussed in [15]. In the special case of second moment, [16,17] introduce the notion of bounded normal approximation, suggested by Berry–Esseen theorem, in the context of highly reliable Markovian systems. Related diagnostic robustness check of confidence interval coverage is discussed in [18] with so-called coverage functions. Secondly, recall that our criteria in this section all presume that the costs of replication are comparable between importance sampler and crude Monte Carlo. This might not always be true, i.e. the cost of generating each sample from importance sampling may increase disproportionately with the rarity parameter. The paper [19] discusses this issue and suggests the use of so-called relative efficiency. The paper [20] justifies theoretically how to include computational effort per sample rightly into efficiency measures. Finally, [21] provides a survey of the above robustness issues in rare-event estimators.

2.2. Basic properties of importance sampling estimators

Importance sampling is a variance reduction technique that is often applied to design an estimator, R_n, with efficiency characteristics such as those described by Definitions 1–4. We shall review basic properties of importance sampling next; for a more detailed discussion see for example [22–24]. The basic idea is to introduce a probability measure $\widetilde{P}(\cdot)$ such that the likelihood ratio or Radon–Nikodym derivative between the nominal (original) probability measure, $P(\cdot)$, and $\widetilde{P}(\cdot)$ is well defined on the event A_n of interest. Then we can simply let

$$R_n(\omega) = \frac{\mathrm{d}P}{\mathrm{d}\widetilde{P}}(\omega) I(\omega \in A_n).$$

As usual, we use ω to denote the underlying random outcome which is simulated according to the probability measure $P(\cdot)$. We use $\widetilde{E}(\cdot)$ to denote the expectation operator associated to the probability measure $\widetilde{P}(\cdot)$. Note that

$$\widetilde{E}R_n = \int_{A_n} \frac{\mathrm{d}P}{\mathrm{d}\widetilde{P}}(\omega)\,\mathrm{d}\widetilde{P}(\omega) = P(A_n) = \alpha_n,$$

so R_n is an unbiased estimator, and

$$\widetilde{E}(R_n^2) = \int_{A_n} \left(\frac{\mathrm{d}P}{\mathrm{d}\widetilde{P}}(\omega)\right)^2 \mathrm{d}\widetilde{P}(\omega) = \int_{A_n} \frac{\mathrm{d}P}{\mathrm{d}\widetilde{P}}(\omega)\,\mathrm{d}P(\omega) = E(R_n). \tag{3}$$

In principle one can design an importance sampling estimator with zero variance. Indeed, if we let

$$\widetilde{P}(\cdot) = P_n^*(\cdot) := P(\cdot \mid A_n),$$

then

$$R_n^* := \frac{\mathrm{d}P}{\mathrm{d}P_n^*}(\omega) I(\omega \in A_n) = P(A_n),$$

which is clearly an exact estimate of $\alpha_n = P(A_n)$. Of course, assuming that one has access to $P^*(\cdot)$ defeats the point of using simulation to estimate $P(A_n)$. However, the underlying lesson behind the characterization of the zero-variance change-of-measure as the conditional distribution given the rare event of interest is that one can use asymptotic theory to describe $P_n^*(\cdot)$ as $n \nearrow \infty$. As we shall review in the next section, this is often the starting point in the design of efficient importance sampling estimators.

3. On large deviations techniques and importance sampling

In order to explain the interplay between large deviations techniques and the design of importance sampling estimators we shall concentrate on models whose fundamental building blocks are given by random walks. Many questions of interest in Operations Research can be posed in terms of sample-path properties of random walks; we will see some examples in the sequel. In order to guide our discussion we shall concentrate on several specific instances underlying a generic random walk $S = (S_k : k \geq 1)$, defined via

$$S_k = X_1 + \cdots + X_k,$$

where the X_i's are iid random variables. We assume that $\psi(\theta) = \log E \exp(\theta X_i)$ is finite in a neighborhood of the origin over the real axis. We use $\dot{\psi}(\cdot)$ to denote the derivative of $\psi(\cdot)$. Given any parameter $a \in R$, we will also assume that a solution θ_a to a root equation such as $\dot{\psi}(\theta_a) = a$ exists whenever we require such a solution. The role of the parameter a will depend on the context of the examples and applications that we shall discuss.

3.1. Large deviations results for random walks

We shall take advantage of the following fundamental result in the theory of large deviations, known as Mogulskii's theorem (see for instance [25, p. 176 Theorem 5.1.2]).

Theorem 1. *Consider the continuous approximation* $Y_n(t) = S_{\lfloor nt \rfloor}/n + (t - \lfloor nt \rfloor/n) X_{\lfloor nt \rfloor + 1}$ *obtained by joining the values of the random walk at the lattice points and by scaling space by* $1/n$ *and time by n. Consider the space* $C[0, T]$ *for any* $T > 0$*, with the topology generated by the uniform norm and let us write* $AC[0, T]$ *to denote the space of absolutely continuous function in* $[0, T]$*. Then, for any closed set C*

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log P(Y_n(\cdot) \in C) \leq -\inf\{I(x(\cdot)) : x(\cdot) \in AC[0, T] \cap C, x(0) = 0\},$$

and for any open set O

$$\underline{\lim_{n\to\infty}} \frac{1}{n} \log P(Y_n(\cdot) \in O) \geq -\inf\{I(x(\cdot)) : x(\cdot) \in AC[0, T] \cap O, x(0) = 0\},$$

where

$$I(x(\cdot)) = \int_0^T J(\dot{x}(s))\,\mathrm{d}s$$

and

$$J(z) = \sup_{\theta \in \mathbb{R}} [\theta z - \psi(\theta)].$$

Remark 1. Given a set A we often write

$$I(A) = \inf\{I(x(\cdot)) : x(\cdot) \in AC[0, T] \cap A,\ x(0) = 0\}.$$

The previous result provides rigorous support for the heuristic approximation $P(Y_n(\cdot) \approx y) \approx \exp(-nI(y))$. In most applications in Operations Research, and in particular in Examples 1 and 2 to be discussed below, we are interested in a set A such that

$$P(Y_n(\cdot) \in A) = \exp(-nI(A) + o(n)). \tag{4}$$

That is, sets for which the limits in the previous theorem coincide when applied to the interior and the closure of A.

Now, the solution $x^*(\cdot)$ to the calculus of variations problem defining $I(A)$ is often unique and describes the law of large numbers of $Y_n(\cdot)$ given that $Y_n(\cdot) \in A$ and in this sense we

say that $x^*(\cdot)$ approximates the conditional distribution of $Y_n(\cdot)$ given $Y_n(\cdot) \in A$. This is easy to see using the earlier heuristic approximation (a rigorous verification is also almost direct and can be found in [26, p. 61 Lemma 4.2]). Indeed, note that if $x^* \neq y \in A$, then

$$P\left(Y_n(\cdot) \approx y | Y_n(\cdot) \in A\right) \approx \exp\left(-nI(y) + nI(A)\right) = \exp\left(-nI(y) + nI\left(x^*\right)\right) \longrightarrow 0$$

as $n \to \infty$, since x^* is the unique optimizer of $I(A)$.

Given that x^* approximates the conditional distribution of $Y_n(\cdot)$ given the event of interest it makes sense to bias the paths to track x^*. In order to do this one uses a very particular importance sampling mechanism called exponential tilting, which is implemented increment-by-increment in the simulation via

$$P_{\theta_k}(X_k \in \mathrm{d}x) = \exp(\theta_k x - \psi(\theta_k)) P(X_k \in \mathrm{d}x). \tag{5}$$

If X_k has a density with respect to the Lebesgue measure, say $f(\cdot)$, then $P(X_k \in \mathrm{d}x) = f(x)\,\mathrm{d}x$. The notation adopted in the previous display is simply a mechanism that allows us to consider general distributions (including discrete or mixtures of discrete and continuous distributions). We will use $E_{\theta_k}(\cdot)$ to denote the expectation operator associated to $P_{\theta_k}(\cdot)$. The parameter θ_k might depend on $(X_1, \ldots, X_{k-1})$ but it is not allowed to depend on future observations. For example, if under $P(\cdot)$, X_k is standard Gaussian, then under $P_{\theta_k}(\cdot)$, X_k is normal with mean θ_k and unit variance. On the other hand, if under $P(\cdot)$, X_k is exponentially distributed with unit mean, then under $P_{\theta_k}(\cdot)$ (for $\theta_k < 1$), X_k is exponentially distributed with mean $1/(1-\theta_k)$.

In order to use exponential tilting to track the optimal path $x^*(\cdot)$ it is natural to select for the k-th increment a value θ_k such that

$$E_{\theta_k} X_k = \dot{x}^*(k/n). \tag{6}$$

Since

$$\log E_{\theta_k} \exp(\eta X_k) = \log E \exp((\eta + \theta_k) X_k - \psi(\theta_k)) = \psi(\eta + \theta_k) - \psi(\theta_k),$$

we conclude that $E_{\theta_k} X_k = \dot{\psi}(\theta_k)$ and therefore Eq. (6) is equivalent to

$$\dot{\psi}(\theta_k) = \dot{x}^*(k/n). \tag{7}$$

Note that selecting θ_k via (7) gives rise to a *state-independent* importance sampling estimator because the simulation of the k-th increment does not depend on the position S_k of the random walk.

A natural question that might arise in the reader's mind is why shall we use exponential tilting as a biasing mechanism? One could think, for instance, to simply shift the means of the increments. This, as we saw earlier, is equivalent to exponential tilting in the Gaussian case, but it is *not so* in the case of exponential random variables. Moreover, in the later case, the change-of-measure induced by shifting might not be admissible because one could violate absolute continuity (the supports of the distributions might not even match) and the likelihood ratio might not be even well defined. The fundamental reason behind the use of exponential tilting arises from entropy considerations. Recall that the entropy between two probability measures $P(\cdot)$ and $\widetilde{P}(\cdot)$ is defined via

$$H(\widetilde{P} \parallel P) = \widetilde{E}\left[\log\left(\frac{\mathrm{d}\widetilde{P}}{\mathrm{d}P}\right)\right]. \tag{8}$$

Note that $H(\widetilde{P} \parallel P) \geq 0$ and $H(\widetilde{P} \parallel P) = 0$ if and only if $P = \widetilde{P}$, so $H(\cdot)$ serves as a measure of "discrepancy" between two probability measures. Exponential tilting arises as a natural biasing mechanism because, in connection to (6), the solution to the optimization problem

$$\min\{H(\widetilde{P} \parallel P) : \widetilde{P}(\cdot) \text{ such that } \widetilde{E} X_k = \dot{x}^*(k/n)\}$$

is precisely given by $\widetilde{P}(\cdot) = P_{\theta_k}(\cdot)$. A proof of this fact can be found in [27] p. 409 and 421. The reader might wish to verify this fact at least in the case of distributions with finite support, in which case is an elementary exercise in convex optimization.

Another reason that motivates the use of exponentially tilted distributions is that one can sometimes show that the distribution of $(X_1, \ldots, X_k)$ given a large deviations event of interest converges to the optimal exponential tilting (i.e. selecting each tilting parameter according to (7)); see for instance [23] Chapter VI Section 5 and [28].

3.2. Applications of large deviations to rare-event simulation

We are now ready to apply the approach discussed in the previous subsection to a couple of examples. Our goal is to investigate the extent to which the tracking of the optimal path via state-independent exponential tilting allows us to obtain a provably efficient importance sampling estimator. Example 1 illustrates a situation in which weak efficiency is obtained and, as we shall see, Example 2 shows an instance in which *no* efficiency is guaranteed.

The first example is classical in insurance and is of significant historical importance in rare-event simulation because it was among the first problems for which an explicit connection to large deviations theory was used in order to construct and analyze efficient importance sampling estimators (see [29]).

Example 1 (*Using Large Deviations For Efficient Importance Sampling*). Suppose that $S_0 = 0$ and let $\tau_n = \inf\{k \geq 0 : S_k > n\}$. Assume that $EX_k < 0$ and consider the problem of computing

$$\alpha_n = P(\tau_n < \infty)$$

as $n \nearrow \infty$. In order to motivate this example note that the surplus process of a insurance company at time $t > 0$ in continuous time can be represented via the process

$$Y(t) = n + pt - \sum_{j=1}^{N(t)} Z_j, \tag{9}$$

where $N(\cdot)$ represents a renewal process that counts the number of arrivals up to time t, $p > 0$ is a premium rate paid continuously in time, $n := Y(0)$ is the initial surplus, and the Z_j's represent iid claim sizes—which are assumed to be positive and independent of the process $N(\cdot)$. Eventual ruin, namely the event that $Y(\cdot)$ hits a level below zero at some point into the future, can only occur at arrival times. Therefore, if A_k is the arrival of the k-th claim into the system then eventual ruin occurs if and only if there exists $k \geq 1$ such that $Y(A_k) < 0$. In particular, if we write $A_k = T_1 + \cdots + T_k$, where the T_j's are iid inter-arrival times, then ruin happens if and only if $\tau_n < \infty$, where $X_j = Z_j - pT_j$.

We now compute the optimal path introduced in Theorem 1, which involves solving the calculus of variation problem. Recall that we are assuming $EX_k = \dot{\psi}(0) < 0$. We also will assume that there exists $\theta^* > 0$ such that $\psi(\theta^*) = 0$ and $\dot{\psi}(\theta^*) < \infty$. We are naturally assuming that X_k is non-deterministic (otherwise $\alpha_n = 0$), therefore $\psi(\cdot)$ is strictly convex and θ^* is unique (see [30, p. 352]). Moreover, also by convexity we have that $\dot{\psi}(\theta^*) > 0$.

We need to find $x^*(\cdot)$ that solves the problem

$$\min\left\{\int_0^\infty I(\dot{x}(s))\mathrm{d}s : x(0) = 0, x(t) \geq 1 \text{ for some } t > 0\right\}.$$

Since $I(z) \geq 0$ it is evident that the optimal path must be such that $I(\dot{x}^*(t)) = 0$ for $t > T^*$, where $T^* = \inf\{t \geq 0 : x^*(t) \geq 1\}$. This forces $\dot{x}^*(t) = \dot{\psi}(0)$ for $t > T^*$. Therefore, we need to minimize $\int_0^T I(\dot{x}(s))\mathrm{d}s$, where $T = \inf\{t \geq 0 : x(t) \geq 1\}$. Now, the function

$I(\cdot)$ is convex because it is the Legendre transform of the convex function $\psi(\cdot)$ (see for example, [31, Part III, Section 12]). So, by convexity of $I(\cdot)$ and by Jensen's inequality we have that

$$T\frac{1}{T}\int_0^T I(\dot{x}(s))\mathrm{d}s \geq TI\left(\frac{1}{T}\int_0^T \dot{x}(s)\mathrm{d}s\right) = TI(1/T). \tag{10}$$

Optimizing the right hand side of (10) over T we then must select T^* in order to satisfy the first order conditions of optimality, namely,

$$-T^*\frac{\mathrm{d}I(1/T^*)}{\mathrm{d}T^*} = I\left(1/T^*\right). \tag{11}$$

Now we will use the previous equation to characterize the optimal path. Note that

$$I(z) = z\theta_z - \psi(\theta_z),$$

where $\dot{\psi}(\theta_z) = z$. This implies that $\mathrm{d}I(z)/\mathrm{d}z = \theta_z$ and therefore (11) is equivalent to

$$\frac{1}{T^*}\theta_{1/T^*} = \frac{1}{T^*}\theta_{1/T^*} - \psi(\theta_{1/T^*}) \Rightarrow \psi(\theta_{1/T^*}) = 0.$$

The previous equation implies that $\theta_{1/T^*} = \theta^*$ and therefore these calculations suggest that $x^*(t) = t\dot{\psi}(\theta^*)$ for $t \leq T^* = 1/\dot{\psi}(\theta^*)$. To complete the argument we must show that the lower bound in (10) is achieved by our selection, but this is clearly the case given that our candidate path is linear up to T^* and therefore its derivative is constant.

The corresponding importance sampling algorithm then consists in simulating each increment X_k according to the exponentially tilted distribution induced by the tilting parameter θ^*. Of course, one stops the simulation until time τ_n which is finite with probability one because under the importance sampling distribution $E_{\theta^*}X_k = \dot{\psi}(\theta_*) > 0$. A single replication of the importance sampling estimator is then given by

$$R_n := \prod_{k=1}^{\tau_n}\frac{\mathrm{d}P}{\mathrm{d}P_{\theta^*}}(X_k)I(\tau_n < \infty) = \exp(-\theta^* S_{\tau_n}).$$

We now proceed to verify weak efficiency. We first need to obtain the asymptotics for α_n. One can use the previous optimality argument for x^* combined with Theorem 1 to conclude that

$$\alpha_n = \exp(-nI(x^*) + o(n)) = \exp(-n\theta^* + o(n))$$

as $n \to \infty$ (in particular this corresponds to the limit (4)). On the other hand, using the expression for the second moment given in (3) we conclude that

$$\begin{aligned} ER_n = E[\exp(-\theta^* S_{\tau_n})I(\tau_n < \infty)] &\leq \exp(-n\theta^*)P(\tau_n < \infty) \\ &= \exp(-2n\theta^* + o(n)). \end{aligned}$$

The previous expression therefore implies that if there exists $\theta^* > 0$ such that $\psi(\theta^*) = 0$ and $\dot{\psi}(\theta^*) \in (0, \infty)$, then R_n is a weakly efficient estimator. Now, in fact, assuming only a mild regularly condition, for instance if the X_k's have a continuous distribution, it turns out that one can apply renewal theory to the increments of the strictly increasing ladder heights (i.e. the increments of the sequence of new records of the random walk) to conclude that R_n is in fact a strongly efficient estimator (see [30, p. 375 Theorem 7.1]). If we have that $\dot{\psi}(\theta^*) = \infty$, then strong efficiency is violated but one can obtain weak efficiency or polynomial complexity depending on the tail of the distribution of X_k.

We now move on to study Example 2. We have selected this example because it illustrates that direct application of state-independent importance sampling, guided by direct, yet sensible, large deviations principles, might not yield efficient estimators.

Example 2 (*Lack of Efficiency from Tracking the Optimal Path in a State-Independent Way*)**.** Given $a < 0 < b$ and $S_0 = 0$ we are interested in efficient estimation via simulation of the probability

$$\alpha_n = P[\min_{0\leq k\leq n} S_k \leq an, S_n > nb]. \tag{12}$$

Such probability can be interpreted as the price of a digital knock-in option in the absence of interest rates. That is, the price of an option that upon exercise pays one unit at time n if and only if both $S_n > nb$ and, in addition, the random walk hits a level less or equal than an before time n. (See [32] for more applications of importance sampling to option pricing.)

For concreteness, in order to facilitate the discussion, let us assume that X_k is standard Gaussian. The calculus of variations problem that must be solved takes the form

$$\inf\left\{\frac{1}{2}\int_0^1 \dot{x}(s)^2\,\mathrm{d}s : x(0) = 0, \min_{0\leq u\leq 1} x(u) \leq a, x(1) \geq b\right\}. \tag{13}$$

The solution x^* attaining the infimum in (13) can be computed explicitly as

$$x^*(t) = s_1 tI(t \leq t_0) + [s_2(t - t_0) - a]I(t_0 < t \leq 1),$$

where

$$s_1 = a/t_0, \qquad s_2 = (b-a)/(1-t_0), \qquad t_0 = -a/(b-2a).$$

Verifying that this indeed is the optimal path is done using Jensen's inequality and convexity arguments similar in spirit to those used in Example 1. In turn, we obtain that

$$\alpha_n = \exp(-n(b-2a)^2/2 + o(n)) \tag{14}$$

as $n \to \infty$ by evaluating $I(x^*)$. Alternatively, given that we are working with Gaussian increments, we can replace the random walk by Brownian motion $B(\cdot)$ and then, using the reflection principle, we can conclude that

$$\begin{aligned} P(\min_{0\leq t\leq n} B(t) \leq an, B(n) > bn) &= P(B(n) \leq -(b-2a)n) \\ &= \exp(-n(b-2a)^2/2 + o(n)) \end{aligned}$$

as $n \to \infty$, obtaining the asymptotic rate indicated earlier for α_n.

Since we are assuming Gaussian increments, the tilting parameter for the increment X_k, namely θ_k^*, matches $\dot{x}^*(k/n)$ and this yields

$$\begin{aligned} \theta_k^* &= -(b-2a)I(-na/(b-2a) > k) \\ &\quad + \frac{(b-a)I(-na/(b-2a) \leq k)}{1-(-a)/(b-2a)} \\ &= -(b-2a)I(-na/(b-2a) > k) \\ &\quad + (b-2a)I(-na/(b-2a) \leq k). \end{aligned}$$

The importance sampling estimator then takes the form

$$\begin{aligned} R_n &= \exp(-[\theta_0^* S_{k_0} - k_0\theta_0^{*2}/2] - \theta_n^*(S_n - S_{k_0})) \\ &\quad \times \exp((n-k_0)\theta_n^{*2}/2)I(\min_{0\leq k\leq n} S_k \leq an, S_n > bn), \end{aligned}$$

where $k_0 = \lfloor -an/(b-2a) \rfloor$.

We now show that R_n might not be efficient. We write $\widetilde{P}(\cdot)$ to denote the probability measure induced by sampling the X_k's independently according to the exponential tilting parameter θ_k^*. First we develop a lower bound for $\widetilde{E}[R_n^2]$ by constructing a suitable set under which the path deviates a distance of size $\varepsilon > 0$ (in supremum norm) from the optimal trajectory. In order to simplify the notation let $\theta^* = (b-2a)$ and then note that

$$\begin{aligned} R_n^2 &= \exp(-n\theta^{*2})\exp\left(2\theta^*\left[S_{k_0} - \left(S_n - S_{k_0}\right) + n\theta^*\right]\right) \\ &\quad \times I(\min_{0\leq k\leq n} S_k \leq an, S_n > bn). \end{aligned}$$

We write $A_n = \{\min_{0\le k\le n} S_k \le an, S_n > bn\}$ and introduce the sets

$$B_\varepsilon(n) = A_n \cap \left\{S_{k_0} + n\theta^* \ge n\varepsilon\right\},$$
$$C_\varepsilon(n) = B_\varepsilon(n) \cap \left\{S_{k_0} - \left(S_n - S_{k_0}\right) + n\theta^* \ge n\varepsilon\right\}.$$

Then,

$$\begin{aligned}\widetilde{E}[R_n^2] &= \exp\left(-n\theta^{*2}\right)\widetilde{E}\left[\exp\left(2\theta^*\left[S_{k_0} - \left(S_n - S_{k_0}\right) + n\theta^*\right]\right); A_n\right]\\ &\ge \exp\left(-n\theta^{*2}\right)\widetilde{E}\left[\exp\left(2\theta^*\left[S_{k_0} - \left(S_n - S_{k_0}\right) + n\theta^*\right]\right); B_\varepsilon(n)\right]\\ &\ge \exp\left(-n\theta^{*2}\right)\exp\left(4\theta^*\varepsilon n\right)\widetilde{P}\left(C_\varepsilon(n)\right).\end{aligned}$$

Next, because $C_\varepsilon(n)$ involves a deviation of size ε in supremum norm relative to the optimal trajectory, it is not difficult to verify that

$$\lim_{n\to\infty}\frac{\log\widetilde{P}\left(C_\varepsilon(n)\right)}{n} = -2\theta^*\varepsilon + O\left(\varepsilon^2\right).$$

So,

$$\lim_{n\to\infty}\frac{1}{n}\log\widetilde{E}[R_n^2] \ge -\theta^{*2} + 2\theta^*\varepsilon + O\left(\varepsilon^2\right) \ge -\theta^{*2} + \delta, \tag{15}$$

if one picks ε small enough such that $2\theta^*\varepsilon + O\left(\varepsilon^2\right) \ge \delta > 0$ for some $\delta > 0$.

Combining (14) together with (15) we conclude that R_n is not in general even weakly efficient.

3.3. Notes on state-independent importance sampling and large deviations

Theory of large deviations from a function space and sample path perspective is comprehensively documented in [25]. On more specific application areas, [33] discusses large deviations in queueing and computer science, while [34] focuses on applications in statistical mechanics. Ref. [35] provides an alternative approach, based on control theory and weak convergence analysis, to develop large deviations results.

The types of importance samplers discussed in Examples 1 and 2 originate from the work of [29] in the context of sequential analysis. Another early reference in this spirit is [36]. The paper [37] develops and proves several general results on importance samplers for Markov chains and semi-Markov processes. Level-crossing probabilities on random walks, as discussed in Examples 1 and 2, have important connections to insurance and performance evaluation of queueing systems. For the former, readers are referred to [38,39] and Chapter X of [40]. The paper [41] studies the extension to Markov-modulated walks on general state space. For queueing, [28] discusses exponential change-of-measure in several applications, and points out a duality link to steady-state estimation via level-crossing problems. Other related papers on change-of-measure from a random walk perspective applied to queueing models include [42–44].

The fact that reasonable importance samplers based on large deviations results can lead to efficiency deterioration was first pointed out in the 90's. The paper [1] studies the problem in the context of tandem queues, which is further analyzed in [45]. The issue is elaborated and strengthened in several examples in [2]. Ref. [46] and Chapter 5 of [47] discuss such inefficiency from so-called dominating point geometry in a general setting, while [48] extends the conditions using Varadhan's lemma. The recent work of [6] aims to remedy such problems in generality, drawing upon the theory of subsolution in Isaacs equation (this will be our main topic of discussion in the next section).

For general review on rare-event simulation and exponential change-of-measure, readers are referred to [49,47] and Chapter VI of [23]. The work of [50] provides a comprehensive discussion on the active research directions and list of references. For ease of explanation of recent advances and as supplement to [50], in this paper we focus mainly on random walk problems, and we emphasize that the scope of applications for importance sampling is much wider. For more specific references on general queueing applications, readers are referred to [51,52]. In the context of infinite and many-server systems especially, recent work include [53–55]. Another related application area is the reliability estimation of highly dependable Markovian systems, of which [51] again provides an excellent list of earlier references. More recent documentation is in [56]. Other applications of importance sampling are in finance (see for example [32]), physics (see for example [57]), systems biology (see [58]), combinatorics and counting (see [59,60,12]), among others.

We also want to mention that besides importance sampling, another powerful tool for variance reduction in rare-event problems is known as the splitting algorithm. Rather than modifying the underlying probability distribution, splitting algorithm divides the simulation trajectory space into levels (more precisely a nested sequence of subsets), with the hope that the success rate of passing through each level stands high even though the target probability is rare. Sample trajectories then split or branch out at each level to enhance arrivals at the next level, and finally up to the target set of interest. Hence the weights of the samples are based on the calculation of the associated branching process instead of a change-of-measure. Interested readers can go to [61] and Chapter VI Section 9 of [23]. The use of large deviations in analyzing splitting algorithms is studied in [62,63]. After that, [64] proposes a subsolution approach which is also closely related to our next section, while recently [65] discusses such algorithms in counting and computer science. Related methodology for the case of a fixed large deviations parameter, based on Feynman–Kac flows which accommodate most of the particle methods used in practice, is studied in [66].

Finally, besides drawing upon the theory of large deviations, importance samplers, with the goal of minimizing estimation variance, can also be implemented from an empirical perspective. These are schemes whose parameters are unknown and are updated from past samples in the simulation (note that the set of parameters can be huge, such as all the entries of a transition matrix in a Markov chain). As opposed to the optimal exponential change-of-measure in Examples 1 and 2 (and also the state-dependent samplers in the next sections) whose exponential parameters can be solved by looking at the associated optimization or variational problems, these empirical Monte Carlo schemes do not have analytical solution to the change-of-measure parameters, and hence empirical measurement is needed. Under appropriate assumptions, they can be proved to reduce variance or lead to exponential convergence of the estimate. Some important approaches in such setting are adaptive Monte Carlo (see e.g. [67,68]), the cross-entropy method (see [69–72]), and stochastic approximation based method (see [73,74]). Moreover, recent work of [75] considers an interesting importance sampling with resampling approach for Markov-modulated problems that avoids the need for computing the associated eigenfunctions.

4. State-dependent techniques for light-tailed systems

4.1. Solutions to a deterministic control problem and efficient importance sampling

In order to deal with the types of issues illustrated by Example 2 above [5] proposed a control theoretic approach for the problem of designing efficient importance sampling estimators. We will formally characterize the solution to a deterministic control problem proposed by Dupuis and Wang [5] in the setting of multidimensional random walks.

In this section we assume that $(X_k : k \geq 1)$ is a sequence of iid rv's taking values in R^d and such that $\psi(\theta) = \log E \exp(\langle X_k, \theta \rangle)$ is finite everywhere. We then let $S_k = X_1 + \cdots + X_k$ and consider a scaling similar to that introduced in Mogulskii's theorem. We let

$$Y_n(t) = y + S_{\lfloor nt \rfloor}/n,$$

for some $y \in R^d$. Let B be a closed set with non-empty interior and define $\tau_B(n) = \inf\{t > 0 : Y_n(t) \in B\}$. We assume that B is attainable in the sense that $P(\tau_B(n) < \infty | Y_n(0) = y) > 0$ for each $y \in R^d$.

Consider the question of computing via simulation for two disjoint sets B and C; in addition to the regularity imposed on B we will impose an additional large deviations requirement below. We are interested in computing

$$\begin{aligned}\alpha_n(y) &= P_y(\tau_B(n) < \tau_C(n), \tau_B(n) < \infty) \\ &:= P(\tau_B(n) < \tau_C(n), \tau_B(n) < \infty | Y_n(0) = y).\end{aligned}$$

We assume that B and C are regular in the sense that

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_n(y) = -I_{B,C}(y),$$

where

$$I_{B,C}(y) = \inf \left\{ \int_0^t J(\dot{x}(s))\, ds : x(0) = y, x(t) \in B \text{ for some } t < \infty \,\&\, x(s) \notin C \text{ if } s < t \right\}.$$

See for instance the text of [35, p. 66] or [25, p. 176].

We are interested in choosing a suitable importance sampling estimator with optimal performance in the sense of minimizing its second moment. The class of policies that we consider are based on exponential tilting so that the estimator ultimately takes the form

$$R_n = \exp\left(- \sum_{k=1}^{n\tau_B(n)} \langle \theta_k, X_k \rangle + \sum_{k=1}^{n\tau_B(n)} \psi(\theta_k) \right) I(\tau_B(n) < \tau_C(n), \tau_B(n) < \infty), \tag{16}$$

where θ_k is adapted to the sequence $\left(X_j : j \geq 1\right)$ in the sense that θ_k is allowed to depend only on $X_1, \ldots, X_{k-1}$. The selection of exponential tiltings as the family of controls is motivated by the discussion involving the optimization problem (8).

Let $V_n(y)$ denote the optimal second moment of R_n with y being the initial position of the random walk. It follows immediately that the HJB (Hamilton–Jacobi–Bellman) equation corresponding to finding the optimal policy (i.e. the optimal sequence of θ_k's) to minimize the second moment of R_n takes the form

$$\begin{aligned}V_n(y) &= \inf_\theta E_\theta[\exp(-2\langle \theta, X \rangle + 2\psi(\theta)) V_n(y + X/n)] \\ &= \inf_\theta E[\exp(-\langle \theta, X \rangle + \psi(\theta)) V_n(y + X/n)]\end{aligned} \tag{17}$$

for $y \notin B \cup C$ and subject to the boundary condition that $V_n(y) = 1$ for $y \in B$ and $V_n(y) = 0$ if $y \in C$. If there is an optimal policy $(\theta_k^* : k \geq 1)$ generated by the previous HJB equation, then generally it would be the case that

$$\begin{aligned}V_n(y) &= E \exp\left(- \sum_{k=1}^{n\tau_B(n)} \langle \theta_k^*, X_k \rangle + \sum_{k=1}^{n\tau_B(n)} \psi(\theta_k^*) \right) I(\tau_B(n) < \tau_C(n), \tau_B(n) < \infty) \\ &\leq E \exp\left(- \sum_{k=1}^{n\tau_B(n)} \langle \theta_k, X_k \rangle + \sum_{k=1}^{n\tau_B(n)} \psi(\theta_k) \right) I(\tau_B(n) < \tau_C(n), \tau_B(n) < \infty),\end{aligned}$$

for any adapted policy $(\theta_k : k \geq 1)$ (adapted in the sense described earlier below display (16)).

The large deviations scaling suggests writing $V_n(y) = \exp(-nH_n(y))$ and thus we should expect $H_n(y) \to H(y)$ as $n \to \infty$ for some function $H(y)$. If we proceed using this postulated limit in the previous HJB equation, after taking logarithms, formally and without being careful about underlying smoothness assumptions and errors incurred, we arrive at the approximation

$$\begin{aligned}-nH(y) &\approx \min_\theta \log E[\exp(-\langle \theta, X \rangle + \psi(\theta) - nH(y + X/n))] \\ &\approx \min_\theta \log E[\exp(-\langle \theta, X \rangle \\ &\qquad + \psi(\theta) - nH(y) - \langle \nabla H(y), X \rangle)].\end{aligned} \tag{18}$$

Equivalently, we have that

$$\begin{aligned}0 &\approx \min_\theta \log E[\exp(-\langle \theta, X \rangle + \psi(\theta) - \langle \nabla H(y), X \rangle)] \\ &= \min_\theta \log \exp(\psi(\theta) + \psi(-\nabla H(y) - \theta)) \\ &= \min_\theta [\psi(\theta) + \psi(-\nabla H(y) - \theta)].\end{aligned} \tag{19}$$

First order optimality conditions imply that at the optimal value $\theta^*(y)$ one has

$$\nabla \psi(\theta^*(y)) = \nabla \psi(-\nabla H(y) - \theta^*(y)),$$

which yields $\theta^*(y) = -\nabla H(y)/2$ and therefore we conclude that Eq. (19) can be expressed as

$$2\psi(-\nabla H(y)/2) = 0, \tag{20}$$

subject to the boundary conditions (inherited from (17)) equal to $H(y) = 0$ for $y \in B$ and $H(y) = \infty$ for $y \in C$. Eq. (20) is an instance of a so-called Isaacs equation.

On the other hand, we have assumed

$$\alpha_n(y) = \exp(-nI_{B,C}(y) + o(n))$$

as $n \to \infty$ and, by conditioning on the first increment of the random walk, we also have that

$$\alpha_n(y) = E\alpha_n(y + X/n) \tag{21}$$

subject to the constraints that $\alpha_n(y) = 1$ for $y \in B$ and $\alpha_n(y) = 0$ if $y \in C$. Proceeding to analyze equality (21) formally as we did for the discussion leading to (20), we conclude that

$$\begin{aligned}\exp(-nI_{B,C}(y) + o(n)) &= E \exp(-nI_{B,C}(y + X/n)) \\ &\approx E \exp(-nI_{B,C}(y) - \langle \nabla I_{B,C}(y), X \rangle).\end{aligned}$$

Equivalently, taking logarithms, we arrive at

$$\psi(-\nabla I_{B,C}(y)) = 0,$$

subject to the boundary conditions implied by $\alpha_n(y)$, namely, $I_{B,C}(y) = 0$ for $y \in B$ and $I_{B,C}(y) = \infty$ if $y \in C$. These considerations, together with our analysis leading to (20) yield that $H(y) = 2I_{B,C}(y)$ and therefore, under suitable regularity conditions we obtain that applying importance sampling with exponential tilting given by the tilting parameter $\theta^*(y) = -\nabla I_{B,C}(y)$ allows us to obtain an asymptotically optimal estimator. Sufficient conditions required to rigorously substantiate this result are given in the following theorem formulated by Dupuis and Wang [5] in the case of large deviations problems for S_n/n.

Theorem 2 (*Adaptation From [5] Theorem 2.1*). *Let S_n, Y_n and $\psi(\theta)$ be defined as in the beginning of this section. Consider the problem of estimating $P(S_n/n \in A | S_0 = 0)$ where A is a closed set that satisfies*

$$\inf_{y \in A^\circ} I(y) = \inf_{y \in A} I(y)$$

and $I(y) = \sup_{\theta \in \mathbb{R}^d} [\langle \theta, y \rangle - \psi(\theta)]$. Let us use ∇_x for the gradient of a function with respect to the component x. Suppose that there

is a classical sense solution corresponding to the Isaacs equation analogous to (20) given by

$$\partial_t H(x,t) = 2\psi(\nabla_x H(x,t)/2)$$

subject to the boundary condition $\lim_{t\searrow 0} H(x,t) = \infty \times I(x \notin A)$ *(by convention we consider* $\infty \times 0 = 0$*). The optimal state-dependent exponential tilting scheme with k-step exponential parameter* $\theta_k^* = \nabla_x H(S_k/n, k/n)$ *generates the estimator*

$$\exp\left(\sum_{k=1}^{n}(-\langle\theta_{k-1}^*, X_k\rangle + \psi(\theta_{k-1}^*))\right) I(S_n/n \in A).$$

Let $V_n(0)$ *be the second moment of this estimator and let* $H_n(0)$ *be such that* $V_n(0) = \exp(-nH_n(0))$*. We have that*

$$\lim_{n\to\infty} H_n(0) = 2\inf_{y\in A} I(y) = H(0).$$

The problem with the previous result is that it is rarely applicable in most situations of interest, especially because it is often the case that there is no classical solution to the Isaacs equation. Other connections are discussed in [76]. The function $H(y) = 2I_{B,C}(y)$ typically provides a solution in a weak sense (a so-called viscosity subsolution). The extent to which this condition can be relaxed and the type of optimality results that can be obtained beyond weak efficiency is not fully understood.

4.2. Subsolutions to a deterministic control problem and efficient importance sampling

We take as starting point the HJB equation introduced in (17). The next lemma provides an inequality whose solution gives an upper bound for the value function in (17).

Lemma 1. *Suppose that one finds a non-negative function* $U(\cdot)$ *such that*

$$\begin{aligned} U_n(y) &\geq E[\exp(-\langle\theta, X\rangle + \psi(\theta))U_n(y + X/n)] \\ &\geq \inf_{\theta} E[\exp(-\langle\theta, X\rangle + \psi(\theta))U_n(y + X/n)], \end{aligned}$$

where $y \notin B\cup C$ *and subject to the boundary condition that* $U_n(y) \geq 1$ *for* $y \in B$*. Then,*

$$U_n(y) \geq E\left[\exp\left(-\sum_{j=1}^{\tau_{B,C}}\langle\theta_j, X_j\rangle + \sum_{j=1}^{\tau_{B,C}}\psi(\theta_j)\right) I(\tau_{B,C} < \infty, \tau_B < \tau_C)\right]$$

for any adapted policy $(\theta_j : j \geq 1)$*, where* $\tau_{B,C} = n(\tau_B \wedge \tau_C)$ *and* τ_B, τ_C *are defined as in the beginning of Section 4.1.*

Proof. Recall $\tau_{B,C} = n \times (\tau_B(n) \wedge \tau_C(n))$ and fix any policy $(\theta_k : k \geq 0)$ for which

$$U_n(y) \geq E[\exp(-\langle\theta, X\rangle + \psi(\theta))U_n(y + X/n)],$$

and $\tau_{B,C} < \infty$. Consider the process

$$M_{k+1} = U_n(S_{\tau_{B,C}\wedge(k+1)}/n)\exp\left(-\sum_{j=1}^{\tau_{B,C}\wedge(k+1)}\langle\theta_j, X_j\rangle + \sum_{j=1}^{\tau_{B,C}\wedge(k+1)}\psi(\theta_j)\right).$$

We claim that $(M_k : k \geq 0)$ is a supermartingale. Note that

$$\begin{aligned} E[M_{k+1}|(S_j : 1 \leq j \leq k)] &= E[M_{k+1}|(S_j : 1 \leq j \leq k)]I(\tau_{B,C} > k) \\ &\quad + E[M_{k+1}|(S_j : 1 \leq j \leq k)]I(\tau_{B,C} \leq k). \end{aligned}$$

It follows that on $\{\tau_{B,C} > k\}$,

$$\begin{aligned} E[M_{k+1}|(S_j : 1 \leq j \leq k)] &= \exp\left(-\sum_{j=1}^{k}\langle\theta_j, X_j\rangle + \sum_{j=1}^{k}\psi(\theta_j)\right) EU_n(S_{k+1}/n)\exp(-\langle\theta, X_{k+1}\rangle + \psi(\theta_{k+1})) \\ &\leq \exp\left(-\sum_{j=1}^{k}\langle\theta_j, X_j\rangle + \sum_{j=1}^{k}\psi(\theta_j)\right) U_n(S_k/n). \end{aligned}$$

Evidently, on $\{\tau_{B,C} \leq k\}$,

$$\begin{aligned} &E[M_{k+1}|(S_j : 1 \leq j \leq k)] \\ &\quad = U_n(S_{\tau_{B,C}}/n)\exp\left(-\sum_{j=1}^{\tau_{B,C}}\langle\theta_j, X_j\rangle + \sum_{j=1}^{\tau_{B,C}}\psi(\theta_j)\right). \end{aligned}$$

Therefore,

$$\begin{aligned} &E[M_{k+1}|(S_j : 1 \leq j \leq k)] \\ &\quad \leq \exp\left(-\sum_{j=1}^{k}\langle\theta_j, X_j\rangle + \sum_{j=1}^{k}\psi(\theta_j)\right) U_n(S_k/n)I(\tau_{B,C} > k) \\ &\qquad + U_n(S_{\tau_{B,C}}/n)\exp\left(-\sum_{j=1}^{\tau_{B,C}}\langle\theta_j, X_j\rangle + \sum_{j=1}^{\tau_{B,C}}\psi(\theta_j)\right) I(\tau_{B,C} \leq k) \\ &\quad = M_k \end{aligned}$$

and, if $S_0 = ny$,

$$\begin{aligned} U_n(y) &\geq E\left[U_n(S_{\tau_{B,C}\wedge k}/n)\exp\left(-\sum_{j=1}^{\tau_{B,C}\wedge k}\langle\theta_j, X_j\rangle + \sum_{j=1}^{\tau_{B,C}\wedge k}\psi(\theta_j)\right)\right] \\ &\geq E\left[U_n(S_{\tau_{B,C}}/n)\exp\left(-\sum_{j=1}^{\tau_{B,C}}\langle\theta_j, X_j\rangle + \sum_{j=1}^{\tau_{B,C}}\psi(\theta_j)\right) I(\tau_{B,C} \leq k)\right]. \end{aligned}$$

By Fatou's lemma, we obtain that

$$U_n(y) \geq E\left[U_n(S_{\tau_{B,C}}/n)\exp\left(-\sum_{j=1}^{\tau_{B,C}}\langle\theta_j, X_j\rangle + \sum_{j=1}^{\tau_{B,C}}\psi(\theta_j)\right) I(\tau_{B,C} < \infty)\right].$$

Using the boundary condition on B and the fact that $U_n(y)$, being non-negative, must satisfy that $U_n(y) \geq 0$ on C we obtain

$$U_n(y) \geq E\left[\exp\left(-\sum_{j=1}^{\tau_{B,C}}\langle\theta_j, X_j\rangle + \sum_{j=1}^{\tau_{B,C}}\psi(\theta_j)\right) I(\tau_{B,C} < \infty, \tau_B < \tau_C)\right]. \quad \square$$

Using a similar formal argument as the one introduced in the previous section in (18) we write $U_n(y) = \exp(-nG_n(y))$ and

postulate $G_n(y) \to G(y)$ as $n \to \infty$ for some function $G(y)$. If we proceed using this postulated limit in the inequality obtained in the previous lemma, after taking logarithms, we arrive at

$$\begin{aligned}-nG(y) &\gtrapprox \min_\theta \log E[\exp(-\langle\theta, X\rangle + \psi(\theta) - nG(y + X/n))] \\ &\approx \min_\theta \log E[\exp(-\langle\theta, X\rangle + \psi(\theta) - nG(y) \\ &\quad - \langle\nabla G(y), X\rangle)].\end{aligned}$$

We then conclude, as in the analysis leading to (23), that

$$0 \geq \min_\theta[\psi(\theta) + \psi(-\nabla G(y) - \theta)]. \tag{22}$$

First order optimality conditions imply that at the optimal value $\theta^*(y)$ one has

$$\nabla\psi(\theta^*(y)) = \nabla\psi(-\nabla G(y) - \theta^*(y)),$$

which yields $\theta^*(y) = -\nabla G(y)/2$ and therefore we conclude that inequality (22) can be expressed as

$$0 \geq 2\psi(-\nabla G(y)/2), \tag{23}$$

for $y \notin B \cup C$ and subject to the boundary conditions (inherited from Lemma 1) equal to $G(y) \leq 0$ for $y \in B$. Inequality (23) is the corresponding subsolution to the Isaacs equation developed in (20). In order to conclude asymptotic optimality for $\alpha_n(y_0)$ (here we let y_0 as the initial position) we must have $G(y_0) \geq 2I_{B,C}(y_0)$. Sufficient conditions required to rigorously substantiate this result are given in the following theorem.

Theorem 3 (*Adaptation from [6] Theorem 8.1*)**.** *Let X_k and S_k be defined as usual, with $\psi(\theta) = E\exp(\langle\theta, X_k\rangle) < \infty$ for any $\theta \in \mathbb{R}^d$. Suppose that $G(y)$ satisfies (22) with the boundary condition $G(y) \leq 0$ for $y \in B$. Let $\theta^*(y) = -\nabla G(y)/2$. Then, the state-dependent sampler using k-th step exponential parameters $\theta^*(S_k/n)$ given by*

$$R_n = \exp\left(\sum_{k=1}^{\tau_{B,C}} \left(-\langle\theta^*(S_{k-1}/n), X_k\rangle + \psi(\theta^*(S_{k-1}/n))\right)\right) I(\tau_{B,C} < \infty, \tau_B < \tau_C)$$

has second moment satisfying

$$\liminf_{n\to\infty} -\frac{1}{n}\log E[R_n^2] \geq G(y_0)$$

where $S_0 = y_0$. Consequently, if $G(y_0) \geq 2I_{B,C}(y_0)$, then R_n is an asymptotically optimal estimator.

We now apply the previous result to a couple of examples.

Example 3 (*Multidimensional First Passage Time Problems*)**.** Assume that $\mu = EX_i \neq 0$ and consider vectors $a_1, \ldots, a_m \in R^d$ such that $\langle\mu, a_i\rangle < 0$ for all $i \in \{1, \ldots, m\}$, where $\langle x, y\rangle$ is used to denote the inner product between the vectors $x, y \in R^d$. Define the event $B = \cup_{i=1}^m \{y\colon \langle y, a_i\rangle \geq 1\}$ and consider the first passage time probability

$$\alpha_n(y) = P(\tau_B(n) < \infty | Y_n(0) = y),$$

where $\tau_B(n) = \inf\{t \geq 0\colon Y_n(t) \in B\}$. Suppose that for each $i \in \{1, \ldots, m\}$ there exists $\theta_i^* \neq 0$ such that $\psi(a_i\theta_i^*) = 0$ and $\langle a_i, \nabla\psi(a_i\theta_i^*)\rangle < \infty$. Just as we explained in Example 1, since $Ea_i^T X_k < 0$ we must have that $\theta_i^* > 0$ and also that $\langle a_i, \nabla\psi(a_i\theta_i^*)\rangle \in (0, \infty)$. This problem is the natural generalization of Example 1 and can also be interpreted in terms of an insurance formulation in which one computes a ruin probability of an insurance company with several lines of business and for which one is allowed to borrow some resources from one line to another one. This formulation is discussed in [77].

We can take advantage of Example 1 in order to obtain the asymptotics for α_n. In particular, define $\tau_i(n) = \inf\{t > 1\colon \langle Y(t), a_i\rangle \geq 1\}$ for each $i \in \{1, \ldots, m\}$ and note that one has the elementary inequalities

$$\begin{aligned}&\max_{i=1}^m P(\tau_i(n) < \infty | Y_n(0) = y) \\ &\leq \alpha_n(y) \leq \sum_{i=1}^m P(\tau_i(n) < \infty | Y_n(0) = y).\end{aligned} \tag{24}$$

Evidently, $P(\tau_i(n) < \infty | Y_n(0) = y)$ is equivalent to the one dimensional first passage time probability discussed in Example 1. To see this, one simply needs to work with the one dimensional random walk with iid increments $(X_1(i), X_2(i), \ldots)$ defined via $X_k(i) = \langle X_k, a_i\rangle$. We then obtain that

$$P(\tau_i(n) < \infty | Y_n(0) = y) = \exp[-n\theta_i^*(1 - \langle y, a_i\rangle) + o(n)] \tag{25}$$

as $n \to \infty$ and therefore,

$$\lim_{n\to\infty} \frac{1}{n}\log\alpha_n(y) = -\min_{i=1}^m \theta_i^*(1 - \langle y, a_i\rangle).$$

The natural state-independent technique, analogous to that of Example 1, consists in computing j such that

$$\theta_j^*(1 - \langle y, a_j\rangle) = \min_{i=1}^m \theta_i^*(1 - \langle y, a_i\rangle)$$

and applying exponential tilting according to the tilting parameter $a_j\theta_j^*$. Unfortunately, just as we showed in Example 2, it turns out that this strategy is not guaranteed to yield an efficient estimator; such a counter-example can be constructed using a two dimensional Gaussian random walk using a lower bound analogous to that developed for Example 2; the details are left as an exercise to the reader.

We now use the subsolution approach to construct an importance sampling change-of-measure for this problem. Ultimately we need to find a function $G(\cdot)$ satisfying the inequality (23) together with the corresponding boundary condition that $G(y) \leq 0$ for $y \in B$. Moreover, since in the end we are after an asymptotic upper bound for the second moment of the importance sampling estimator and we wish to achieve the same decay rate as $\alpha_n(y)^2$, it makes sense to use the union bound (24) and (25) to postulate

$$\exp(-nG(y)) = \left(\sum_{i=1}^m \exp[-n\theta_i^*(1 - \langle y, a_i\rangle)]\right)^2.$$

We now verify that our selection is indeed a valid subsolution. First, note that

$$\nabla G(y) = -2\sum_{j=1}^m \theta_j^* a_j p_j(y, n),$$

where

$$p_j(y, n) = \frac{\exp[-n\theta_i^*(1 - \langle y, a_i\rangle)]}{\sum_{i=1}^m \exp[-n\theta_i^*(1 - \langle y, a_i\rangle)]}.$$

Note that for each y, $(p_1(y, n), \ldots, p_m(y, n))$ can be interpreted as a probability vector. Moreover, observe that as $n \to \infty$, the probability vector degenerates and concentrates its mass on the indexes corresponding to the tilting parameters θ_j^* that drive the most likely way in which the random walk hits the target set. Now, using Jensen's inequality and convexity of $\psi(\cdot)$ we conclude that

$$\begin{aligned}\psi(-\nabla G(y)/2) &= \psi\left(\sum_{j=1}^m \theta_j^* a_j p_j(y, n)\right) \\ &\leq \sum_{j=1}^m p_j(y, n)\psi(\theta_j^* a_j) = 0.\end{aligned}$$

In the last equality we used that $\psi(\theta_j^* a_j) = 0$ for each $j = 1, \ldots, m$. Now, if $y \in B$, then there exists i such that $1 - \langle y, a_i \rangle \leq 0$ and therefore evidently $\exp(-nG(y)) \geq 1$ or, equivalently, $G(y) \leq 0$. We then conclude that the importance sampling scheme generated by using as exponential tilting

$$\theta_k(S_{k-1}/n) = \sum_{j=1}^{m} \theta_j^* a_j p_j(S_{k-1}/n, n) \tag{26}$$

for $k/n < \tau_B(n)$ is asymptotically optimal.

Remark 2. Note that we have restricted ourselves to importance sampling estimators of exponential tilting form. Nevertheless, the form of (26), which is expressed in terms of a mixture, suggests considering mixtures of exponential tiltings, or in other words, distributions of the form

$$Q(X_k \in \mathrm{d}y) = \sum_{j=1}^{m} p_j(S_{k-1}/n, n) \exp(\theta_j^* a y) P(X_k \in \mathrm{d}y). \tag{27}$$

Dupuis and Wang [6] also study these types of "generalized" controls which are asymptotically equivalent (in the sense of achieving also weak efficiency) to those in (26), see also [7]. The paper [78] shows that using Lyapunov inequalities the sampler (27) actually yields strong efficiency.

The subsolution approach is better appreciated in situations in which the system exhibits piecewise random walk dynamics, as it happens in the setting of Jackson networks. The next example illustrates this situation in the context of a two dimensional random walk.

Example 4 (*A First-Passage Time Problem With Discontinuous Statistics*)**.** Let $(Z_k : k \geq 1)$ be a sequence of iid two dimensional standard Gaussian random variables (i.e. having zero mean and identity covariance matrix). Let $\mu_1 = (\cos(9\pi/8), \sin(9\pi/8))^T$ and $\mu_2 = (\cos(11\pi/8), \sin(11\pi/8))^T$. We let $W = (W_k : k \geq 0)$ be a two dimensional process defined via

$$\begin{aligned} W_{k+1} &= W_k + Z_{k+1} + I(W_k(1) < W_k(2))\mu_1 + I(W_k(1) \\ &\quad \geq W_k(2))\mu_2, \end{aligned}$$

with $W_0 = ny$, where $y = (y(1), y(2))^T$ is such that $y(1) < 1$ and $y(2) < 1$. Define $Y_n(t) = W_{\lfloor nt \rfloor}/n$, $B = \{x = (x(1), x(2))^T : x(1) \geq 1 \text{ or } x(2) \geq 1\}$ and consider the problem of efficiently estimating via simulation

$$\alpha_n(0) = P(\tau_B(n) < \infty | Y_n(0) = 0),$$

where $\tau_B(n) = \inf\{t > 0 : Y_n(t) \in B\}$.

Although the process $(W_k : k \geq 0)$ is not a random walk, note that it behaves like a Gaussian random walk above and below the 45° line (i.e. the line $y(1) = y(2)$); only the drift changes in each of these two regions. We have selected the drift parameters to simplify our analysis and flesh out the main issue arising in constructing the subsolution in the current setting. A picture illustrating the drift of the process in the whole space is given next (see Fig. 1).

In this case it is not difficult to convince ourselves that if $y(1) < y(2) < 1$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_n(y) = 2\mu_1(2)(1 - y(2)).$$

and if $1 > y(1) \geq y(2)$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_n(y) = 2\mu_2(1)(1 - y(1)).$$

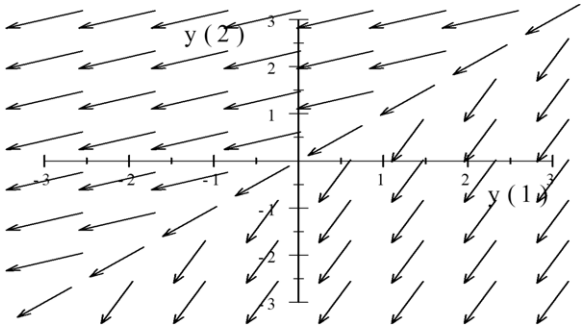

Fig. 1. Vector field of fluid paths.

The previous limits correspond to the first passage time problem of a one dimensional random walk with Gaussian increments having negative drift equal to $\mu_1(2) = \sin(9\pi/8) = \mu_2(1)$.

A similar development to that leading to (23) leads to the subsolution inequality for $G(\cdot)$ given by

$$2\psi(y, -\nabla G(y)/2) \leq 0,$$

where, for $y = (y(1), y(2))^T$,

$$\begin{aligned} \psi(y, \theta) &= \log E \exp(\langle Z_1, \theta \rangle \\ &\quad + I(y(1) \geq y(2)) \langle \mu_1, \theta \rangle + I(y(1) < y(2)) \langle \mu_2, \theta \rangle) \\ &= \|\theta\|^2/2 + I(y(1) \geq y(2)) \langle \mu_1, \theta \rangle \\ &\quad + I(y(1) < y(2)) \langle \mu_2, \theta \rangle, \end{aligned}$$

and the corresponding boundary condition $G(y) \leq 0$ for $y \in B$.

Now, let us first attempt a construction of the subsolution similar to that given in Example 3, namely by considering the function $\widetilde{G}(\cdot)$ such that

$$\begin{aligned} \widetilde{G}(y) &= -\frac{2}{n} \log(\exp[2n\mu_1(2)(1 - y(2))] \\ &\quad + \exp[2n\mu_2(1)(1 - y(1))]). \end{aligned}$$

Note that in this case we have that

$$-\nabla \widetilde{G}(y)/2 = \widetilde{p}_1(y, n)\theta_1^* + \widetilde{p}_2(y, n)\theta_2^*, \tag{28}$$

where

$$\widetilde{p}_1(y, n) = \frac{\exp[2n\mu_1(2)(1 - y(2))]}{\exp[2n\mu_1(2)(1 - y(2))] + \exp[2n\mu_2(1)(1 - y(1))]};$$

$$\theta_1^* = -(0, 2\mu_1(2)),$$

and

$$\widetilde{p}_2(y, n) = \frac{\exp[2n\mu_2(1)(1 - y(1))]}{\exp[2n\mu_1(2)(1 - y(2))] + \exp[2n\mu_2(1)(1 - y(1))]};$$

$$\theta_2^* = -(2\mu_2(1), 0).$$

Observe that

$$\begin{aligned} \psi(y, -\nabla \widetilde{G}(y)/2) &\leq \widetilde{p}_1(y, n)\psi(y, \theta_1^*) + \widetilde{p}_2(y, n)\psi(y, \theta_2^*) \\ &\leq 2\widetilde{p}_1(y, n) I(y(1) \geq y(2)) \\ &\quad + 2\widetilde{p}_2(y, n) I(y(2) > y(1)). \end{aligned}$$

Therefore, if $|y(1) - y(2)| \geq \delta$ for some $\delta > 0$ then

$$\psi(y, -\nabla \widetilde{G}(y)/2) \leq 2\exp(2\delta n \mu_1(2)). \tag{29}$$

So, while the subsolution inequality is not satisfied, the size of the violation converges to zero exponentially fast as $n \to \infty$. On the other hand, we can easily deduce that

$$\psi(y, \theta_1^*)\widetilde{p}_1(y, n) + \psi(y, \theta_2^*)\widetilde{p}_2(y, n) > \delta > 0$$

on a region of size $O(1/n)$ around the 45° line, namely the line $y(1) = y(2)$. While this region might look small, note that in the

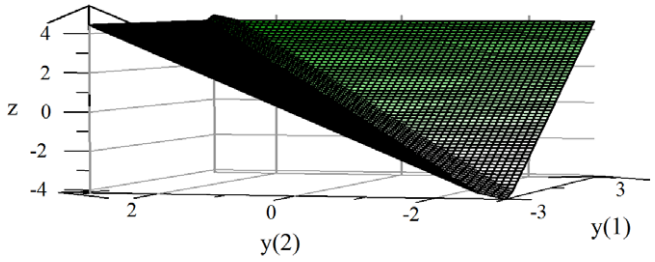

Fig. 2. Plot of the function $\max\left(-4\mu_1(2)y(1), -4\mu_1(2)y(2), 2\left\langle\theta', y\right\rangle + \Delta\right)$.

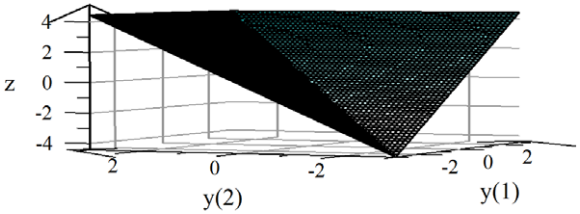

Fig. 3. Plot of the function $\widetilde{F}_0(y) = -4\mu_1(2)\max(y(1), y(2))$.

original scaling of the process W the region is of order $O(1)$, so the process might spend a significant amount of time (in fact a random amount of time of order $O(1)$) in a region where the likelihood ratio of the importance sampling distribution is not appropriately controlled.

The main difficulty with $\widetilde{G}(y)$ arises precisely because of the discontinuous dynamics of the process W right on the line $y(1) = y(2)$. In order to deal with this problem we note that

$$\lim_{n\to\infty} \widetilde{G}(y) = -2\max\{\langle\theta_1, y\rangle + \mu_2(1), \langle\theta_2, y\rangle + \mu_1(2)\} = 2\mu_2(1)\max\{y(2), y(1)\} - 2\mu_2(1). \tag{30}$$

We will need to slightly modify the function $\widetilde{G}(\cdot)$ right on the set $y(1) = y(2)$ and in such a way that one can obtain a smooth function $G(\cdot)$ whose gradient can be represented as a weighted sum, just as in (28). In order to do this, define

$$\widetilde{f}_1(y) = 2\left\langle\theta_1^*, y\right\rangle, \qquad \widetilde{f}_2(y) = 2\left\langle\theta_2^*, y\right\rangle,$$

and define the convex function

$$\widetilde{F}_0(y) = \max\{\widetilde{f}_1(y), \widetilde{f}_2(y)\} = -4\mu_2(1)\max\{y(2), y(1)\},$$

which is just a translation to the origin of the limit obtained in (30). The plot of $\widetilde{F}_0(\cdot)$ is given in Fig. 3.

The idea is to introduce an affine function $\widetilde{f}_0(y) = 2\left\langle\theta_0^*, y\right\rangle + \Delta$, for some $\Delta > 0$, so that we can define corresponding probabilities $p_0(y, n)$, $p_1(y, n)$ and $p_2(y, n)$, associated to θ_0^*, θ_1^* and θ_2^*, respectively, with the property that $p_1(y, n), p_2(y, n) = o(p_0(y, n)\exp(-\varepsilon n))$ for some $\varepsilon := \varepsilon(\Delta) > 0$ uniformly over y such that $|y(1) - y(2)| \le \delta := \delta(\Delta)$ and such that $\psi(y, \theta_0^*) \le 0$ for all y. We also wish the behavior of $p_0(y, n), p_2(y, n) = o(p_1(y, n)\exp(-\varepsilon n))$ uniformly over $y(2) > y(1)+\delta$ and similarly $p_0(y, n), p_1(y, n) = o(p_2(y, n)\exp(-\varepsilon n))$ uniformly over $y(1) > y(2) + \delta$. If in addition, we select $\psi(\theta_0, y) \le 0$, just as we obtained in the analysis (29), we will be violating the subsolution only by an exponentially small amount in n, which as it turns out, is enough to conclude asymptotic optimality.

A natural approach that comes to mind to construct such a function $\widetilde{f}_0(\cdot)$ is by considering a supporting hyperplane to the function $\widetilde{F}_0(\cdot)$ right along the axis $y(1) = y(2)$. Let us denote such supporting hyperplane by $f(y) = 2\langle\eta, y\rangle$. By shifting such hyperplane an amount Δ we obtain a picture such as the one shown in Fig. 2.

All subgradients corresponding to supporting hyperplanes strictly along $y(1) = y(2)$ are potential candidates for $2\theta_0^*$. However, as

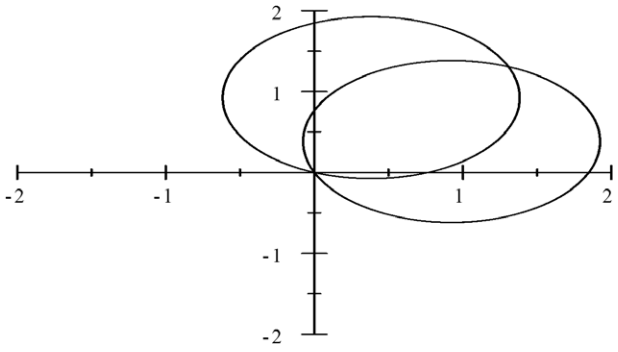

Fig. 4. Plot of the level curves $\{\theta : \psi(\theta, y) = 0\}$ for $y = (y(1), y(2))$ with $y(1) < y(2)$ and $y(1) > y(2)$.

we indicated before, we also wish to select $\psi(\theta_0^*, y) \le 0$. So, it is useful to plot the functions $\psi(\cdot, y)$ for $y(2) > y(1)$ and $y(2) < y(1)$ respectively, as we do next.

The intersection of the areas enclosed by both circles shown in Fig. 4 above correspond to the points θ for which $\psi(\theta, y) \le 0$, for each $y \in R^2$. A natural selection for θ_0^* (although clearly not the only one) is $\theta_0^* = (-\mu_2(1), -\mu_2(1))^T$. Clearly, since $-\mu_2(1) \in (0, 1)$ we have that $\psi(\theta_0^*, y) < 0$ for each $y \in R^2$ and such a selection indeed corresponds to the supporting hyperplane that makes the plot of Fig. 2 symmetric around the line $y(1) = y(2)$.

We now have all the necessary elements required to define our subsolution. In particular, by virtue of our considerations leading to the definition of $\widetilde{F}_0(\cdot)$ and our selection of θ_0^*, we let

$$G(y) = -\frac{2}{n}\log(\exp[n(\widetilde{f}_1(y) + \mu_2(1))] + \exp[n(\widetilde{f}_2(y) + \mu_2(1))] + \exp[n(\widetilde{f}_0(y) + \mu_2(1))]).$$

The importance sampling scheme is generated by using as exponential tilting parameter at time k:

$$\theta_k(\widetilde{W}_{k-1}/n) = \sum_{j=0}^{2} \theta_j^* p_j(\widetilde{W}_{k-1}/n, n),$$

where $p_j(\widetilde{W}_{k-1}/n, n)$ is as given in Box I below. Under the importance sampling distribution the process evolves according to the Markov chain

$$\widetilde{W}_{k+1} = \widetilde{W}_k + \widetilde{Z}_{k+1} + \theta_k(W_k/n), \qquad \widetilde{W}_0 = 0,$$

where $(\widetilde{Z}_k : k \ge 1)$ is a sequence of iid two dimensional standard Gaussian random variables. The importance sampling estimator, namely,

$$R_n = \exp\left(-\sum_{k=1}^{n\tau_B(n)} \theta_k(\widetilde{W}_{k-1}/n)(\widetilde{W}_k - \widetilde{W}_{k-1}) + \sum_{k=1}^{n\tau_B(n)} \psi(\widetilde{W}_{k-1}/n, \theta_k(\widetilde{W}_{k-1}/n))\right) I(\tau_B(n) < \infty),$$

satisfies

$$\lim_{n\to\infty} \frac{\log ER_n}{\log \alpha_n(0)} \ge 2 + \Delta/\mu_2(1).$$

One can select $\Delta := \Delta_n \longrightarrow 0$ at a suitable speed in order to conclude asymptotic optimality. These types of issues are discussed in the next subsection on remarks and further considerations on state-dependent importance sampling for light-tailed systems.

$$p_j(\widetilde{W}_{k-1}/n, n) = \frac{\exp[n(\widetilde{f}_j(\widetilde{W}_{k-1}/n) + \mu_2(1))]}{\exp[n(\widetilde{f}_1(\widetilde{W}_{k-1}/n) + \mu_2(1))] + \exp[n(\widetilde{f}_2(\widetilde{W}_{k-1}/n) + \mu_2(1))] + \exp[n(\widetilde{f}_0(\widetilde{W}_{k-1}/n) + \mu_2(1))]},$$

whenever $k/n < \tau_B(n)$.

Box I.

4.3. Notes on state-dependent importance sampling for light-tailed systems

The discussion so far in this section is taken mainly from a series of papers by [5,79,6]. It is important to note, however, that we have restricted ourselves to exponential changes-of-measure. They formulate optimal state-dependent twisting as an HJB equation that leads to a differential game and a corresponding Isaacs equation in [5]. The case for uniformly recurrent Markov chain is studied in [79], and the general method of finding subsolution in a variety of examples is shown in [6]. Moreover, [7] studies the subsolution approach to the tandem queue problem spotted by [1], as well as some generalizations. Importance sampling for Jackson network using similar approach is done in [80]. The use of viscosity solution in large deviations analysis of queueing networks can be seen in [81]. Recently, rare event simulation estimators for measure-valued light-tailed systems are developed in [55,82]. Developing Isaacs equations in these settings appears to be a very challenging task.

Other related approaches and general conditions for asymptotically optimal importance samplers are studied in [83,46,84,48]. Sharper analysis (beyond weak efficiency) of state-dependent (subsolution based) importance sampling estimators is given in [55].

5. Large deviations and importance sampling for heavy-tailed random walks

We now move to the study of Monte Carlo methods for random walks with heavy-tailed increments. Large and moderate deviations for heavy-tailed random walks are studied in the classical papers of [85–90]. More recent work on first passage and other related quantities include [91–94]. The paper [95] studies conditional limit theorems and their applications to insurance risk analysis.

Let us start by discussing classical simulation techniques in this setting. This discussion will allow us to explain the basic principles that govern the occurrence of rare events when heavy tails are present and it will also allow us to set the stage for state-dependent techniques.

5.1. Classical simulation techniques for heavy-tailed random walks

Let us start by collecting some useful definitions related to heavy-tailed systems (see, for instance, [96]).

Definition 5. A non-negative random variable X is said to be subexponential if given m independent copies $X_1, X_2, \ldots, X_m$ of X we have $P(X_1 + X_2 + \cdots + X_m > x) \sim mP(X > x)$ as $x \nearrow \infty$.

The subexponential property, as we shall see, provides insight into how large deviations tend to occur in heavy-tailed models.

Proposition 1. *Suppose that $S_m = X_1 + \cdots + X_m$ where the X_i's are iid non-negative subexponential rv's.*

$$P\left(\max_{1\le j\le m} X_j > n | S_m > n\right) \longrightarrow 1 \tag{31}$$

as $n \nearrow \infty$. Moreover, for each Borel set $A \subset R^d$, define $\widehat{P}_n(\cdot)$ via

$$\widehat{P}_n((X_1, \ldots, X_m) \in A) = \sum_{j=1}^{m} P((X_1, \ldots, X_m) \in A | X_j > n)\frac{1}{m},$$

then

$$\sup_A |P((X_1, \ldots, X_m) \in A | S_m > n) - \widehat{P}_n((X_1, \ldots, X_m) \in A)| \longrightarrow 0$$

as $n \nearrow \infty$.

Proof. The first part follows by definition of subexponentiality. Indeed, it follows by Bonferroni inequalities that

$$P(S_m > x) \sim m\overline{F}(x) \sim P\left(\max_{1\le j\le m} X_j > x\right) \quad \text{as } x \nearrow \infty. \tag{32}$$

For the second part, first note that

$$\begin{aligned} &P((X_1, \ldots, X_m) \in A | S_m > n) \\ &= P\left((X_1, \ldots, X_m) \in A | \max_{1\le j\le m} X_j > n\right) \\ &\quad \times \frac{P\left(\max_{1\le j\le m} X_j > n\right)}{P(S_m > n)} \\ &\quad + \frac{P\left((X_1, \ldots, X_m) \in A, \max_{1\le j\le m} X_j \le n, S_m > n\right)}{P(S_m > n)}. \end{aligned}$$

Noting that

$$\begin{aligned} &\frac{P\left((X_1, \ldots, X_m) \in A, \max_{1\le j\le m} X_j \le n, S_m > n\right)}{P(S_m > n)} \\ &\le \frac{P\left(\max_{1\le j\le m} X_j \le n, S_m > n\right)}{P(S_m > n)} \to 0 \end{aligned}$$

as $n \to \infty$ and using (32) we conclude that

$$\left| P((X_1, \ldots, X_m) \in A | S_m > n) - P\left((X_1, \ldots, X_m) \in A | \max_{1\le j\le m} X_j > n\right) \right| \to 0$$

uniformly over all Borel sets A as $n \to \infty$. Similarly, it follows that

$$\sup_A \left| P\left((X_1, \ldots, X_m) \in A | \max_{1\le j\le m} X_j > n\right) - \widehat{P}_n((X_1, \ldots, X_m) \in A) \right| \longrightarrow 0$$

and therefore we conclude our result. □

The previous proposition illustrates the so-called "catastrophe principle" behind extreme behavior of heavy-tailed systems, which postulates that large deviations are caused by extremes in one or few components. In this case, the sum is large

because one single component, namely the maximum, is large. In previous sections we saw that a convenient change-of-measure for importance sampling is suggested by studying the asymptotic conditional distribution of the underlying process given the rare event of interest. Now, consider applying the idea behind such conditional description to a simple example. Suppose that we are interested in estimating $P(S_m > n)$ efficiently as $n \nearrow \infty$. Proposition 1 indicates that an asymptotic conditional description of the X_j's given $S_m > n$ assigns zero mass to sample paths for which all the random variables are less than n. As a consequence, the natural asymptotic description of the X_j's given $S_m > n$ is singular with respect to the nominal (original) distribution and therefore, contrary to the light-tailed case, a direct importance sampling approach is not feasible. This feature was observed by [4], which provides an extended discussion of the difficulties that are inherent to the design of efficient simulation estimators for heavy-tailed systems.

Perhaps the most relevant special subclass of subexponential random variables is given by those whose right tail is regularly varying, which we define next.

Definition 6. A random variable X with tail distribution $\overline{F}(\cdot) = P(X > \cdot) = 1 - F(\cdot)$ has a regularly varying right tail with index $\alpha > 0$ if $\overline{F}(\beta x)/\overline{F}(x) \longrightarrow \beta^{-\alpha}$ as $x \nearrow \infty$ for any $\beta > 0$. Similarly, X has a regularly varying left tail with index $\alpha > 0$ if $-X$ has a regularly varying right tail with index $\alpha > 0$.

A function $L(\cdot)$ is slowly varying at infinity if $L(\beta x)/L(x) \longrightarrow 1$ as $x \to \infty$ for each $\beta > 0$. Therefore, if X has regularly varying right tail with index $\alpha > 0$ we often write $\overline{F}(t) = t^{-\alpha}L(t)$.

In addition to regularly varying distributions, subexponential random variables include models that incorporate Weibull-type tails with index $\gamma \in (0, 1)$, which satisfy $\overline{F}(x) \sim c\exp(-\beta x^{\gamma})$ for $c, \beta > 0$ as $x \nearrow \infty$. Other popular special models include log-normal, log-gamma and t distributions (the last two models are particular cases of regular variation).

We summarize a few important properties of regularly varying distributions that will be used in our future development; see for instance, [97] or [96].

Theorem 4. *Suppose that X possesses a regularly varying right tail with index $\alpha > 0$; in particular, $P(X > t) := \overline{F}(t) = t^{-\alpha}L(t)$. Then,*
(i) *(Karamata's theorem) If $\alpha > 1$, then*

$$\int_t^{\infty} \overline{F}(s)\mathrm{d}s = \frac{t^{-\alpha+1}L(t)}{\alpha - 1}(1 + o(1))$$

as $t \to \infty$.

(ii) *(Breiman's theorem) Let Z be positive random variable such that $EZ^{\alpha+\varepsilon} < \infty$ for some $\varepsilon > 0$, then*

$$P(ZX > t) = t^{-\alpha}L(t)EZ^{\alpha}(1 + o(1))$$

as $t \to \infty$.

(iii) *(Pareto conditional excess tail) For any $x > 0$ we have that*

$$P(X - nx > nt | X > nx) \to \frac{1}{(1 + t/x)^{-\alpha}}$$

as $n \to \infty$.

(iv) *(Long tails) For any $x \in R$*

$$\frac{P(X > t + x)}{P(X > t)} \longrightarrow 1$$

as $t \to \infty$.

The first provably efficient rare-event simulation algorithms for heavy-tailed random walk problems focused on the problem of estimating

$$\alpha_n = P(S_m > n), \tag{33}$$

efficiently as $n \to \infty$. This calculation is motivated by the ruin problem explained in Example 2 in the case of Poisson arrivals. In this case, there exists a representation, known as Pollaczek–Khintchine's formula (see [30, p. 237]) which allows to express the probability of eventual ruin, given that the initial reserve of the company is n, as $P(S_M > n)$, where M is a suitably defined geometric random variable independent of an appropriately defined random walk $(S_k : k \geq 1)$.

The first weakly efficient algorithm for estimating (33) was designed by [98] for regularly varying increments and it was based on a conditional Monte Carlo idea (see also [4]). They note that it is straightforward to have direct access to the distribution of $X_{(m)} := \max_{1 \leq j \leq m} X_j$. Using the fact that the event $\{S_m > n\}$ belongs to the single jump domain as $n \to \infty$ it is natural to condition on the first $m - 1$ order statistics $X_{(1)}, \ldots, X_{(m-1)}$ corresponding to the X_j's ($j \leq m$) (i.e. integrating out the contribution of $X_{(m)}$ which is the most relevant one). Following this idea, [98] provided the following result.

Proposition 2. *Assume that X_1 has a density and a regularly varying right tail with index $\alpha > 0$. Define $S_{(m-1)} = X_{(1)} + \cdots + X_{(m-1)}$ and set*

$$Z_0(n) = P\left(S_m > n | X_{(1)}, \ldots, X_{(m-1)}\right) = \frac{\overline{F}\left(\left(n - S_{(m-1)}\right) \vee X_{(m-1)}\right)}{\overline{F}\left(X_{(m-1)}\right)}.$$

Then, $Z_0(n)$ is a logarithmically efficient estimator for $P(S_m > n)$ as $n \nearrow \infty$.

Proof. Note that

$$EZ_0(n)^2 = E[Z_0(n)^2 ; X_{(m-1)} \leq n/2] + E[Z_0(n)^2 ; X_{(m-1)} > n/2].$$

Now, let us denote the density of X_1 by $f(\cdot)$. It follows that the density of $X_{(m-1)}$, which we shall denote by $f_{(m-1)}(\cdot)$ satisfies

$$f_{m-1}(x) = m(m-1)F(x)^{m-2}\overline{F}(x)f(x).$$

Therefore,

$$E(Z_0(n)^2 ; X_{(m-1)} > n/2) \leq P\left(X_{(m-1)} > n/2\right) = O(\overline{F}(n)^2)$$

as $n \nearrow \infty$. In addition,

$$E(Z_0(n)^2 ; X_{(m-1)} \leq n/2) = \int_0^{n/m} E(Z_0(n)^2 ; X_{(m-1)} \in \mathrm{d}x) + \int_{n/m}^{n/2} E(Z_0(n)^2 ; X_{(m-1)} \in \mathrm{d}x).$$

Observe that on $X_{(m-1)} < n/m$ we have $S_{(m-1)} < (m-1)n/m$ and therefore (by Karamata's theorem)

$$\int_0^{n/m} E(Z_0(n)^2 ; X_{(m-1)} \in \mathrm{d}x) \leq \int_0^{n/m} \frac{m^2\overline{F}(n/m)^2 f(x)}{\overline{F}(x)}\mathrm{d}x = O(\overline{F}(n)^2 \log(n)).$$

The last integral piece (from n/m to $n/2$) is handled as in the previous display. Logarithmic efficiency then follows as a result of the previous estimates. □

Regular variation has been used extensively in the previous result, so it may not be surprising that the estimator $Z_0(n)$, in the presence of other types of heavy-tailed assumptions such as Weibull tails, fails to be logarithmic efficient. Improved Monte Carlo estimators have been recently proposed by [99]. In particular, they observe that in order to reduce some uncertainty, one can also consider the index corresponding to the largest jump and note that

$$\begin{aligned} P(S_m > n) &= \sum_{j=1}^{m} E\left(P\left(S_m > n, X_{(m)} = X_j | X_1, \right.\right. \\ &\qquad \left.\left. \ldots, X_{j-1}, X_{j+1}, \ldots, X_m\right)\right) \\ &= mE\left(P\left(S_m > n, X_{(m)} = X_m | X_1, \ldots, X_{m-1}\right)\right). \end{aligned}$$

Therefore, a natural conditional Monte Carlo estimator that one can consider is

$$\begin{aligned} Z_1(n) &= mP\left(S_m > n, X_{(m)} = X_m | X_1, \ldots, X_{m-1}\right) \\ &= m\overline{F}\left((n - S_{m-1}) \vee \max(X_j : j \le m-1)\right). \end{aligned}$$

The paper [99] proves the following result (the analysis is similar to that given in the proof of Proposition 2 and therefore the details are omitted).

Proposition 3. *Assume that X_1 has a density and regularly varying right tails. Then, the estimator $Z_1(n)$ is strongly efficient as $n \nearrow \infty$. In addition, if X_1 has Weibull-type tails with index $\gamma \in (0, .58)$ then $Z_1(n)$ is logarithmically efficient.*

The previous proposition indicates that $Z_1(n)$ can only be guaranteed to be efficient if γ is sufficiently small. In the design of efficient estimators in the heavy-tailed setting it often occurs that models with less heavy tails than, say regularly varying, tend to require more and more information of all the components (not only the largest one). It is partly due to this feature that the majority of the efficient estimators developed for heavy-tailed systems assume special characteristics (such as regular variation or Weibullian-type tails). Notable exceptions are the algorithms by Blanchet and Glynn [9], Blanchet and Li [123] which are designed to work basically for all subexponential distributions. Additional conditional Monte Carlo algorithms have been proposed for the transient distribution of an $M/G/1$ queue by [100]. Their estimator is proved to be strongly efficient for regularly varying input.

The paper [4] discusses the difficulties of applying importance sampling to heavy-tailed problems—an important problem concerns the singularity issue that we mentioned in the paragraph after the proof of Proposition 1. Nevertheless, they also study ideas that give rise to provably efficient algorithms. For instance, they note that if the X_j's have a density and regularly varying tails with index $\alpha > 0$, then a logarithmically efficient importance sampling scheme for estimating $P(S_m > n)$ is obtained by sampling the X_j's in an iid fashion according to the tail distribution $\widetilde{P}\left(X_j > x\right) = 1/\log(e + x)$ for $x \ge 0$. Logarithmic efficiency for this estimator is very easy to obtain by means of Karamata's theorem (see, for instance, [23, p. 176]). This selection of importance sampling distribution biases the increments to induce very heavy-tailed distributions and therefore over-samples such sample paths for which several (and not only the maximum) components contribute to the occurrence of the rare event. However, this importance sampler, although logarithmically efficient, does not seem to perform well in practice as reported by [23, p. 176].

Another importance sampling approach was suggested by [101] and is based on applying exponential tilting type ideas via the hazard rate corresponding to the X_j's. A basic observation behind Juneja and Shahabuddin's hazard rate tilting approach is the fact that if the X_j's have a positive density, then $P\left(X_j > t\right) = \exp(-\Lambda(t))$, where $\Lambda(t) = \int_0^t \lambda(s)\,ds$ and $\lambda(\cdot)$ is the hazard rate of X_j. In particular, if T is exponentially distributed with mean 1, then $\Lambda^{-1}(T)$ is a copy of X_j and therefore, for appropriate θ, we can define hazard rate tilting densities, $f_\theta(\cdot)$, via

$$f_\theta(x) = \frac{\exp(\theta\Lambda(x)) f(x)}{E\exp\left(\theta\Lambda\left(X_j\right)\right)} = \exp(\theta\Lambda(x)) f(x)(1-\theta).$$

The corresponding hazard rate importance sampling estimator for $P(S_m > n)$ takes the form

$$Z_2(n) = \frac{\exp\left(-\sum_{j=1}^{m}\theta\Lambda\left(X_j\right)\right)}{(1-\theta)^m} I(S_m > n).$$

Using $E_\theta(\cdot)$ to denote the expectation operator assuming that the X_j's are iid with density $f_\theta(\cdot)$ we obtain

$$\begin{aligned} E_\theta(Z_2(n)^2) &= (1-\theta)^{-2m} E_\theta \\ &\quad \times \left[\exp\left(-2\sum_{j=1}^{m}\theta\Lambda\left(X_j\right)\right); S_m > n\right]. \end{aligned}$$

Assuming that $\Lambda(\cdot)$ is a concave function (as is the case for standard Pareto and Weibull random variables), then we obtain that if $S_m > n$ then

$$\sum_{j=1}^{m}\Lambda\left(X_j\right) \ge \Lambda\left(\sum_{j=1}^{m} X_j\right) = \Lambda(S_m) \ge \Lambda(n) \tag{34}$$

and therefore, if $\theta \ge 0$, we obtain

$$E_\theta(Z_2(n)^2) \le (1-\theta)^{-2m}\exp(-2\theta\Lambda(n)).$$

Taking $\theta = 1 - \eta/\Lambda(n)$ for some $\eta > 0$ yields the following result.

Proposition 4. *If the X_j's are non-negative and have a concave cumulative hazard rate function $\Lambda(\cdot)$ then $Z_2(n)$ is a logarithmically efficient estimator for $P(S_m > n)$ as $n \nearrow \infty$.*

The previous hazard rate tilting strategy has been improved for the case in which the number of increments follows a geometric distribution as in the $M/G/1$ setting. Such approaches involve suitable translation of the function $\Lambda(\cdot)$ applied when tilting. This adjustment is convenient in order to apply the concavity argument given in (34) at a more convenient location relative to the rare event $S_m > n$ (see [50] and references therein). The hazard rate tilting idea has inspired further studies in the field because it tries to develop rare-event simulation methodology through a structure that resembles that of light-tailed input systems (via exponential tiltings). Nevertheless, virtually all estimators that take advantage of this idea utilize the random walk structure substantially and some sort of subadditivity argument on $\Lambda(\cdot)$ as we did in (34).

5.2. Large deviations for heavy-tailed random walks

As we illustrated by means of the catastrophe principle, the large deviations theory for heavy-tailed random walks has a completely different qualitative characteristics than its light-tailed counter-part. Intuitively, one can explain this because in light-tailed situations large deviations occur, as we have seen, because of a "conspiratorial effect" among all the increments; each of them follows a specific biased (exponentially tilted) distribution that tracks the "optimal" trajectory. In the heavy-tailed setting, in contrast, large deviations occur because only one or few jumps contribute to the occurrence of the rare event; the rest of the jumps evolve according to the original / nominal distribution. This intuition has been made rigorous in considerable generality only in situations in which only one jump contributes to the occurrence of the rare event in the regularly varying setting [102]. (We shall

refer to this setting as "the single jump domain".) A few examples of interest in which large deviations happen by the occurrence of more than one jump ("multiple jump domain") have been rigorously studied mostly in the setting of queueing systems (see [103–105]).

Throughout the rest of our technical discussion we shall concentrate on regularly varying distributions. We will provide further remarks on other types of heavy-tailed increment distributions at the end of the section.

In spite of not having a large deviations theory that is as comprehensive as that of light-tailed systems, in the heavy-tailed case there is a powerful heuristic that allows to "guess" a plausible form for large deviations results and the associated "optimal path" to the rare event. This heuristic, which is known as the "fluid heuristic", is based on elementary Law of Large Numbers analysis and basic facts of regularly varying distributions. Since, as we shall explain, the methodology for constructing the algorithms can be used to rigorously justify (up to a constant) the heuristics, we content ourselves with explaining how to apply the fluid heuristic in a few examples and point to relevant references when available.

In order to guide our discussion we shall revisit a couple of examples studied earlier in the light tailed case.

Example 5 (*First Passage Time Probabilities For Heavy-Tailed Random Walk*)**.** We consider once again the surplus process introduced in Example 1. In this case, however, we shall assume that the claim sizes, namely the Z_j's, are regularly varying with index $\alpha > 1$.

We define $X_j = Z_j - pT_j$, where T_j (the j-th inter-arrival time), is exponentially distributed with unit mean. Assume that $EX_k = -\mu < 0$. Suppose that $S_0 = y$, let $\tau_n = \inf\{k \geq 0 : S_k > n\}$ and consider the problem of approximating the ruin probability

$$\alpha_n(y) = P(\tau_n < \infty | S_0 = y)$$

as $n \nearrow \infty$.

Because the service times are heavy-tailed we expect that only a few jumps would contribute to the rare event. By the geometry of the problem, however, it is natural to expect that only one jump will contribute to ruin and thus we are in the single jump domain.

The fluid heuristic consists in substituting the random walk by its fluid or Law of Large Numbers behavior prior to the big jump. The associated fluid path in this case is $S_{\lfloor nt \rfloor}/n \approx y(t) = y - \mu t$ and therefore the fluid heuristic suggests

$$\begin{aligned}\alpha_n(y) &\sim \sum_{k=0}^{\infty} P(X_{k+1} + y(k) > n) \\ &\sim \int_0^{\infty} P(X > n + \mu t - y)\, \mathrm{d}t\end{aligned}$$

as $n \nearrow \infty$, thereby neglecting the contribution of more than one jump and fluctuations beyond the Law of Large Numbers.

Letting $u = n + \mu t - y$ we obtain

$$\alpha_n(y) \sim \frac{1}{\mu} \int_{n-y}^{\infty} P(X > u)\, \mathrm{d}y.$$

as $n \nearrow \infty$. This approximation turns out to be correct even beyond the assumption of regularly varying service times (see [106–109]).

We now discuss an example in the multiple jump domain.

Example 6 (*A Multiple Jump Large Deviations Event*)**.** We revisit Example 2, now in the context of heavy-tailed random variables. In particular, we assume that $(X_k : k \geq 1)$ is a sequence of iid random variables t distributed with $v > 1$ degrees of freedom and we are interested in approximating the probability

$$\alpha_n = P[\min_{0 \leq k \leq n} S_k \leq an, S_n > nb]$$

as $n \to \infty$, where $S_k = X_1 + \cdots + X_k$ and $a < 0 < b$.

The associated fluid path in this case is $S_k/n \approx y(k) := 0$ and therefore, given the geometry of the problem, it is not difficult to convince ourselves that in the current situation the most likely path leading to the rare event involves two jumps, therefore, we are in the multiple jump domain.

The fluid heuristic then suggests

$$\begin{aligned}\alpha_n &\sim \sum_{1 \leq k_1 < k_2 \leq n} P(X_{k_1} + y(k_1 - 1) < an) \\ &\quad \times P(X_{k_2} + y(k_2) - y(k_1) + X_{k_1} > nb | X_{k_1} \\ &\quad + y(k_1 - 1) < an) \\ &\sim \frac{n(n-1)}{2} P(X_1 < an) P(X_2 + X_1 > nb | X_1 < an).\end{aligned}$$

Once again, note that we replace the path in-between jumps by the fluid trajectory and neglect the contributions of more than two jumps. We can further simplify the previous expression using the fact that for each $y < 0$,

$$P(X_1 < yn + an | X_1 < an) \to P(Z_a < y + a) := \frac{1}{(y/a + 1)^v},$$

as $n \to \infty$. The previous limit indicates that the distribution of X_1/n given that $X_1/n < a$ can be well approximated by $Z_a \leq a$ with the above distribution. This observation, combined with the heuristic approximation developed earlier in turn suggests

$$\alpha_n \sim \frac{n(n-1)}{2} P(X_1 < an) P(X_2 > n(b - Z_a)),$$

where Z_a is independent of X_2. By the symmetry of the t distribution it follows that $P(X_1 < an) = P(X_1 > |a|\, n)$. Further, since a t distribution with v degrees of freedom is regularly varying with index v, it follows that $P(X_1 > |a|\, n) = |a|^{-v} P(X_1 > n)(1 + o(1))$ as $n \to \infty$. In this case, evidently, we have that $0 < 1/(b - Z_a) < 1/(b - a)$ and therefore we can apply Breiman's theorem to conclude that

$$P(X_2 > n(b - Z_a)) = E(b - Z_a)^{-v} P(X_2 > n)(1 + o(1)),$$

as $n \to \infty$. We then conclude the following plausible asymptotic approximation,

$$\alpha_n \sim n^2 P(X_1 > n)^2 \frac{|a|^{-v}}{2} E(b - Z_a)^{-v} = \Theta(n^2 P(X_1 > n)^2),$$

as $n \to \infty$. We have not been able to locate a reference for this asymptotic result in the literature, but it is not too difficult to show rigorously its validity.

5.3. Conditional distributions and implications for simulation

As we discussed in Section 2, it is useful to understand the conditional distribution of the random walk given the occurrence of the rare event of interest because such conditional distribution provides the optimal change-of-measure (in terms of variance minimization). A natural starting point, as we followed in the light-tailed setting would be to obtain a conditional limit theorem for the X_k's given the rare event of interest. The asymptotic conditional distribution would then be a natural candidate for an importance sampling distribution. Let us pursue this idea in the context of Example 5.

Once again, we use the fluid heuristic to obtain an asymptotic result for the conditional distribution of $(X_1, \ldots, X_k)$ for any fixed $k > 0$ given that $\tau_n < \infty$ and that $S_0 = 0$ as $n \to \infty$. Given $-\infty < x_j^- < x_j^+ < \infty$ for $j \in \{1, \ldots, k\}$ we note that

$$\begin{aligned}&P\left(x_1^- \leq X_1 \leq x_1^+, \ldots, x_k^- \leq X_k \leq x_k^+ | \tau_n < \infty\right) \\ &\quad = \frac{P\left(x_1^- \leq X_1 \leq x_1^+, \ldots, x_k^- \leq X_k \leq x_k^+, \tau_n < \infty\right)}{\alpha_n(0)}\end{aligned}$$

$$= \frac{E[I(x_1^- \le X_1 \le x_1^+, \ldots, x_k^- \le X_k \le x_k^+)\alpha_n(S_k)]}{\alpha_n(0)}.$$

Note that $\alpha_n(y)$ is increasing as a function of y. Therefore, letting $s_k^- = x_1^- + \cdots + x_k^-$ and $s_k^+ = x_1^+ + \cdots + x_k^+$,

$$\begin{aligned} P(x_1^- \le X_1 \le x_1^+, \ldots, x_k^- \le X_k \le x_k^+) &\times \frac{\alpha_n(s_k^-)}{\alpha_n(0)} \\ &\le P\left(x_1^- \le X_1 \le x_1^+, \ldots, x_k^- \le X_k \le x_k^+ | \tau_n < \infty\right) \\ &\le P(x_1^- \le X_1 \le x_1^+, \ldots, x_k^- \le X_k \le x_k^+) \times \frac{\alpha_n(s_k^+)}{\alpha_n(0)}. \end{aligned} \tag{35}$$

As we have seen in Example 5,

$$\alpha_n(y) \sim G(n-y) := \frac{1}{\mu}\int_{n-y}^{\infty} P(X > t)\, dt$$

as $n - y \to \infty$. Now, as a consequence of Karamata's theorem (see for example [97, p. 17]), $G(\cdot)$ is also regularly varying (i.e. exhibits basically power-law tail behavior) and therefore it follows easily that $G(n-y)/G(n) \longrightarrow 1$ as $n \to \infty$ for each y. We then conclude (taking the limit in (35)) that the asymptotic conditional distribution of $(X_1, \ldots, X_k)$ given that $\tau_n < \infty$ remains unchanged as $n \to \infty$.

The fact that the asymptotic conditional distribution of $(X_1, \ldots, X_k)$ does not provide a useful description that can be used as an importance sampling distribution in the prelimit is one of the main reasons that makes the design of efficient importance sampling estimators for heavy-tailed systems a challenging problem. This observation lies at the center of the discussion in the paper by [4], which provides a number of counter-examples, in the spirit of the previous analysis, illustrating the challenges that arise in the design of efficient Monte Carlo methods in heavy-tailed settings. In the same vein, the papers of [3,110] show that no state-independent importance sampling estimator can achieve weak efficiency.

Let us now take a more direct approach to approximate the increment X_{k+1} given the current position $S_k = s$ of the random walk at time k and given that $k < \tau_n < \infty$. We note (adopting the notation introduced in (5)) that

$$\begin{aligned} &P(X_{k+1} \in dy | \tau_n \in [k+1, \infty), S_k = s) \\ &\quad = P(X_{k+1} \in dy, \tau_n = k+1 | \tau_n \in [k+1, \infty), S_k = s) \\ &\qquad + P(X_{k+1} \in dy, \tau_n > k+1 | \tau_n \in [k+1, \infty), S_k = s). \end{aligned}$$

Now,

$$\begin{aligned} &P(X_{k+1} \in dy, \tau_n = k+1 | \tau_n \in (k, \infty), S_k = s) \\ &\quad = p_n^*(s) \frac{P(X_{k+1} \in dy)\, I(y > n-s)}{P(X_{k+1} > n-s|s)}, \end{aligned} \tag{36}$$

where

$$p_n^*(s) = \frac{P(X_{k+1} > n-s)}{\alpha_n(s)}.$$

Using the approximation in Example 5, we have that

$$p_n^*(s) \sim p(n-s) := \mu \frac{P(X_{k+1} > n-s)}{\int_{n-s}^{\infty} P(X_{k+1} > t)\, dt} \sim \frac{\mu(\alpha-1)}{(n-s)}$$

as $n - s \to \infty$. On the other hand, we have that

$$\begin{aligned} &P(X_{k+1} \in dy, \tau_n > k+1 | \tau_n \in [k+1, \infty), S_k = s) \\ &\quad = \frac{P(X_{k+1} \in dy, \tau_n \in [k+2, \infty) | S_k = s)\, I(y \le n-s)}{\alpha_n(s)} \\ &\quad = \frac{P(X_{k+1} \in dy)\, I(y \le n-s)}{P(X_{k+1} \le n-s)} \\ &\qquad \times P(X_{k+1} \le n-s) \frac{\alpha_n(s+y)}{\alpha_n(s)}. \end{aligned} \tag{37}$$

If we apply the fluid heuristic locally, that is by replacing X_{k+1} by its mean, namely $EX_{k+1} = -\mu$, we arrive at

$$\begin{aligned} \frac{\alpha_n(s+X_{k+1})}{\alpha_n(s)} &\approx \frac{\alpha_n(s-\mu)}{\alpha_n(s)} \\ &\approx \frac{(n-s+\mu)^{-\alpha+1}}{(n-s)^{-\alpha+1}} \approx [1+\mu/(n-s)]^{-\alpha+1} \\ &= 1 - \frac{\mu(1-\alpha)}{n-s} + o\left(\frac{1}{n-s}\right). \end{aligned}$$

Therefore, applying the previous local version of the fluid heuristic into (37) and combining this with (36) we arrive at the non-rigorous approximation

$$\begin{aligned} &P(X_{k+1} \in dy | \tau_n \in [k+1, \infty), S_k = s) \\ &\quad \approx \frac{P(X_{k+1} \in dy)\, I(y > n-s)}{P(X_{k+1} \le n-s)} p(n-s) \\ &\qquad + \frac{P(X_{k+1} \in dy)\, I(y \le n-s)}{P(X_{k+1} \le n-s)} [1 - p(n-s)]. \end{aligned} \tag{38}$$

It turns out that this approximation is indeed valid in total variation [111], and it can be used to derive other conditional limit theorems, such as the conditional distribution of τ_n given that $\tau_n < \infty$. For instance, applying the fluid heuristic we have that if $\tau_n > nt$, then $S_k \approx -\mu k$ whenever $k \le nt$ and therefore

$$\begin{aligned} P(\tau_n > nt | \tau_n < \infty) &\approx \prod_{k=1}^{nt} (1 - p(n+\mu k)) \\ &\approx \exp\left(-\sum_{k=1}^{nt} \frac{\mu(\alpha-1)}{n+\mu k}\right) \\ &\approx \exp\left(-\int_0^{nt} \frac{\mu(\alpha-1)}{n+\mu s} ds\right) \\ &= \frac{1}{(1+\mu t)^{\alpha-1}} \end{aligned} \tag{39}$$

as $n \to \infty$. This result and analogous joint approximation results involving τ_n and other quantities of interest such as the overshoot and undershoot at the time of ruin have been derived, using ladder height theory, by [95]; related extensions in continuous time are given in [112]. The heuristic approach taken to derive (39) and other conditional limit theorems is made rigorous in [111]. It is important to note that if $1 < \alpha < 2$, then the asymptotic conditional distribution of the normalized hitting time τ_n/n given that $\tau_n < \infty$ has infinite mean.

5.4. State-dependent importance sampling techniques for heavy-tailed random walks

In the previous section we argued, using a local version of the fluid heuristic, that a family of importance sampling distributions based on a simple mixture appears suitable to design provably efficient estimators. The family suggested by (38) is a mixture of two components, one involving the occurrence of the rare event in question and a second component corresponding to the nominal/original distribution of the increment distribution. The form of the mixture is consistent with the way in which rare events occur in a heavy-tailed setting. This type of mixture family was introduced by [113] in the context of estimating $P(S_m > n)$ as $n \to \infty$ for fixed m. In this context they were able to design a strongly efficient estimator for $P(S_m > n)$. In fact, their estimator was shown to give a relative error of at most $\varepsilon > 0$ for any fixed $\varepsilon > 0$. Their proof technique is based on a weak convergence analysis, using Proposition 1, which is facilitated substantially because m is held fixed and independent of n. We now quickly illustrate the approach proposed by [113].

Example 7 (*Weak Convergence Approach And State-Dependent Importance Sampling*)**.** We shall estimate $\alpha_n = P(S_m > n)$ efficiently as $n \to \infty$ for fixed m. For simplicity we shall assume that the X_k's are non-negative random variables with density $f(\cdot)$ and tail distribution function $\overline{F}(\cdot)$. The family of importance sampling distributions that we consider is the following. Given that $S_{k-1} = s$ for $1 \leq k < m$, then the next increment is sampled according to the density

$$f_k(x|s) = \frac{p_k(s) f(x) I(x > a(n-s))}{\overline{F}(a(n-s))} + \frac{(1-p_k(s)) f(x) I(x \leq a(n-s))}{F(a(n-s))}, \tag{40}$$

where $p_k(s) \in (0,1)$ is selected appropriately and $a \in (0,1)$. The parameter a, as we shall see, is introduced in order to facilitate a dominated convergence argument. If $k = m$ (i.e. $S_{m-1} = s$), then the increment is sampled according to the law of X_m given that $X_m > n - s$. Dupuis et al. [113] obtains a limiting control problem that allows to optimally select the p_k's. Note that $p_k(s)$ represents the probability that the rare event is caused in the k-th step given that $S_{k-1} = s < n$. Therefore, if $k \leq m-1$, we would like to select

$$\begin{aligned} p_k(s) &\approx P(S_k > n | S_m > n, S_{k-1} = s) \\ &= \frac{P(X_k > n-s)}{P(S_m > n | S_{k-1} = s)} \approx \frac{P(X_k > n-s)}{(m-k+1)\overline{F}(n-s)} \\ &\approx \frac{1}{m-k+1}. \end{aligned}$$

Motivated by this observation, we shall propose $p_k(s) = 1/(m-k+1)$ for $1 \leq k \leq m-1$ if $s \leq n$ and $p_k(s) = 1$ is $s > n$. We then conclude that the importance sampling estimator, R_n, satisfies

$$\begin{aligned} &R_n(X_1, \ldots, X_m) \\ &= \frac{f(X_1)}{f_1(X_1|S_0)} \times \frac{f(X_2)}{f_2(X_2|S_1)} \times \cdots \\ &\quad \times \frac{f(X_m)}{f_m(X_m|S_{m-1})} I(S_m > n) \\ &= \prod_{k=1}^{m-1} \left[\frac{\overline{F}(a(n-S_{k-1})) I(X_k > a(n-S_{k-1}))}{p_k(S_{k-1})} \right. \\ &\quad \left. + \frac{F(a(n-S_{k-1})) I(X_k \leq a(n-S_{k-1}))}{(1-p_k(S_{k-1}))} \right] \\ &\quad \times \left[I(S_{m-1} < n) \overline{F}(a(n-S_{m-1})) + I(S_{m-1} \geq n) \right]. \end{aligned}$$

Let $\widetilde{P}(\cdot)$ be the probability measure induced by the proposed importance sampling distribution; we use $\widetilde{E}(\cdot)$ to denote the associated expectation operator. Then

$$\begin{aligned} &\frac{1}{P(S_m > n)^2} \widetilde{E}[R_n(X_1, \ldots, X_m)^2] \\ &= \frac{1}{P(S_m > n)} E[R_n(X_1, \ldots, X_m) | S_m > n]. \end{aligned}$$

We wish to use Proposition 1 to approximate the previous expectation. In order to do this we first will argue that $R_n/P(S_m > n)$ is bounded by some constant uniformly as $n \to \infty$. Note that there exists $1 \leq k \leq m$ such that $X_k > a(n - S_{k-1})$. Otherwise, if for all $1 \leq k \leq m$, we have that $S_k - S_{k-1} = X_k \leq a(n - S_{k-1})$ or equivalently we have that

$$S_k \leq an + (1-a) S_{k-1},$$

which iterating yields

$$\begin{aligned} S_k &\leq an + (1-a)[an + (1-a) S_{k-2}] \\ &\leq an + a(1-a)n + a(1-a)^2 n + \cdots + a(1-a)^{k-1} n \\ &\leq n(1 - (1-a)^m) < n. \end{aligned}$$

Now let $\upsilon_n = \inf\{k \geq 1 : X_k > a(n - S_{k-1})\}$ and note that on $\upsilon_n = i$ we have that $S_{i-1} \leq n(1-(1-a)^m)$

$$\begin{aligned} R_n I(\upsilon_n = i) &\leq \overline{F}\left(a(1-a)^m n\right) 2^m I(\upsilon_n = i) \\ &\quad \times \prod_{k=1}^{m-1} \frac{1}{\min(1/(m-k+1), (1-1/(m-k+1)))}. \end{aligned}$$

Therefore, we have that

$$\begin{aligned} \frac{R_n(X_1, \ldots, X_m)}{P(S_m > n)} &\leq \frac{\overline{F}(a(1-a)^m n)}{P(S_m > n)} \\ &\quad \times \prod_{k=1}^{m-1} \frac{2}{\min(1/(m-k+1), (1-1/(m-k+1)))} = O(1) \end{aligned}$$

as $n \to \infty$. We then conclude thanks to Proposition 1 that

$$\begin{aligned} &\frac{1}{P(S_m > n)} E[R_n(X_1, \ldots, X_m) | S_m > n] \\ &= \sum_{j=1}^{m} \frac{E[R_n(X_1, \ldots, X_m) | X_j > n]}{m P(S_m > n)} (1 + o(1)) \end{aligned}$$

as $n \to \infty$. Evaluating the previous expectation we conclude that

$$\begin{aligned} &\sum_{j=1}^{m} \frac{E[R_n(X_1, \ldots, X_m) | X_j > n]}{m P(S_m > n)} \\ &\quad \to \frac{a^{-\alpha}}{m^2} \left(\sum_{j=1}^{m} \frac{1}{1/(m-j+1)} \times \prod_{k=1}^{j-1} \frac{1}{(1-1/(m-k+1))} \right) \\ &\quad = a^{-\alpha}. \end{aligned}$$

We can select $a \in (0,1)$ arbitrarily close to one in order to get an asymptotically negligible relative error. Note however, that it is crucial to select $a \in (0,1)$ in order to ensure that $R_n/P(S_m > n)$ remains bounded. This suggests that $a \approx 1$ might not be a good option for a pre-determined n.

As it can be appreciated in the previous development, the fact that m is bounded substantially simplifies the use of the weak convergence approach. An alternative technique, which can be directly applied to Examples 7 and 8 and leverages the fluid heuristic is the use of Lyapunov inequalities, which were introduced in [9] and we discuss in the next lemma. This technique is closely related and in some sense parallel to the subsolution approach explained earlier in the light-tailed setting.

Lemma 2. *Suppose that there exists $r(x,y) > 0$ satisfying*

$$\int r^{-1}(x,y) P(S_j \in \mathrm{d}y | S_{j-1} = x) = 1$$

for all x and $j \in \{1, 2, \ldots\}$. Then, we can define a Markov transition kernel via $K(x, \mathrm{d}y) = r^{-1}(x,y) P(S_j \in \mathrm{d}y | S_{j-1} = x)$ and an importance sampling estimator of the form

$$R = \prod_{j=1}^{T_A \wedge T_B} r(S_{j-1}, S_j) I(T_A < T_B).$$

Suppose that there exists a non-negative function $G(\cdot)$ and a constant $\rho \in (0, \infty)$ such that

$$E[r(x, S_1) G(S_1) | S_0 = x] \leq G(x) \tag{41}$$

for $x \notin A \cup B$ and $G(x) \geq \rho$ for $x \in A$. Then, $E[R | S_0 = x] \leq G(x)/\rho$.

The proof of the previous result is completely analogous to Lemma 1 and therefore is omitted. We now explain how to apply the previous result, combined with the fluid heuristics to design efficient importance sampling estimators in the context of Example 5.

Example 8 (*Lyapunov Inequalities And State-Dependent Importance Sampling*)**.** We shall apply importance sampling following the analysis in the previous section. We distinguish two situations, first, if the current position s is sufficiently far away from the boundary n, say if $n-s>\Delta$ for some $\Delta>0$, then the intuition developed in the previous section is accurate and then we apply importance sampling using the mixture family.

However, in the second situation, if s is sufficiently close to n we do not need to apply importance sampling. Also, if we use the previous family when s is sufficiently close to n, we might create variance because the mixture sampling distribution is derived using the large deviations regime, so turning off the sampling when s is sufficiently close to n is a sensible defense mechanism. We introduce a boundary layer $0\le n-s\le\Delta$ for some $\Delta>0$ and if s lies in this layer no importance sampling is applied.

Our family, then, takes the form

$$\begin{aligned}&\widetilde{P}(X_{k+1}\in \mathrm{d}y|S_k=s)\\&=\frac{P\left(X_{k+1}\in \mathrm{d}y\right)I(y>a(n-s))}{P(X_{k+1}>a(n-s))}p(n-s)I(n-s>\Delta)\\&\quad+\frac{P\left(X_{k+1}\in \mathrm{d}y\right)I(y\le a(n-s))}{P\left(X_{k+1}\le a(n-s)\right)}\\&\quad\times[1-p(n-s)]I(n-s>\Delta)\\&\quad+P\left(X_{k+1}\in \mathrm{d}y\right)I(n-s\le\Delta).\end{aligned}$$

The parameter $a\in(0,1)$ is selected in order to enforce uniform integrability conditions, just as we noted in Example 7. Following the intuition developed earlier we propose

$$p(n-s)=\frac{\theta P(X>a(n-s))}{\int_{n-s}^{\infty}P(X>t)\mathrm{d}t}$$

for some $\theta>0$. In fact, we expect $\theta\approx\mu$ to be optimal in the sense of variance minimization.

For notational convenience define

$$H(n-s)=\int_{n-s}^{\infty}P(X>t)\mathrm{d}t.$$

We now need to construct a valid Lyapunov function. We wish to prove strong efficiency, so it is natural to suggest

$$G(n-s)=\min(\kappa H^2(n-s),1)=O(\alpha_n(s)^2)$$

for some $\kappa>0$. The strategy is then to select (given $a\in(0,1)$) θ, κ and Δ in order to satisfy the Lyapunov inequality. If we can show that there exists a selection of these parameters that satisfies (41), then, given that the corresponding boundary condition in Lemma 2, namely

$$\begin{aligned}G(n-s)\ge\rho&:=\min\left(\kappa\int_0^{\infty}P(X>t)\mathrm{d}t,1\right)\\&=\min(\kappa EX^{+},1)>0\end{aligned}$$

is satisfied, we would conclude strong efficiency of the estimator.

Note that the inequality (41) holds trivially if $G(n-s)=1$. Indeed, in this case, since $G(x)\le1$ for every $x\in R$, the inequality (41) takes the form

$$EG(n-s-X)\le G(n-s)=1.$$

Now observe that $G(n-s)<1$ holds if and only if $H(n-s)<1/\kappa^{1/2}$, which in turn holds if and only if $n-s>H^{-1}(1/\kappa^{1/2})$. Therefore we can simply choose $\Delta=H^{-1}(1/\kappa^{1/2})$ and simply select θ and κ. In other words, we shall apply importance sampling using the mixture family if and only if our current position s is such that $G(n-s)<1$.

In order to proceed with the verification of inequality (41) we henceforth assume that $G(n-s)<1$. In this case inequality (41) takes the form

$$J_1+J_2\le1,\tag{42}$$

where

$$\begin{aligned}J_1&=\frac{E[G(n-s-X);X>a(n-s)]P(X>a(n-s))}{\kappa H^2(n-s)p(n-s)},\\J_2&=\frac{E[G(n-s-X);X\le a(n-s)]P(X\le a(n-s))}{\kappa H^2(n-s)[1-p(n-s)]}.\end{aligned}$$

We first bound the term J_1. Observe, noting that $G(\cdot)\le1$, and using the form of $p(n-s)$, we obtain that

$$J_1\le\frac{P(X>a(n-s))}{\theta\kappa H(n-s)}=\frac{p(n-s)}{\theta^2\kappa}.$$

Then we analyze the term J_2. Note that

$$J_2\le\frac{E[H^2(n-s-X);X\le a(n-s)]}{H^2(n-s)[1-p(n-s)]}.$$

Using the following Taylor development with remainder, which comes from the elementary formula using $t=x+yu$

$$\begin{aligned}H^2(x+y)-H^2(x)&=\int_x^{x+y}2H(t)\dot{H}(t)\,\mathrm{d}t\\&=2E[yH(x+yU)\dot{H}(x+yU)],\end{aligned}$$

where U is uniformly distributed in [0,1]. Therefore, we have, letting $x=(n-s)$ and $y=-X$,

$$\begin{aligned}&\frac{E[H^2(n-s-X)I(X\le a(n-s))]}{H^2(n-s)}\\&=P(X\le a(n-s))-2\\&\quad\times\frac{E[X\dot{H}(n-s-XU)H(n-s-XU);X\le a(n-s)]}{H^2(n-s)},\end{aligned}$$

where X and U are independent. Now,

$$\begin{aligned}&-\frac{E[X\dot{H}(n-s-XU)H(n-s-XU);X\le a(n-s)]}{H^2(n-s)}\\&=\frac{E[X\overline{F}(n-s-XU)H(n-s-XU);X\le a(n-s)]}{H^2(n-s)}.\end{aligned}$$

We are assuming that $G(n-s)<1$, or equivalently, that $n-s>H^{-1}(1/\kappa^{1/2})=\Delta$. On the other hand, we have that

$$X\frac{\overline{F}(n-s-XU)H(n-s-XU)I(X\le a(n-s))}{H^2(n-s)}\sim X\frac{\alpha-1}{(n-s)}$$

almost surely as $n-s\nearrow\infty$ and also we have, because of regular variation and Karamata's theorem that there exists a constant $K\in(0,\infty)$ such that

$$\begin{aligned}&|X|\frac{\overline{F}(n-s-XU)H(n-s-XU)I(X\le a(n-s))}{H^2(n-s)}\\&\le|X|\frac{\overline{F}((n-s)(1-a))H((n-s)(1-a))}{H^2(n-s)}\le\frac{K|X|}{n-s}.\end{aligned}$$

Consequently, if κ (or equivalently $\Delta=H^{-1}(1/\kappa^{1/2})$) is chosen sufficiently large, we conclude, applied the dominated convergence theorem that for each $\varepsilon>0$

$$J_2\le1+2EX\frac{(\alpha-1)}{n-s}(1-\varepsilon)$$

as long as $G(n-s) < 1$. Combining our estimates for J_1 and J_2 and choosing κ sufficiently large so that

$$p(n-s) \leq \theta a^{-\alpha} \frac{(\alpha-1)(1+\varepsilon)}{n-s},$$
$$\frac{1}{1-p(n-s)} \geq 1 + \theta a^{-\alpha} \frac{(\alpha-1)(1-\varepsilon)}{n-s},$$

we arrive at

$$\begin{aligned} J_1 + J_2 &\leq \frac{(\alpha-1)(1+\varepsilon)}{\theta\kappa(n-s)} a^{-\alpha} + \left(1 - 2\mu\frac{(\alpha-1)}{n-s}(1-\varepsilon)\right) \\ &\quad \times \left(1 + \theta a^{-\alpha}\frac{(\alpha-1)(1-\varepsilon)}{n-s}\right) \\ &\leq a^{-\alpha}\frac{(\alpha-1)(1+\varepsilon)}{\theta\kappa(n-s)} + \left(1 - 2\mu\frac{(\alpha-1)}{n-s}(1-\varepsilon)\right. \\ &\quad \left. + \theta a^{-\alpha}\frac{(\alpha-1)(1-\varepsilon)}{n-s}\right). \end{aligned}$$

In order to enforce (42) we then need to ensure that

$$a^{-\alpha}\frac{1+\varepsilon}{\theta\kappa} - 2\mu(1-\varepsilon) + a^{-\alpha}\theta(1-\varepsilon) < 0,$$

which clearly feasible given that $\mu > 0$. Note that it is crucial that $-\mu = EX < 0$, otherwise it would be impossible to select the parameters appropriately. This makes sense given that when $EX \geq 0, \alpha_n(y) = 1$ and it would be therefore impossible to bound the second moment of the estimator by a function that decreases to zero as $n \to \infty$, as we are proposing.

Note that since $\varepsilon > 0$ is arbitrarily small, κ can be selected arbitrarily large and a can be chosen arbitrarily close to one, note that the feasible region for parameter selection for θ becomes

$$\theta < 2\mu.$$

It turns out that by selecting $\theta = \mu$ we can make the relative error arbitrarily small. Selecting $\theta \approx 0$ implies a low frequency of the jumps and therefore an increase in the termination time of the algorithm. In fact, if θ is chosen too low, one can obtain (just as we saw in (39)) an infinite expected termination time. One can opt, however, to increase θ, as long as $\theta < 2\mu$ in order to increase the frequency of the big jumps and therefore make the termination time of the algorithm shorter. Considering for instance an extreme case, we saw that if $\alpha \in (1, 2)$ then the optimal change-of-measure (in the sense of variance minimization) has infinite expected termination time (that is, whenever $\alpha \in (1, 2), E(\tau_n | \tau_n < \infty) = \infty$). By choosing $\theta \approx 2\mu$, one can show that the termination time of the importance sampling estimator constructed by means of the previous Lyapunov procedure, which maintains strong efficiency, is finite as long as $\alpha \in (3/2, 2)$. These optimality results are discussed in [111].

5.5. Notes on importance sampling with heavy tails

Detailed expositions on large deviations asymptotics for heavy-tailed random variables can be found in [96,97] and Chapter IX of [40]. General discussion on importance samplers with heavy tails is covered in Chapter VI of [23,50].

As mentioned earlier this section, the various challenges of rare-event simulation with heavy tails are pointed out in [4,110]. The former focuses on the singularity of conditional measure with respect to the original measure and the difficulty of finding suitable importance sampler, while suggesting provably efficient conditional Monte Carlo scheme. The latter suggests the non-existence of state-independent importance samplers for some first passage problems. Conditional Monte Carlo is also studied in [98] in the context of insurance, and further improved in [99]. It is also used in [114] for sums of dependent random variables. The hazard rate twisting method that we discussed in Section 5.1 appears in [115,101,116], as well as [117] in the context of finance and [118] in single-server queues. Another version of this method is studied in [119] using cross-entropy, while [120] suggests a closely related transform method that essentially equates hazard rate twisting. Finally, [121] considers a different type of twisting, by scaling the variance of the underlying variables, for sums of correlated lognormals.

Regarding state-dependent importance samplers, as we have discussed, [113] suggests the weak convergence approach for regularly varying sums and the use of a two-mixture sampler for each increment distribution. Two mixtures are no longer enough for other heavy tails, such as Weibull. The appropriate family of mixtures in such settings is developed in [111]. The Lyapunov approach is proposed and applied in a variety of contexts by [11] in single-server queues, [10,122] in regularly varying random walks, [9] in random walk maximum and [123] in random sum. The last two papers in particular aim to construct samplers under the assumptions of general subexponential distributions. The paper [124] further extends the method to the case of Markov modulation. Mixture-type algorithm is also used in [125] for computing quantile and other risk measures that involve regularly varying tails. Recently, [126] proposes sequential importance sampling technique with resampling that appears to exhibit good empirical performance in terms of variance reduction. Moreover, [127] considers an adaptive mechanism based on cross-entropy methods to guide the selection of provably efficient importance sampling estimators for heavy-tailed sums.

Acknowledgments

We are grateful to the referees for their useful comments which helped to improve the exposition in this paper. In addition, support from the NSF through the grants DMS-0806145, DMS-0846816 and DMS-1069064 is gratefully acknowledged.

References

[1] P. Glasserman, S. Kou, Analysis of an importance sampling estimator for tandem queues, ACM TOMACS 5 (1995) 22–42.
[2] P. Glasserman, Y. Wang, Counter examples in importance sampling for large deviations probabilities, Annals of Applied Probability 7 (1997) 731–746.
[3] A. Bassamboo, S. Juneja, A. Zeevi, On the inefficiency of state-independent importance sampling in the presence of heavy-tails, Operations Research Letters 34 (2006) 521–531.
[4] S. Asmussen, K. Binswanger, B. Hojgaard, Rare events simulation for heavy-tailed distributions, Bernoulli 6 (2000) 303–322.
[5] P. Dupuis, H. Wang, Importance sampling, large deviations, and differential games, Stochastic and Stochastic Reports 76 (2004) 481–508.
[6] P. Dupuis, H. Wang, Subsolutions of an Isaacs equation and efficient schemes of importance sampling, Mathematics of Operations Research 32 (2007) 723–757.
[7] P. Dupuis, A. Sezer, H. Wang, Dynamic importance sampling for queueing networks, The Annals of Applied Probability 17 (2007) 1306–1346.
[8] P. Dupuis, K. Leder, H. Wang, Large deviations and importance sampling for a tandem network with slowdown, QUESTA 57 (2007) 71–83.
[9] J. Blanchet, P. Glynn, Efficient rare-event simulation for the maximum of a heavy-tailed random walk, The Annals of Applied Probability 18 (2008) 1351–1378.
[10] J. Blanchet, J.C. Liu, State-dependent importance sampling for regularly varying random walks, Advances in Applied Probability 40 (2008) 1104–1128.
[11] J. Blanchet, P. Glynn, J.C. Liu, Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed $G/G/1$ queue, QUESTA 57 (2007) 99–113.
[12] J. Blanchet, Efficient importance sampling for binary contingency tables, The Annals of Applied Probability 19 (2009) 949–982.
[13] S. Juneja, Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error, QUESTA 57 (2007) 115–127.
[14] P. L'Ecuyer, M. Mandjes, B. Tuffin, Importance sampling in rare event simulation, in: G. Rubino, B. Tuffin (Eds.), Rare Event Simulation Using Monte Carlo Methods, Wiley, West Sussex, United Kingdom, 2009, pp. 17–38. (chapter 2).

[15] P. L'Ecuyer, J. Blanchet, B. Tuffin, P. Glynn, Asymptotic robustness of estimators in rare-event simulation, ACM TOMACS 20 (2010) 6:1–6:41.
[16] B. Tuffin, Bounded normal approximation for the simulation of highly reliable Markovian systems, Journal of Applied Probability 36 (1999) 974–986.
[17] B. Tuffin, On numerical problems in simulation of highly reliable Markovian systems, in: Proceedings of the 1st International Conference on Quantitative Evaluation of SysTems (QEST), IEEE Computer Society Press, Los Alamitos, CA, 2004, pp. 156–164.
[18] L.W. Schruben, A coverage function for interval estimators of simulation response, Management Science 26 (1980) 18–27.
[19] H. Cancela, G. Rubino, B. Tuffin, New measures of robustness in rare event simulation, in: Proceedings of the 2005 Winter Simulation Conference, IEEE Press, 2005, pp. 519–527.
[20] P. Glynn, W. Whitt, The asymptotic efficiency of simulation estimators, Operations Research 40 (1992) 505–520.
[21] P. Glynn, G. Rubino, B. Tuffin, Robustness properties and confidence interval reliability issues, in: Rare Event Simulation Using Monte Carlo Methods, West Sussex, United Kingdom, 2009, pp. 63–84. (chapter 4).
[22] R.Y. Rubinstein, D.P. Kroese, Simulation and the Monte Carlo Method, second ed. John Wiley, 2009.
[23] S. Asmussen, P. Glynn, Stochastic Simulation: Algorithms and Analysis, Springer-Verlag, New York, NY, USA, 2007.
[24] J. Liu, Monte Carlo Strategies in Scientific Computing, Springer, 2001.
[25] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, second ed. Springer, New York, 1998.
[26] A. Ganesh, N. O'Connell, D. Wischik, Big Queues, Springer-Verlag, Berlin Heidelberg, Germany, 2004.
[27] T.M. Cover, J.A. Thomas, Elements of Information Theory, second ed. Wiley, 2006.
[28] S. Asmussen, R. Rubinstein, Steady state rare events simulation in queueing models and its complexity properties, in: J. Dshalalow (Ed.), Advances in Queueing, CRC Press, 1995, pp. 429–461.
[29] D. Siegmund, Importance sampling in the Monte Carlo study of sequential tests, The Annals of Statistics 3 (1976) 673–684.
[30] S. Asmussen, Applied Probability and Queues, Springer-Verlag, New York, NY, USA.
[31] R.T. Rockafellar, Convex Analysis, Princeton University Press, 1970.
[32] P. Glasserman, Monte Carlo Methods in Financial Engineering, Springer-Verlag, New York, 2004.
[33] A. Shwartz, A. Weiss, Large Deviations for Performance Analysis. Queues, Communications, and Computing, Chapman & Hall, London, UK, 1995.
[34] R.S. Ellis, Entropy, Large Deviations and Statistical Mechanics, Springer-Verlag, Berlin Heidelberg, Germany, 1985.
[35] P. Dupuis, R.S. Ellis, A Weak Convergence Approach to the Theory of Large Deviations, Wiley, New York, 1997.
[36] J.S. Sadowsky, Large deviations theory and efficient simulation of excessive backlogs in a $GI/GI/m$ queue, IEEE Transactions on Automatic Control 36 (1991) 1383–1394.
[37] P. Glynn, D. Iglehart, Importance sampling for stochastic simulations, IEEE Transactions on Automatic Control 35 (1989) 1367–1392.
[38] S. Asmussen, Conjugate processes and the simulation of ruin problems, Stochastic Processes and their Applications 20 (1985) 213–229.
[39] T. Lehtonen, H. Nyrhinen, Simulating level-crossing probabilities by importance sampling, Advances in Applied Probability 24 (1992) 858–874.
[40] S. Asmussen, Ruin Probabilities, World Scientific, London, UK, 2000.
[41] J. Collamore, Importance sampling techniques for the multidimensional ruin problem for general markov additive sequences of random vectors, The Annals of Applied Probability 12 (2002) 382–421.
[42] D.P. Kroese, V.F. Nicola, Efficient estimation of overflow probabilities in queues with breakdowns, Performance Evaluation 36–37 (1999) 471–484.
[43] S. Parekh, J. Walrand, Quick simulation of rare events in networks, IEEE Transactions on Automatic Control 34 (1989) 54–66.
[44] C.S. Chang, P. Heidelberger, S. Juneja, P. Shahabuddin, Effective bandwidth and fast simulation of ATM intree networks, Performance Evaluation 20 (1994) 45–65.
[45] P.T. de Boer, Analysis of state-independent importance-sampling measures for the two-node tandem queue, ACM Transactions on Modeling and Computer Simulation 16 (2006) 225–250.
[46] J.S. Sadowsky, On Monte Carlo estimation of large deviations probabilities, The Annals of Applied Probability 6 (1996) 399–422.
[47] J. Bucklew, Introduction to Rare Event Simulation, Springer-Verlag, New York, NY, 2004.
[48] A.B. Dieker, M. Mandjes, On asymptotically efficient simulation of large deviation probabilities, Advances in Applied Probability 37 (2005) 539–552.
[49] J. Bucklew, Large Deviation Techniques in Decision, Simulation and Estimation, Wiley, New York, NY, USA, 1990.
[50] S. Juneja, P. Shahabuddin, Rare event simulation techniques: an introduction and recent advances, in: S.G. Henderson, B.L. Nelson (Eds.), Simulation, Handbooks in Operations Research and Management Science, Elsevier, Amsterdam, The Netherlands, 2006.
[51] P. Heidelberger, Fast simulation of rare events in queueing and reliability models, ACM Transactions on Modeling and Computer Simulation (TOMACS) 5 (1995) 43–85.
[52] J. Blanchet, M. Mandjes, Rare event simulation for queues. in: Rare Event Simulation Using Monte Carlo Methods, West Sussex, United Kingdom, 2009, pp. 87–124. (chapter 5).
[53] R. Szechtman, P. Glynn, Rare event simulation for infinite server queues. in: E. Yucesan, C. H. Chen, J. L. Snowdon, and J. M. Charnes, Eds., Proceedings of the 2002 Winter Simulation Conference, pp. 416–423, 2002.
[54] A. Ridder, Importance sampling algorithms for first passage time probabilities in the infinite server queue, European Journal of Operational Research 199 (2009) 176–186.
[55] J. Blanchet, P. Glynn, H. Lam, Rare event simulation for a slotted time $M/G/s$ model, Queueing System 63 (2009) 33–57.
[56] G. Rubino, B. Tuffin, Markovian models for dependability analysis, in: G. Rubino, B. Tuffin (Eds.), Rare Event Simulation Using Monte Carlo Methods, Wiley, West Sussex, United Kingdom, 2009, pp. 125–144. (chapter 6).
[57] P. Grassberger, Go with the winners: a general Monte Carlo strategy, Computer Physics Communications 147 (2002) 64–70.
[58] W. Sandmann, Rare event simulation methodologies in systems biology, in: Rare Event Simulation Using Monte Carlo Methods, Wiley, West Sussex, United Kingdom, 2009, pp. 243–266. (chapter 11).
[59] R.Y. Rubinstein, D.P. Kroese, The Cross-Entropy Method, Springer, New York, NY, 2004.
[60] M. Bayati, J. Kim, A. Saberi, A sequential algorithm for generating random graphs, in: Approximation, Randomization and combinatorial Optimization. Algorithms and Techniques, Lecture Notes in Computer Science 4627 (2007) 326–340.
[61] P. L'Ecuyer, F. Le Gland, P. Lezaud, B. Tuffin, Splitting techniques, in: G. Rubino, B. Tuffin (Eds.), Rare Event Simulation Using Monte Carlo Methods, Wiley, West Sussex, United Kingdom, 2009, pp. 39–62. (chapter 3).
[62] P. Glasserman, P. Heidelberger, P. Shahabuddin, T. Zajic, A large deviations perspective on the effiency of multilevel splitting, IEEE Transactions on Automated Control (1998) 1666–1679.
[63] P. Glasserman, P. Heidelberger, P. Shahabuddin, T. Zajic, Multilevel splitting for estimating rare event probabilities, Operations Research 47 (1999) 585–600.
[64] T. Dean, P. Dupuis, Splitting for rare event simulation: a large deviation approach to design and analysis, Stochastic Processes and their Applications 119 (2009) 562–587.
[65] R.Y. Rubinstein, Randomized algorithms with splitting: why the classic randomized algorithms do not work and how to make them work, Methodology and Computing in Applied Probability 12 (2010) 1–50.
[66] P. del Moral, Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications, Springer-Verlag, New York, USA, 2004.
[67] C. Kollman, K. Baggerly, D. Cox, R. Picard, Adaptive importance sampling on discrete Markov chains, Annals of Applied Probability 9 (1999) 391–412.
[68] P.Y. Desai, P.W. Glynn, A Markov chain perspective on adaptive Monte Carlo algorithms, Proceedings of 2001 Winter Simulation Conference 9 (2001) 391–412.
[69] P.T. De Boer, V.F. Nicola, R.Y. Rubinstein, Adaptive importance sampling simulation of queueing networks, in: K. Kang, J.A. Joines, R.R. Barton, P.A. Fishwick (Eds.), Proceedings of 2000 Winter Simulation Conference, IEEE Press, Piscataway, New Jersey, 2000, pp. 646–655.
[70] P.T. de Boer, D. Kroese, S. Mannor, R. Rubinstein, A tutorial on the cross-entropy method, Annals of Operations Research 134 (2005) 19–67.
[71] R.Y. Rubinstein, Optimization of computer simulation models with rare events, European Journal of Operations Research 99 (1997) 89–112.
[72] R.Y. Rubinstein, Rare-event simulation via cross-entropy and importance sampling. Second Workshop on Rare Event Simulation, RESIM 99, pp. 1–17, 1999.
[73] T.P.I. Ahamed, V.S. Borkar, S. Juneja, Adaptive importance sampling technique for Markov chains using stochastic approximation, Operations Research 54 (2006) 489–504.
[74] V.S. Borkar, S. Juneja, A.A. Kherani, Performance analysis conditioned on rare events: an adaptive simulation scheme, Communications in Information 3 (2004) 259–278.
[75] H.P. Chan, T.L. Lai, A sequential Monte Carlo approach to computing tail probabilities in stochastic models, The Annals of Applied Probability, 2011 (forthcoming).
[76] E. Weinan, R. Weiqing, E. Vanden-Eijnden, Minimum action method for the study of rare events, Communications on Pure & Applied Mathematics 52 (2004) 637–656.
[77] H. Hult, F. Lindskog, Heavy-tailed insurance portfolios: buffer capital and ruin probabilities, School of ORIE, Cornell University, Technical Report No. 1441, 2006.
[78] J. Blanchet, P. Glynn, K. Leder, On Lyapunov inequalities and subsolutions for efficient importance sampling, 2010. Preprint, http://www.columbia.edu/jb2814/papers/LyapSub7_Dec20_2009.pdf.
[79] P. Dupuis, H. Wang, Dynamic importance sampling for uniformly recurrent Markov chains, The Annals of Applied Probability 15 (2005) 1–38.
[80] P. Dupuis, H. Wang, Importance sampling for Jackson networks, Queueing Systems 62 (2009) 113–157.
[81] P. Dupuis, H. Ishii, H.M. Soner, A viscosity solution approach to the asymptotic analysis of queueing systems, The Annals of Probability 18 (1990) 226–255.
[82] J. Blanchet, H. Lam, Rare-event simulation for many-server queues, Preprint.
[83] J.S. Sadowsky, J.A. Bucklew, On large deviations theory and asymptotically efficient Monte Carlo estimation, IEEE Transactions on Information Theory 36 (1990) 579–588.

[84] H.P. Chan, T.L. Lai, Efficient importance sampling for Monte Carlo evaluation of exceedance probabilities, The Annals of Applied Probability 17 (2007) 440–473.
[85] Y.V. Linnik, On the probability of large deviations for the sums of independent variables, in: Proceedings of 4th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 1961, pp. 283–306.
[86] A.V. Nagaev, Integral limit theorems taking large deviations into account when Cramer's condition does not hold II, Theory of Probability and Its Applications 14 (1969) 193–208.
[87] S. Nagaev, Large deviations of sums of independent random variables, Annals of Probability 7 (1979) 745–789.
[88] L.V. Rozovskii, Probabilities of large deviations of sums of independent random variables, Theory of Probability and Its Applications 34 (1990) 625–644.
[89] L.V. Rozovskii, Probabilities of large deviations on the whole axis, Theory of Probability and its Applications 38 (1993) 53–79.
[90] A.A. Borovkov, K.A. Borovkov, On probabilities of large deviations for random walks. I. Regularly varying distribution tails, Theory of Probability and its Applications 49 (2001) 189–205.
[91] S. Asmussen, Subexponential asymptotics for stochastic processes: extremal behavior, stationary distributions and first passage probabilities, The Annals of Applied Probability 8 (1998) 354–374.
[92] S. Foss, Z. Palmowski, S. Zachary, The probability of exceeding a high boundary on a random time interval for a heavy-tailed random walk, Annals of Applied Probability 15 (2005) 1936–1957.
[93] S. Foss, T. Konstantopoulos, S. Zachary, Discrete and continuous time modulated random walks with heavy-tailed increments, Journal of Theoretical Probability 20 (2007) 581–612.
[94] S. Foss, S. Zachary, The maximum on a random time interval of a random walk with long-tailed increments and negative drift, Annals of Applied Probability 13 (2003) 37–53.
[95] S. Asmussen, C. Kluppelberg, Large deviations results for subexponential tails, with applications to insurance risk, Stochastic Processes and their Applications 64 (1996) 103–125.
[96] P. Embrechts, C. Klpelberg, T. Mikosch, Modelling Extremal Events for Insurance and Finance, Springer-Verlag, New York, 1997.
[97] S.I. Resnick, Extreme Values, Regular Variation, and Point Processes, New York, 1987.
[98] S. Asmussen, K. Binswanger, Simulation of ruin probabilities for subexponential claims, ASTIN Bulletin 27 (1997) 297–318.
[99] S. Asmussen, D. Kroese, Improved algorithms for rare event simulation with heavy tails, Advances in Applied Probability 38 (2006) 545–558.
[100] L. Rojas-Nandayapa, S. Asmussen, Efficient simulation of finite horizon problems in queueing and insurance risk, QUESTA 57 (2007) 85–97.
[101] S. Juneja, P. Shahabuddin, Simulating heavy-tailed processes using delayed hazard rate twisting, ACM TOMACS 12 (2002) 94–118.
[102] H. Hult, F. Lindskog, T. Mikosch, G. Samordnitsky, Functional large deviations for multivariate regularly varying random walks, The Annals of Applied Probability 15 (2005) 2651–2680.
[103] A.P. Zwart, Queueing Systems with Heavy Tails, Ph.D. Dissertation, The Netherlands, 2001.
[104] B. Zwart, S. Borst, M. Mandjes, Exact asymptotics for fluid queues fed by multiple heavy-tailed on-off flows, Annals of Applied Probability 14 (2004) 903–957.
[105] S. Foss, D. Korshunov, Heavy tails in multi-server queue, QUESTA 52 (2006) 31–48.
[106] A.G. Pakes, On the tails of waiting-time distribution, Journal of Applied Probability 12 (1975) 555–564.
[107] N. Veraverbeke, Asymptotic behavior of Wiener–Hopf factors of a random walk, Stochastic Processes And Their Applications 5 (1977) 27–31.
[108] P. Embrechts, N. Veraverbeke, Estimates for the probability of ruin with special emphasis on the possibility of large claims, Insurance: Mathematical Economics 1 (1982) 55–72.
[109] J.W. Cohen, On the $M/G/1$ queue with heavy-tailed service time distributions, Technical Report PNA-R9702, CWI, Amsterdam, 1997.
[110] A. Bassamboo, S. Juneja, A. Zeevi, Importance sampling simulation in the presence of heavy tails, in: M.E. Kuhl, N.M. Steiger, F.B. Armstrong, J.A. Joines (Eds.), Proceedings of 2005 Winter Simulation Conference, IEEE Press, Piscataway, New Jersey, 2005, pp. 664–672.
[111] J. Blanchet, J.C. Liu, Efficient simulation and conditional functional limit theorems for ruinous heavy-tailed random walks, 2010. http://www.columbia.edu/jb2814/papers/ConcaveHazard9f.pdf (submitted for publication).
[112] C. Kluppelberg, A. Kyprianou, R. Maller, Ruin probabilities and over-shoots for general levy insurance risk processes, The Annals of Applied Probability 14 (2004) 1766–1801.
[113] P. Dupuis, K. Leder, H. Wang, Importance sampling for sums of random variables with regularly varying tails, ACM TOMACS 17 (2006).
[114] J. Blanchet, L. Rojas-Nandayapa, Efficient simulation of tail probabilities of sums of dependent random variables, Journal of Applied Probability, Special Vol. 48A (2011) 147–165.
[115] S. Juneja, P. Shahabuddin, A. Chandra, Simulating heavy-tailed processes using delayed hazard-rate twisting, in: P.A. Farrington, H.B. Nembhard, D.T. Sturrock, G.W. Evans (Eds.), Proceedings of 1999 Winter Simulation Conference, IEEE Press, Piscataway, New Jersey, 1999, pp. 420–427.
[116] Z. Huang, P. Shahabuddin, A unified approach for finite dimensional, rare-event Monte Carlo simulation, in: R.G. Ignalls, M.D. Rossetti, J.S. Smith, B.A. Peters (Eds.), Proceedings of 2004 Winter Simulation Conference, IEEE Press, Piscataway, New Jersey, 2004, pp. 1616–1624.
[117] Z. Huang, P. Shahabuddin, Rare-event, heavy-tailed simulations using hazard function transformations, with applications to value-at-risk, in: S. Chick, P.J. Sanchez, D. Ferrin, D. Morrice (Eds.), Proceedings of 2003 Winter Simulation Conference, IEEE Press, Piscataway, New Jersey, 2003, pp. 276–284.
[118] N.K. Boots, P. Shahabuddin, Simulating $GI/GI/1$ queues and insurance processes with subexponential distributions, in: J.A. Joines, R.R. Barton, K. Kang, P.A. Fishwick (Eds.), Proceedings of 2000 Winter Simulation Conference, IEEE Press, Piscataway, New Jersey, 2000, pp. 656–665.
[119] S. Asmussen, D.P. Kroese, R.Y. Rubinstein, Heavy tails, importance sampling and cross entropy, Stochastic Models 21 (2005) 57–76.
[120] D.P. Kroese, R.Y. Rubinstein, The transform likelihood ratio method for rare-event simulation with heavy tails, Queueing Systems 46 (2004) 317–351.
[121] S. Asmussen, J. Blanchet, S. Juneja, L.R. Nandayapa, Efficient simulation of tail probabilities of sums of correlated lognormals, Annals of Operations Research 189 (2009) 5–23.
[122] J. Blanchet, J.C. Liu, Importance sampling in ruin problems for multidimensional regularly varying random walks, Journal of Applied Probability 47 (2010) 301–322.
[123] J. Blanchet, C. Li, Efficient rare-event simulation for heavy-tailed compound sums, ACM TOMACS Transactions in Modeling and Computer Simulation 21 (2011) 1–10.
[124] J. Blanchet, H. Lam, Efficient state-dependent importance sampling for Markov-modulated heavy-tailed random walk, Working Paper.
[125] H. Hult, J. Svensson, Efficient calculation of risk measures by importance sampling—the heavy tailed case. arXiv:0909.3335v1 [math.PR], 2009.
[126] H.P. Chan, S. Deng, T.L. Lai, Rare-event simulation of heavy-tailed random walks by sequential importance sampling and resampling, Preprint.
[127] J. Blanchet, Y. Shi, Efficient rare event simulation for heavy-tailed systems via cross entropy. S. Jain, R.R. Creasey, J. Himmelspach, K. P. White, and M. Fu, (Eds.), in: Proceedings of 2011 Winter Simulation Conference. IEEE Press (in press).