

STATE-DEPENDENT IMPORTANCE SAMPLING FOR REGULARLY VARYING RANDOM WALKS

JOSE H. BLANCHET* AND

JINGCHEN LIU,** *Columbia University*

Abstract

Consider a sequence $(X_k : k \geq 0)$ of regularly varying independent and identically distributed random variables with mean 0 and finite variance. We develop efficient rare-event simulation methodology associated with large deviation probabilities for the random walk $(S_n : n \geq 0)$. Our techniques are illustrated by examples, including large deviations for the empirical mean and path-dependent events. In particular, we describe two efficient state-dependent importance sampling algorithms for estimating the tail of S_n in a large deviation regime as $n \nearrow \infty$. The first algorithm takes advantage of large deviation approximations that are used to mimic the zero-variance change of measure. The second algorithm uses a parametric family of changes of measure based on mixtures. Lyapunov-type inequalities are used to appropriately select the mixture parameters in order to guarantee bounded relative error (or efficiency) of the estimator. The second example involves a path-dependent event related to a so-called knock-in financial option under heavy-tailed log returns. Again, the importance sampling algorithm is based on a parametric family of mixtures which is selected using Lyapunov bounds. In addition to the theoretical analysis of the algorithms, numerical experiments are provided in order to test their empirical performance.

Keywords: State dependent; importance sampling; regularly varying; random walk; heavy tail; strong efficiency; rare-event simulation

2000 Mathematics Subject Classification: Primary 60G50; 60J05; 68W40

Secondary 60G70

1. Introduction

Our goal is to develop efficient rare-event simulation estimators for large deviation probabilities of regularly varying random walks (an estimator for such large deviation probabilities is said to be strongly efficient if its coefficient of variation remains bounded as the target probability approaches 0).

The development of analytical or numerical methods for random walks (RWs) often suggest useful techniques that are applicable to more general models. A classical example that illustrates how the ideas underlying the structure of RWs can provide insight into more general settings is given by Cramer's theorem and its connection to large deviations theory for systems with light-tailed input. Loosely speaking, these are systems in which there is a law of large numbers or 'averaging principle' operating directly to random variables that have a finite moment generating function (see, for instance, Dembo and Zeitouni (1998, Chapter 5)). Cramer's theorem is

Received 3 January 2008; revision received 28 August 2008.

* Postal address: Department of Industrial Engineering and Operations Research, Columbia University, S. W. Mudd Building, 500 West 120th Street, New York, NY 10027-6699. Email address: blanchet@fas.harvard.edu

** Postal address: Department of Statistics, Columbia University, 1255 Amsterdam Avenue, Room 1030, New York, NY 10027.

a celebrated result that captures the exact exponential rate of decay for the probability that the empirical mean of n light-tailed independent and identically distributed (i.i.d.) random variables (RVs) deviates from the (population) mean by amounts of order $O(1)$ as $n \nearrow \infty$. The principles underlying Cramer's theorem have found far reaching generalizations, pioneered by the work of Donsker and Varadhan (1975), not only to random walks (in the form, for instance, of Sanov's theorem—see Dembo and Zeitouni (1998, Section 3.1)) but also to much more general systems and processes. In addition, it is of great value that some of the basic proof techniques underlying the ideas behind Cramer's theorem in turn allow us to develop a rich set of computational tools that are useful to estimate large deviation probabilities efficiently via importance sampling. Indeed, it is well known that from the proof of asymptotic lower bounds in traditional large deviations results we can obtain a change of measure that often (but unfortunately not always—see Glasserman and Wang (1997) and Glasserman and Kou (1995)) can be proved to be efficient; Dupuis and Wang (2004), (2005) have developed techniques to guarantee the efficiency of exponential changes of measure. For excellent expositions of rare-event simulation methodology, see Bucklew (2004, Chapter 4), Juneja and Shahabuddin (2006), and Asmussen and Glynn (2007, Chapter VI).

RWs are used to build a large variety of stochastic models such as queueing models and insurance/reservoir processes; see Asmussen (2003, Chapter 3). Some of these models are used to study phenomena that exhibit certain stylized features that are well captured by heavy-tailed characteristics in the input data; see Adler *et al.* (1998). This motivates the large deviation analysis of RWs with heavy-tailed features. Recently, Hult and Lindskog (2005) and Hult *et al.* (2005) have developed sample path large deviations for regularly varying processes and other results that parallel some of the development reached by the light-tailed counterpart. However, contrary to the light-tailed situation, the proof techniques do not lead themselves directly to the design of efficient importance sampling algorithms for rare-event simulation. Other large deviations results have recently been obtained for heavy-tailed processes with Markovian structure; see, for instance, Foss and Korshunov (2006) and Foss *et al.* (2007). Our techniques here can also be applied to these types of processes, this research will be reported elsewhere.

The connection between large deviations theory for heavy-tailed systems and importance sampling has received substantial attention in recent years. The work by Asmussen *et al.* (2000) provides a number of examples that indicates some of the problems that arise when trying to apply importance sampling ideas based on large deviations theory for systems with heavy-tailed characteristics; see also Bassamboo *et al.* (2005) for additional issues arising in rare-event simulation for systems with heavy tails. Perhaps the simplest way to see the types of issues that arise in rare-event simulation of systems with heavy-tailed features stems from the fact that often, in the heavy-tailed setting, the large deviations behavior of the system is caused by one or a few components with 'extreme' behavior, while the rest of the system is operating in 'normal' circumstances. Implementing a change of measure with such characteristics is not entirely obvious. The reason is that the contribution to the variances from the sample paths that lie in the large deviations domain for which none or more than just a few components exhibit 'extreme' behavior is nonnegligible. It is then important to have a good understanding of the contribution of the sample paths that will add to the variance significantly, but are negligible in terms of the most likely asymptotic behavior of the system.

Recently, state-dependent importance sampling (SDIS) has proved to be useful in developing efficient simulation estimators. For instance, Blanchet and Glynn (2007) took advantage of well-known asymptotics in the design of an SDIS algorithm for the tail of the maximum of

an RW with heavy-tailed increments, including regularly varying, Weibull, lognormal, and many other types of heavy-tailed distributions. Blanchet and Glynn (2007) also introduced the use of Lyapunov inequalities to test the efficiency of SDIS schemes. Blanchet *et al.* (2007) developed an efficient SDIS algorithm for the tail of the steady-state delay in the G/G/1 queue with regularly varying input. Blanchet *et al.*'s (2007) algorithm is applied to the regenerative ratio representation for steady-state measures whose main ideas can be applied to more general queueing systems such as multiserver queues; see Blanchet *et al.* (2008). Blanchet and Liu (2006) also considered rare-event simulation for the empirical mean of i.i.d. regularly varying RVs. Blanchet and Liu (2006) developed an algorithm that is close in spirit to that of Blanchet and Glynn (2007), in the sense that their importance sampling scheme takes advantage of well-known asymptotics (although in contrast to Blanchet and Glynn (2007), the change of measure proposed by Blanchet and Liu (2006) is time inhomogeneous). In this work we discuss important implementation issues that were left open in Blanchet and Liu's (2006) development, such as the additional complexity involved in generating paths under Blanchet and Liu's (2006) proposed change of measure. In addition, we propose a different SDIS algorithm that copes with the problem of variate generation under their change of measure. It is important to note that the new algorithm requires only an asymptotic lower bound for the probability to claim the decaying rate of the theoretical limit. Both the asymptotic upper bound for such probability and the proof of efficiency of the importance sampling algorithm are obtained via Lyapunov inequalities. Thereby, the large deviations analysis and importance sampling are connected in the same spirit as in the light-tailed situations discussed previously.

In order to describe more precisely the types of results that we develop, let $S_0 = 0$ and define $S_n = X_1 + \cdots + X_n$, where the X_i s have regularly varying right tails with index $\alpha > 2$ (i.e. $P(X > xt)/P(X > t) \rightarrow x^{-\alpha}$ as $t \nearrow \infty$ for each $x \in (0, \infty)$). Assume that the X_i s have mean 0 and finite variance. We will develop an efficient importance sampling methodology for estimating large deviation probabilities, such as $P(S_n > b)$ when n and $b \geq cn^{1/2+\varepsilon}$ (for any $\varepsilon > 0$) and $P(\min_{k \leq n} S_k \leq -a, S_n > b)$, assuming, for instance, that $a = n\gamma > 0$ and $b = cn > 0$ for some γ and $c > 0$. The last probability can be interpreted as the chance of exercising a financial call option that is activated only if the underlying price hits a minimum threshold before maturity—such a financial contract is a knock-in option, more precisely, it is a down-and-in call option.

As we indicated before, in order to prove efficiency of our estimator, we do not require to have exact large deviations (although such results obviously exist for the empirical mean and are due to Nagaev (1969a), (1969b); see also Rozovskii (1989)). It suffices to have an asymptotic lower bound for the probability of interest. In order to develop an upper bound for the second moment of our estimator (and, by Jensen's inequality, an upper bound for the probability of interest), we take advantage of a Lyapunov inequality. We believe that the technique can potentially be sharpened to obtain exact asymptotics (rather than just so-called 'big O' asymptotics), but we do not pursue this line of research here; we content ourselves with proving efficiency of the scheme and with providing a method that allows us to connect large deviations estimates with rare-event simulation via importance sampling, which is one of the main points that we aim to convey.

Some remarkable features of our results here concern the spatial scales and the efficiency criteria that are applicable to algorithms relative to the standard results that are well known for the light-tailed counterpart. In the light-tailed case, the standard exponential change of measure obtained directly from the proof of the lower bound in Cramer's theorem yields an algorithm that can be used to estimate $P(S_n > b)$, say if $c_0 n^{1/2+\varepsilon} \leq b \leq c_1 n$ (for any $c_0, c_1 > 0$ and

$\varepsilon \in (0, \frac{1}{2})$; see Sadowsky and Bucklew (1990) or the text of Bucklew (2004, Chapter 3)). The algorithm can only be proved to be weakly efficient (which means that the coefficient of variation is not bounded, but grows slowly as $n \nearrow \infty$). In contrast, our algorithm for heavy-tailed RWs can be guaranteed to apply throughout the large deviations region, $b \geq \gamma n$ for $\gamma > 0$. The algorithms that we propose here also cover moderate deviations, since we only assume that $b \geq cn^{1/2+\varepsilon}$ for some $c > 0$ and any $\varepsilon > 0$. In addition, our algorithms can be shown to be strongly efficient in the sense that the coefficient of variation remains bounded as $n \nearrow \infty$.

Rare-event simulation for sums of heavy-tailed increments has been the topic of several papers in the literature. However, most of the previous work in this line has considered the case in which n is fixed (n is often allowed to be an RV with exponentially decaying tails) and $b \nearrow \infty$. This situation is motivated by the problem of efficient estimation of the tail distribution of the steady-state waiting time of the M/G/1 queue, which can be expressed as a sum of a random number of positive i.i.d. RVs with a geometric distribution by virtue of the Pollaczek–Khinchine representation. Several efficient algorithms have been developed for the situation in which n is fixed. The first algorithm which applies to the case of regularly varying increments and is based on conditional Monte Carlo was proposed by Asmussen *et al.* (2000), Juneja and Shahabuddin (2002) proposed an importance sampling algorithm based on hazard rate twisting. Asmussen and Kroese (2006) proposed improved estimators that are strongly efficient and have excellent practical performance. Dupuis *et al.* (2006) proposed algorithms that are based on SDIS. Their approach, motivated from a control-theoretic perspective, introduced the use of mixture samplers such as those discussed here. Their development is not designed to cover the case in which n is large and the techniques used to prove efficiency are completely different to the ones that we propose here.

Our general strategy involves three steps. First, we propose a parametric family of importance sampling distributions for the increments of the RW. The parameters of the family are allowed to depend on the current position of the RW. Such a parametric family captures the asymptotic behavior of the RW conditional on the occurrence of a large deviations event. Since we are in a heavy-tailed setting, such a large deviations event is expected to be caused by a single large jump. As a consequence, the parametric family of distributions takes the form of a mixture: with some probability, say p , determined using a Lyapunov inequality, as we will explain, we sample the next increment (given the current state of the RW) conditional on it being large, and with probability $(1 - p)$ we sample basically from the nominal (original) increment distribution. Then, we need to tune the parameters of the mixture distributions. This tuning is done in the second and the third steps of the procedure. These two steps involve solving a Lyapunov inequality that is required to upper bound the second moment of the estimator. Such an inequality is basically a system of linear equations and its solution is called a Lyapunov function. The second step involves guessing an appropriate parametric family of functions from which a Lyapunov function will be selected. Since we are interested in efficiency, which basically implies that the second moment of the estimator should behave at least asymptotically as the square of its first moment, good guesswork for a convenient Lyapunov function involves understanding the behavior of $P(S_n > b \mid S_0 = y)$ (in terms of n , b , and y). The last step of our strategy involves a rigorous verification of the guesswork, that is, testing the Lyapunov inequality. It is in this last procedure that we force both the mixture parameters and the parameters in our Lyapunov function to comply with the inequality in Proposition 1.

The rest of the paper is organized as follows. In Section 2 we summarize some results on large deviations for heavy-tailed RVs and discuss their applications in the context of Blanchet

and Liu’s (2006) algorithm for the empirical mean. In Section 3 we introduce some Lyapunov inequalities for SDIS and discuss some elements behind our strategy, namely, those corresponding to steps one and two outlined previously. The verification of efficiency of our proposed estimator for the empirical mean, which corresponds to the third step of our strategy, is given in Section 4. In Section 5 we apply our strategy to the probability of exercising a call-in financial option. Numerical experiments and implementation issues are given in Section 6.

2. An SDIS approach

Throughout the rest of the paper, we will assume that $(X_k: k \geq 1)$ is a sequence of i.i.d. RVs with mean 0 and unit variance. In addition, we assume that the X_n s possess a regularly varying right tail with index $\alpha > 2$. That is, $\bar{F}(t) = P(X_k > t) = t^{-\alpha}L(t)$, where $L(\cdot)$ is a slowly varying function at ∞ (i.e. $L(tx)/L(t) \rightarrow 1$ as $t \nearrow \infty$ for each $x > 0$).

In this section we will describe a simulatable estimator $Z(b)$ that can be shown to be (strongly) efficient throughout moderate and large deviations regions. More precisely, $Z(b)$ will satisfy $E(Z(b)) = P(S_n > b)$, with $S_0 = 0$ and $S_n = X_1 + \dots + X_n$, and

$$\sup_{\substack{n \geq 1 \\ b \geq \xi(n)}} \frac{E(Z^2(b))}{P(S_n > b)^2} < \infty, \tag{1}$$

where $\xi(n) = c_0 n^{1/2+\varepsilon}$. Strong efficiency is a concept that is widely applied in the rare-event simulation community (see, for instance, Asmussen and Glynn (2007, Chapter VI) and Juneja and Shahabuddin (2006)). In our current context, strong efficiency implies that the number of replications required to obtain an estimator with a prescribed relative precision is uniformly bounded as $n \nearrow \infty$ over $b \geq \xi(n)$.

Using SDIS, Blanchet and Liu (2006) constructed an estimator satisfying (1). Let us summarize the ideas behind their development.

First, let us define $\delta = 1/n$ and write (for $t \in \{0, \delta, 2\delta, \dots, 1 - \delta\}$)

$$Y_\delta(t + \delta) = Y_\delta(t) + \delta X_{t/\delta+1}.$$

Note that $Y_\delta(t) = S_{tn}/n$ and $Y_\delta(0) = y = S_0/n$. Let $\eta_\delta = \delta b \geq c_0$, and set

$$u_\delta(t, y) = P_{(t,y)}(Y_\delta(1) \geq \eta_\delta),$$

where $P_{(t,y)}(\cdot)$ is the conditional probability given $Y_\delta(t) = y$. Using this notation, we have $P(Y(1) > \eta_\delta) = u_\delta(0, 0)$. It is well known that the function $u_\delta(\cdot)$ can be used to describe the zero-variance change of measure. Indeed, $Q_\delta^*(\cdot)$, defined via

$$Q_\delta^*(y_0, y_1, t) dy = P(Y_\delta(t + \delta) \in y_1 + dy_1 \mid Y_\delta(t) = y_0) \frac{u_\delta(t + \delta, y_1)}{u_\delta(t, y_0)},$$

describes a bona fide transition density, which can be verified by conditioning on the next transition of $Y_\delta(\cdot)$ at time t . Note that if path-generation according to $Q_\delta^*(\cdot)$ was feasible then the corresponding importance sampling estimator would be

$$Z^*(b) = \prod_{j=0}^{n-1} \frac{u_\delta(j\delta, Y_\delta(j\delta))}{u_\delta((j+1)\delta, Y_\delta((j+1)\delta))} = \frac{u_\delta(0, 0)}{u_\delta(1, Y_\delta(1))} = u_\delta(0, 0),$$

which estimates the target probability $u_\delta(0, 0)$ with zero variance. Obviously, if we had access to $Q_\delta^*(\cdot)$, there would not be any need for estimating $u_\delta(\cdot)$. Nevertheless, the basic idea is to take advantage of an approximation $v_\delta(\cdot)$ to $u_\delta(\cdot)$ and construct a change of measure based on $v_\delta(\cdot)$. This formal idea, which is known in the simulation literature (see Juneja and Shahabuddin (2006) for a review), was rigorously exploited by Blanchet and Liu (2006). In order to describe their sampling strategy, let us first recall a basic result from the large deviations theory for regularly varying RWs due to Rozovskii (1989).

Theorem 1. *Assume that the X_k s are regularly varying with finite mean and unit variance. Then*

$$u_\delta(0, 0) = \left(\delta^{-1} \bar{F}\left(\frac{\eta_\delta}{\delta}\right) + \bar{\Phi}(\eta_\delta \delta^{-1/2}) \right) (1 + o(1))$$

as $\delta \searrow 0$ uniformly over $\eta_\delta \geq \delta^{1/2}$, where $\bar{\Phi}(y) = P(N(0, 1) \geq y)$.

Using Theorem 1, Blanchet and Liu (2006) suggested using

$$v_\delta(t, y) := \left(\frac{(1-t)}{\delta} \bar{F}\left(\frac{(\eta_\delta - y)}{\delta}\right) + \bar{\Phi}\left(\frac{(\eta_\delta - y)\delta^{-1/2}}{(1-t)^{1/2}}\right) \right) \mathbf{1}((\eta_\delta - y)\delta^{-1/2} > (1-t)^{1/2}) + \mathbf{1}((\eta_\delta - y)\delta^{-1/2} \leq (1-t)^{1/2})$$

to construct a suitable Markov transition kernel that mimics the behavior of $Q_\delta^*(\cdot)$. In particular, Blanchet and Liu (2006) proposed using

$$Q_\delta(y_0, y_1, t) = P(Y_\delta(t + \delta) \in y_1 + dy_1 \mid Y_\delta(t) = y_0) \frac{v_\delta(t + \delta, y_1)}{w_\delta(t, y_0)},$$

where $w_\delta(t, y_0)$ is defined as

$$w_\delta(t, y_0) = E(v_\delta(t + \delta, y_0 + \delta X_1)).$$

Since $v_\delta(\cdot)$ is a suitable approximation to $u_\delta(\cdot)$ in a certain asymptotic sense, it is natural to expect that the ‘local’ likelihood ratio $(w_\delta(t, y_0)/v_\delta(t + \delta, y_1))$ will be well behaved (i.e. close to unity) on a certain region (described in terms of time and space) that represents the domain under which the asymptotics provide a description that is close enough to the behavior of $u_\delta(\cdot)$. So, it is intuitive that the performance of an algorithm based on $Q_\delta(\cdot)$ (in terms of the behavior of the local likelihood ratios and the ability to mimic the zero-variance change of measure) will be very good in the domain under which the asymptotics are guaranteed to also be good.

The algorithm of Blanchet and Liu (2006) follows next. We assume that the X_i s have a regularly varying density $f(\cdot)$ that is continuously differentiable, satisfying $f'(t) \sim t^{-(\alpha+2)} L(t)$ as $t \nearrow \infty$.

Algorithm 1. (Blanchet and Liu (2006).) The parameters $b \geq cn^{1/2+\varepsilon}$, $c > 0$, $\delta = 1/n$, and $\varepsilon > 0$ are given.

- (i) Set $s = 0$, $m = 0$, and $Z_A = 1$.
- (ii) *Step 1.* Sample X from the density $f_{m+1}(\cdot)$ defined via

$$f_{m+1}(x) = f(x) \frac{v_\delta(t + \delta, s + \delta x)}{w_\delta(t, s)}. \tag{2}$$

Update

$$\begin{aligned} Z_A &\leftarrow Z_A \frac{w_\delta(t, s)}{v_\delta(t + \delta, s + \delta X)}, \\ s &\leftarrow s + \delta X, \\ m &\leftarrow m + 1. \end{aligned}$$

- (iii) Repeat step 1 until $s > b - c(n - m)^{1/2+\varepsilon}/2$ or $m = n$.
- (iv) Step 2. **If** $m = n$ then **return** $Z_A \mathbf{1}(s > b)$, **else** generate $n - m$ i.i.d. RVs (X_1, \dots, X_{n-m}) according to the density $f(\cdot)$, evaluate $s' = X_1 + \dots + X_{n-m}$, and **return** $Z_A \mathbf{1}(s' + s > b)$.

The following result summarizes the performance of Blanchet and Liu’s (2006) estimator.

Theorem 2. (Blanchet and Liu (2006).) *Assume that the X_k s have a density $f(\cdot)$ satisfying $f'(t) \sim t^{-(\alpha+2)}L(t)$ with $\alpha > 2$. The estimator Z_A , provided by Algorithm 1, is strongly efficient according to (1).*

The proof of the previous result is given in Blanchet and Liu (2006). Although the previous result shows efficiency of the estimator Z_A , the simulation of the increments via the state-dependent density indicated in step 2 of the algorithm is not straightforward. These issues must be considered in the cost of generating each replication of the estimator Z_A . As we will see in Section 5, each replication can be done in approximately $O(bn^2)$ elementary operations (assuming that we count function evaluations, sums, and products as elementary operations).

3. Lyapunov inequalities and SDIS

In the previous section we described an SDIS algorithm that can be shown to be strongly efficient for the tail probability $P(S_n > b)$ as $n \nearrow \infty$, assuming that $b \geq cn^{1/2+\varepsilon}$. We also discussed, however, some issues related to the path generation and the computation of the normalizing constants that come up when implementing such an SDIS algorithm. An alternative approach consists in restricting our importance samplers to a parametric family for which path generation can be easily implemented. In order to achieve this, we first derive some Lyapunov inequalities that are applicable to general state-dependent importance samplers. We then apply these inequalities to a specific parametric family that is rich enough to mimic the behavior of the zero-variance change of measure in our current setting and for which path generation can be implemented easily.

A general SDIS strategy is characterized by transition densities $K_\delta(\cdot)$ of the form

$$K_\delta(y_0, y_1, t) dy = r_\delta(y_0, y_1, t)^{-1} P(Y_\delta(t + \delta) \in y_1 + dy_1 \mid Y_\delta(t) = y_0), \tag{3}$$

where $r_\delta(\cdot)$ represents the local likelihood ratio (i.e. $r_\delta(\cdot)^{-1}$ is a nonnegative, integrable function that is normalized in such a way that $K_\delta(\cdot)$ is a well-defined (time inhomogeneous) Markov transition density). Simulating according to $K_\delta(\cdot)$ gives rise to an estimator of the form

$$Z_K = \prod_{j=0}^{n-1} r_\delta(Y_\delta(j\delta), Y_\delta((j + 1)\delta), j\delta) \mathbf{1}(Y_\delta(1) > \eta_\delta).$$

The efficiency analysis of the estimator Z_K involves studying its second moment, namely, $E_{(0,y)}^K(Z_K^2)$. Throughout the rest of our discussion, we will use $E_{(t,y)}^K$ to denote the expectation operator in the path space induced by the transition density $K_\delta(y, \cdot, t)$, given that $Y_\delta(t) = y$.

The following result will be useful to bound the second moment of Z_K . Also, let $E_{(t,x)}$ be the conditional expectation given $Y(t) = x$.

Proposition 1. *Suppose that there exist a function $g_\delta: [0, 1) \times \mathbb{R} \rightarrow (0, \infty)$ and a constant $\rho \in (0, \infty)$ such that*

$$E_{(t,x)}(g_\delta(t + \delta, Y_\delta(t + \delta))r(x, Y_\delta(t + \delta), t)) \leq g_\delta(t, x) \tag{4}$$

and $g_\delta(1, x) \geq \rho$ for $x \geq \eta_\delta$. Then,

$$\begin{aligned} \frac{g_\delta(t, x)}{\rho} &\geq E_{(t,x)}^K \left(\prod_{j=t/\delta}^{n-1} r_\delta(Y_\delta(j\delta), Y_\delta((j + 1)\delta), j\delta)^2 \mathbf{1}(Y_\delta(1) \geq \eta_\delta) \right) \\ &= E_{(t,x)} \left(\prod_{j=t/\delta}^{n-1} r_\delta(Y_\delta(j\delta), Y_\delta((j + 1)\delta), j\delta) \mathbf{1}(Y_\delta(1) \geq \eta_\delta) \right). \end{aligned}$$

Proof. For $k \in \{t/\delta, t/\delta + 1, \dots, n\}$, define

$$M_k = g_\delta(k\delta, Y_\delta(k\delta)) \prod_{j=t/\delta}^{k-1} r_\delta(Y_\delta(j\delta), Y_\delta((j + 1)\delta), j\delta),$$

with $M_{t/\delta} = g_\delta(t, x)$. We claim that the process $(M_k: t/\delta \leq k \leq n)$ is a supermartingale with respect to the filtration $(\mathcal{F}_j: t/\delta \leq j \leq n)$, where $\mathcal{F}_j = \sigma(Y_\delta(m\delta): 0 \leq m \leq j)$. In order to see this, just note that

$$\begin{aligned} E(M_{k+1} \mid \mathcal{F}_k) &= \prod_{j=t/\delta}^{k-1} r_\delta(Y_\delta(j\delta), Y_\delta((j + 1)\delta), j\delta) \\ &\quad \times E(g_\delta((k + 1)\delta, Y_\delta((k + 1)\delta))r_\delta(Y_\delta(k\delta), Y_\delta((k + 1)\delta), j\delta) \mid \mathcal{F}_k) \\ &\leq g_\delta(k\delta, Y_\delta(k\delta)) \prod_{j=t/\delta}^{k-1} r_\delta(Y_\delta(j\delta), Y_\delta((j + 1)\delta), j\delta) \\ &= M_k. \end{aligned}$$

The conclusion of the proposition follows immediately after taking expectations and using the fact that $g_\delta(1, x) \mathbf{1}(x \geq \eta_\delta) \geq \mathbf{1}(x \geq \eta_\delta)$.

The main strategy behind our procedure involves the three steps discussed at the end of the introduction. First, we propose a parametric form for the function $r_\delta(\cdot)$. If we pick a parametric family that is powerful enough to capture the behavior of the zero-variance change of measure then $g_\delta(\cdot)$ should behave similarly as $u_\delta(\cdot)^2$ (recall the definition of efficiency given in (1)). The second step in the strategy is to propose a function $g_\delta(\cdot)$, possibly indexed by additional parameters, based on an approximation to $u_\delta(\cdot)^2$, which could be heuristic. Then, in the third and last step, we need to tune various parameters (both in $r_\delta(\cdot)$ and $g_\delta(\cdot)$) in order to force the Lyapunov inequality, (4). Throughout the rest of this section, we will illustrate the first two steps of this approach in the context of regularly varying RWs.

Large deviation probabilities for the empirical mean of regularly varying i.i.d. RVs is achieved by the contribution of a large jump. As a consequence, it is intuitive that a family of

changes of measures which induces this type of behavior is appropriate. Let us define

$$\lambda_\delta := \lambda_\delta(y) = \frac{\eta_\delta - y}{\delta} \quad \text{and} \quad \beta_\delta := \beta_\delta(t) = \frac{1 - t}{\delta}.$$

We then consider, for $a \in (0, 1)$,

$$r_\delta(y_0, y_1, t)^{-1} = p_t \frac{\mathbf{1}(y_1 - y_0 > a\delta\lambda_\delta(y_0))}{\bar{F}(a\lambda_\delta(y_0))} + (1 - p_t) \frac{\mathbf{1}(y_1 - y_0 \leq a\delta\lambda_\delta(y_0))}{F(a\lambda_\delta(y_0))}.$$

In other words, under the transition kernel $K(\cdot)$, given that $Y_\delta(t) = y_0$, the next increment has density $\tilde{f}(\cdot \mid y_0)$ defined as

$$\tilde{f}(x \mid y_0) = p_t \frac{f(x) \mathbf{1}(x > a\lambda_\delta(y_0))}{\bar{F}(a\lambda_\delta(y_0))} + (1 - p_t) \frac{f(x) \mathbf{1}(x \leq a\lambda_\delta(y_0))}{F(a\lambda_\delta(y_0))}$$

for some mixture probability p_t which will eventually be allowed to depend not only on time but also on the current position y_0 , and $a \in (0, 1)$. The parameter a is chosen to lie in $(0, 1)$ because we need to consider the contribution to the variance of such sample paths in which the rare event occurs, because of the occurrence of more than one single jump. The contribution of such sample paths is known to play an important role in the design of efficient rare-event simulation estimators; this has been noted several times since the development of the first efficient rare-event simulation estimator in a heavy-tailed setting in Asmussen *et al.* (2000). This strategy is over sampling those sample paths that take more than one jump to achieve the rare event. Nevertheless, as we will see, the inclusion of $a \in (0, 1)$ will prove to be useful in our technical development and will not preclude our estimator from achieving efficiency. An interesting exercise would be to send $a \nearrow 1$ as $\delta \searrow 0$ at some convenient rate in order to optimize performance.

Once we have fixed a parametric family of importance samplers that seems appropriate, we need to propose a parametric family of functions from which to pick $g_\delta(\cdot)$. For the reasons indicated in our discussion following Proposition 1, our guesswork should be guided by any approximation (rigorous or heuristic) available for $u_\delta(\cdot)^2$. Typically, in heavy-tailed problems, there are a number of fluid heuristics that are useful in order to develop plausible approximations for large deviations probabilities (see, for instance, Anantharam (1989) and Zwart (2001)). We can use such heuristics and then use Proposition 1 to develop a rigorous upper bound and, simultaneously, show strong efficiency. An alternative, when available, is to base the guesswork on rigorous approximations. In this case we have access to an approximation in Rozovskii (1989), which we stated in Theorem 1 and now use to construct our Lyapunov function.

First, let $F_1(\cdot)$ be any distribution function with a continuously differentiable density $f_1(\cdot)$ such that $f_1'(t) \sim t^{-(\alpha+2)}(\alpha + 2)(\alpha + 1)L(t)$ as $t \nearrow \infty$ (where $L(\cdot)$ is a slowly varying function at ∞ , that is, $L(tx)/L(t) \rightarrow 1$ as $t \nearrow \infty$ for each $x > 0$). By Karamata's theorem we have $\bar{F}_1(t) = 1 - F_1(t) \sim t^{-\alpha}L(t)$ as $t \nearrow \infty$ (in particular, $\bar{F}_1(\cdot)$ is slowly varying with index α). Moreover, we will pick $F_1(\cdot)$ so that $F_1(t) \sim F(t)$ as $t \nearrow \infty$. If $F(\cdot)$ is not twice continuously differentiable then we can construct $F_1(\cdot)$ by convolving $F(\cdot)$ with a Gaussian or an exponential kernel—these types of constructions are standard in the analysis of regularly varying functions (see, for instance, Resnick (1987, Chapter 0)). Then, set $\bar{D}_1(t) = (\bar{F}_1(t))^\nu$. Note that, for any given $\varepsilon > 0$, we can always choose $\nu > 0$ such that, for some $\beta_0 > 0$,

$$\bar{D}_1\left(\frac{\lambda_\delta(y)}{\beta_\delta(t)^{1/2+\varepsilon}}\right) < \beta_\delta(t)\bar{F}_1(\lambda_\delta(y)) \tag{5}$$

if $\lambda_\delta(y) \geq \beta_\delta(t)^{1/2+2\varepsilon}$ and $\beta_\delta(t) > \beta_0$. In order to have (5), we need

$$\alpha\nu \left(\frac{\log(\lambda_\delta(y))}{\log(\beta_\delta(t))} - \frac{1}{2} - \varepsilon \right) > \frac{\alpha \log(\lambda_\delta(y))}{\log(\beta_\delta(t))} - 1.$$

We will choose $\nu = \max\{1, (\alpha(1/2 + 2\varepsilon) - 1)/\alpha\varepsilon\} + 1$. The choice of ν is to make sure that $\beta_\delta(t)\bar{F}_1(\lambda_\delta(y))$ is the dominating term in the interesting area. Furthermore, we use the notation $\partial_t \bar{D}_1(t) = -d_1(t)$ throughout the rest of our development. Finally, for $t \in [0, 1 - \delta]$ and $\delta \in (0, 1)$, define

$$h_\delta(t, y) = \beta_\delta(t)\bar{F}_1(\lambda_\delta(y)) + \bar{D}_1\left(\frac{\lambda_\delta(y)}{\beta_\delta(t)^{1/2+\varepsilon}}\right),$$

set $h_\delta(1, y) = \mathbf{1}(\lambda_\delta(y) \leq 0)$, and, for $\kappa > 0$ to be specified later, write

$$g_\delta(t, y) = \min(\kappa h_\delta(t, y)^2, 1). \tag{6}$$

The Gaussian contribution in the approximation suggested by Theorem 1 has been replaced by \bar{D}_1 , such substitution is appropriate because our goal is to bound the second moment of an estimator for $P(S_n > b)$ with $b \geq n^{1/2+\varepsilon}$. Before we continue, it is important to remark on some basic properties of $g_\delta(\cdot)$.

Remark 1. As we will see, the region of points (t, y) where $g_\delta(t, y) < 1$ leads to the most interesting part in the development of the Lyapunov bound. In particular, note that if $g_\delta(t, y) < 1$ then $\bar{F}_1(\lambda_\delta(y)/\beta_\delta(t)^{1/2+\varepsilon}) < 1/\kappa^{1/2}$, which implies that

$$\lambda_\delta(y) > \beta_\delta(t)^{1/2+\varepsilon} m_0(\kappa) \rightarrow \infty \quad \text{as } \kappa \rightarrow \infty,$$

where $m_0(\kappa) = \bar{F}_1^{-1}(\kappa^{-1/2}) \nearrow \infty$ as $\kappa \nearrow \infty$ and $\bar{F}^{-1}(\cdot)$ is the inverse function of $\bar{F}(\cdot)$. This observation makes sure that we can choose κ large enough such that $\lambda_\delta(y)$ is large enough and all the asymptotics hold whenever $g_\delta(\cdot) < 1$.

4. Efficiency via Lyapunov bounds

In this section we discuss the third step of the proposed strategy to develop our efficient SDIS. Recall, as we mentioned in the introduction, that this step involves tuning the various parameters (p_t , a , and κ) in order to apply Proposition 1. The Lyapunov inequality, (4), takes the form

$$\begin{aligned} 1 &\geq \mathbb{E}\left(r_\delta(y, y + \delta X) \frac{g_\delta(t + \delta, y + \delta X)}{g_\delta(t, y)}\right) \\ &= \frac{\bar{F}(a\lambda_\delta(y))}{p_t} J_1 + \frac{F(a\lambda_\delta(y))}{1 - p_t} J_2, \end{aligned}$$

where

$$\begin{aligned} J_1 &= \mathbb{E}\left(\frac{g_\delta(t + \delta, y + \delta X)}{g_\delta(t, y)}; X > a\lambda_\delta(y)\right), \\ J_2 &= \mathbb{E}\left(\frac{g_\delta(t + \delta, y + \delta X)}{g_\delta(t, y)}; X \leq a\lambda_\delta(y)\right). \end{aligned}$$

We now analyze the expectations J_1 and J_2 . First we start with J_1 . Observe that, whenever $g_\delta(t, y) = 1$, we can pick $r_\delta(y, \cdot, t) = 1$, which automatically makes the Lyapunov function

satisfied, since $g_\delta \leq 1$. Picking $r_\delta(y, \cdot, t) = 1$ means that, for such a particular value of (t, y) , we do not apply importance sampling. Given this, we will assume that $g_\delta(t, y) < 1$ throughout the rest of our development. Again, since $g_\delta \leq 1$, after conditioning on the event $\{X > a\lambda_\delta(y)\}$, we obtain

$$\frac{\bar{F}(a\lambda_\delta(y))}{p_t} J_1 \leq \frac{\bar{F}(a\lambda_\delta(y))}{p_t} \frac{\bar{F}(a\lambda_\delta(y))}{g_\delta(t, y)} \leq \frac{\bar{F}(a\lambda_\delta(y)) \bar{F}_1(a\lambda_\delta(y))}{p_t \kappa h_\delta(t, y)^2}. \tag{7}$$

We propose selecting

$$p_t = \theta \frac{\bar{F}(a\lambda_\delta(y))}{h_\delta(t, y)} \tag{8}$$

for some $\theta > 0$ to be chosen later. This selection is sensible because p_t behaves roughly according to the probability that the next increment causes the rare event.

The most interesting part of the analysis involves the term J_2 , which we now discuss. First, note that by applying Taylor’s theorem with remainder we obtain, for each $z \in (-\infty, \infty)$,

$$g_\delta(t + \delta, y + \delta z) = g_\delta(t, y) + E(\partial_t g_\delta(t + \delta U, y + \delta z)\delta + \partial_y g_\delta(t, y)\delta z) + E(\partial_{yy} g_\delta(t, y + U U_1 \delta z)(\delta z)^2),$$

assuming that U and U_1 are independent and uniformly distributed on $(0, 1)$. We then conclude that

$$J_2 \leq 1 + J_{2,1} + J_{2,2} + J_{2,3},$$

where

$$J_{2,1} = E\left(\frac{\partial_t g_\delta(t + \delta U, y + \delta X)}{g_\delta(t, y)} \delta; X \leq a\lambda_\delta(y)\right),$$

$$J_{2,2} = E\left(\frac{\partial_y g_\delta(t, y)}{g_\delta(t, y)} \delta X; X \leq a\lambda_\delta(y)\right),$$

$$J_{2,3} = E\left(\frac{\partial_{yy}^2 g_\delta(t, y + U U_1 \delta X)}{2g_\delta(t, y)} (\delta X)^2; X \leq a\lambda_\delta(y)\right),$$

and U and U_1 are independent of X .

In order to analyze the previous expressions, we require some properties that are summarized in the following lemma. The proof is given at the end of this section. Nevertheless, we simply note that in the proof of this lemma the parameter $a \in (0, 1)$ is particularly useful. Indeed, as we can see from the proof, such a selection of a facilitates a uniform integrability argument that can be applied thanks to the regularly varying property.

Lemma 1. *There exists $\kappa_0 > 0$ such that if $\kappa \geq \kappa_0$ and $g_\delta(t, y) < 1$ then the following assertions hold.*

(i) *It is possible to find constants $\gamma_1, \gamma_2 > 0$ such that*

$$J_{2,1} \leq -\gamma_1 \frac{\bar{F}_1(\lambda_\delta(y))}{h_\delta(t, y)} - \gamma_2 \frac{d_1(\lambda_\delta(y)/\beta_\delta(t)^{1/2+\varepsilon})\lambda_\delta(y)/\beta_\delta(t)^{3/2+\varepsilon}}{h_\delta(t, y)}.$$

(ii) $J_{2,2} \leq 0$.

(iii) There exists $m_3 \in (0, \infty)$ such that

$$|J_{2,3}| \leq m_3 \frac{|\beta_\delta(t) \dot{f}_1(\lambda)| + |\dot{d}_1(\lambda_\delta(y)/\beta_\delta(t)^{1/2+\varepsilon})|/\beta_\delta(t)^{1+2\varepsilon}}{h_\delta(t, y)},$$

where $\dot{f}(x) = df(x)/dx$.

We need to introduce some additional notation in order to strip out the most important ingredients in the construction of our Lyapunov function. Let $m_{f_1}, m_{d_1}, m_F, m_{F_1} \in (0, \infty)$ such that

$$\sup_{t \geq 1} \frac{|\dot{f}_1(t)t^2|}{\bar{F}_1(t)} < m_{f_1}, \quad \sup_{t \geq 1} \frac{|\dot{d}_1(t)t|}{d_1(t)} < m_{d_1}, \tag{9}$$

$$\sup_{t \geq 1} \frac{\bar{F}(at)}{\bar{F}_1(t)} \leq m_F, \quad \sup_{t \geq 1} \frac{\bar{F}_1(at)}{\bar{F}_1(t)} \leq m_{F_1}. \tag{10}$$

The existence of these constants is guaranteed by virtue of regular variation and because the ratios on all the previous expressions are guaranteed to be bounded on any compact set. The next result shows the validity of the Lyapunov function for an appropriate selection of parameters.

Theorem 3. *There exist $a \in (0, 1)$, $\theta > 0$, and $\kappa \geq 0$ such that*

1. $\kappa > 0$ is selected satisfying the assumptions of Lemma 1;
2. $\theta > 0$ and $\kappa > 0$ satisfy the inequality

$$\frac{\gamma_1}{2} \geq 2\theta m_F + \frac{m_{F_1}}{\theta \kappa}$$

and also

$$p_t = \frac{\theta \bar{F}(\lambda_\delta(y))}{h_\delta(t, y)} \leq \frac{1}{2};$$

3. if $g_\delta(t, y) < 1$,

$$\begin{aligned} \frac{\bar{F}(a\lambda_\delta(y))}{p_t} J_1 + \frac{(1 - \bar{F}(a\lambda_\delta))}{(1 - p_t)} J_2 &\leq 1 - \left(\frac{\gamma_1}{2} - 2\theta m_F - \frac{m_{F_1}}{\theta \kappa} \right) \frac{\bar{F}_1(a\lambda_\delta)}{h_\delta} \\ &\leq 1. \end{aligned}$$

Proof. Throughout the proof, we will often drop the arguments in the functions in order to simplify the notation (for instance, we will write h_δ instead of $h_\delta(t, y)$). Lemma 1 yields

$$\begin{aligned} j_2 - 1 &\leq -\gamma_1 \frac{\bar{F}_1(\lambda_\delta)}{h_\delta} - \gamma_2 \frac{d_1(\lambda_\delta/\beta_\delta^{1/2+\varepsilon})\lambda_\delta/\beta_\delta^{3/2+\varepsilon}}{h_\delta} \\ &\quad + m_3 \frac{|\beta_\delta \dot{f}_1(\lambda_\delta)|}{h_\delta} + m_3 \frac{|\dot{d}_1(\lambda_\delta/\beta_\delta^{1/2+\varepsilon})|/\beta_\delta^{1+2\varepsilon}}{h_\delta}. \end{aligned}$$

We need to make sure that the negative contribution is larger than the positive contribution on the right-hand side of the above inequality. As we pointed out in Remark 1, $g_\delta(t, y) < 1$ implies that

$$\lambda_\delta \geq \beta_\delta^{1/2+\varepsilon} m_0(\kappa) \geq m_0(\kappa), \tag{11}$$

where $m_0(\kappa) \nearrow \infty$ as $\kappa \nearrow \infty$. Therefore, using the definition of m_{f_1} given in (9), we obtain

$$|\beta_\delta \dot{f}_1(\lambda_\delta)| \leq \bar{F}_1(\lambda_\delta) \left(m_{f_1} \frac{\beta_\delta}{\lambda_\delta^2} \right).$$

Using (11), we can choose κ large enough such that

$$\frac{\gamma_1 \bar{F}_1(\lambda_\delta)}{2 h_\delta} \geq \frac{m_3}{h_\delta} \bar{F}_1(\lambda_\delta) \left(\frac{m_{f_1}}{m_0(\kappa)^2 \beta_\delta^{2\varepsilon}} \right) \geq \frac{m_3}{h_\delta} \bar{F}_1(\lambda_\delta) \left(m_{f_1} \frac{\beta_\delta}{\lambda_\delta^2} \right) \geq m_3 \frac{|\beta_\delta \dot{f}_1(\lambda_\delta)|}{h_\delta}. \tag{12}$$

Similarly, we will control the contribution of the rest of the positive terms, using to our advantage the negative factor with coefficient γ_2 in Lemma 1(i). Using the definition of m_{d_1} given in (9), we obtain

$$\frac{|\dot{d}_1(\lambda_\delta/\beta_\delta^{1/2+\varepsilon})|}{\beta_\delta^{1+2\varepsilon}} \leq m_{d_1} \frac{d_1(\lambda_\delta/\beta_\delta^{1/2+\varepsilon})}{\lambda_\delta \beta_\delta^{1/2+\varepsilon}} = \frac{d_1(\lambda_\delta/\beta_\delta^{1/2+\varepsilon}) \lambda_\delta m_{d_1} \beta_\delta}{\beta_\delta^{3/2+\varepsilon} \lambda_\delta^2}.$$

Note that $\beta_\delta/\lambda_\delta^2 \leq m_0(\kappa)^{-2}$ (assuming, as we have been doing of course, that $g_\delta(t, y) < 1$). We find that the inequality

$$\gamma_2 \frac{d_1(\lambda_\delta/\beta_\delta^{1/2+\varepsilon}) \lambda_\delta / \beta_\delta^{3/2+\varepsilon}}{h_\delta} \geq m_3 \frac{|\dot{d}_1(\lambda_\delta/\beta_\delta^{1/2+\varepsilon})| / \beta_\delta^{1+2\varepsilon}}{h_\delta} \tag{13}$$

holds if κ is chosen large enough and $g_\delta(t, y) < 1$. Together with (12) and (13), we conclude that, for such selection of κ , as long as $g_\delta(t, y) < 1$,

$$\begin{aligned} J_2 &\leq 1 + J_{2,1} + J_{2,2} + J_{2,3} \\ &\leq 1 - \gamma_1 \frac{\bar{F}_1(\lambda_\delta)}{h_\delta} - \gamma_2 \frac{d_1(\lambda_\delta/\beta_\delta^{1/2+\varepsilon}) \lambda_\delta}{h_\delta \beta_\delta^{3/2+\varepsilon}} + m_3 \frac{|\beta_\delta \dot{f}_1| + |\dot{d}_1(\lambda_\delta/\beta_\delta^{1/2+\varepsilon})| / \beta_\delta^{1+2\varepsilon}}{h_\delta} \\ &\leq 1 - \frac{\gamma_1 \bar{F}_1(\lambda_\delta)}{2 h_\delta}. \end{aligned}$$

Recall that we are after a bound for $(1 - \bar{F}(a\lambda_\delta))J_2/(1 - p_t)$ (which then must be combined with our estimate involving J_1). We can always choose $\theta > 0$ small enough (but uniformly bounded away from 0 as $\delta \rightarrow 0$) so that we can guarantee $p_t \leq \frac{1}{2}$ for all values of (t, y) for which $g_\delta(t, y) < 1$. If $p_t \leq \frac{1}{2}$, we have $(1 - p_t)^{-1} \leq 1 + 2p_t$. Using this inequality, the definition of p_t , the fact that $\bar{F}(a\lambda_\delta) \in (0, 1)$, and also recalling the definition of m_F given in (10), we obtain

$$\begin{aligned} \frac{(1 - \bar{F}(a\lambda_\delta))J_2}{(1 - p_t)} &\leq \frac{J_2}{1 - p_t} \\ &\leq \frac{1 - \gamma_1 \bar{F}(\lambda_\delta)/2h_\delta}{1 - p_t} \\ &\leq \left(1 - \frac{\gamma_1 \bar{F}(\lambda_\delta)}{2 h_\delta} \right) (1 + 2p_t) \\ &\leq 1 - \left(\frac{\gamma_1}{2} - 2\theta m_F \right) \frac{\bar{F}_1(\lambda_\delta)}{h_\delta}. \end{aligned}$$

Now it is time to introduce the contribution of the term involving J_1 . Our selection of p_t yields, using bound (7) and the definition of m_{F_1} given in (10),

$$\frac{\bar{F}(a\lambda_\delta)}{p_t} J_1 \leq \frac{\bar{F}_1(a\lambda_\delta)}{\theta\kappa h_\delta} \leq \frac{m_{F_1} \bar{F}_1(\lambda_\delta)}{\theta\kappa h_\delta}.$$

We then conclude that if $\kappa > 0$ is selected large enough then, for all (t, y) for which $g_\delta(t, y) < 1$, we have

$$\frac{\bar{F}(a\lambda_\delta)}{p_t} J_1 + \frac{(1 - \bar{F}(a\lambda_\delta))J_2}{(1 - p_t)} \leq 1 - \left(\frac{\gamma_1}{2} - 2\theta m_F - \frac{m_{F_1}}{\theta\kappa} \right) \frac{\bar{F}_1(\lambda_\delta)}{h_\delta}.$$

The conclusion of the theorem follows after choosing $\theta > 0$ small enough and $\kappa > 0$ large enough so that

$$\frac{\gamma_1}{2} \geq 2\theta m_{F_1} + \frac{m_{F_1}}{\theta\kappa}.$$

The construction of the Lyapunov bound given in the previous theorem suggests an importance sampling algorithm that we now describe.

Algorithm 2. The parameters $b \geq cn^{1/2+\varepsilon}$, $c > 0$, $\delta = 1/n$, and $\varepsilon > 0$ are given. Select $a \in (0, 1)$ and $\theta, \kappa > 0$ satisfying the conditions of Theorem 3. This can be done analytically following the steps in the development.

(i) Set $s = 0$, $m = 0$, and $Z_B = 1$.

(ii) **for** $k \in \{0, 1, \dots, n - 1\}$ **do**

(a) If $g_\delta(m\delta, s\delta) < 1$ then let

$$p = \theta \frac{\bar{F}(a(b\delta - s)/\delta)}{h_\delta(m\delta, \delta s)},$$

if $g_\delta(m\delta, s\delta) = 1$ then let $p = \bar{F}(a(b\delta - s)/\delta)$. Sample X from the density $f_{m+1}(\cdot)$ defined via

$$f_{m+1}(x) = p \frac{f(x)}{\bar{F}(a(b\delta - s)/\delta)} \mathbf{1}(\delta x > a(\delta b - s)) + (1 - p) \frac{f(x)}{F(a(b\delta - s)/\delta)} \mathbf{1}(\delta x \leq a(\delta b - s)).$$

(b) Update

$$Z_B \leftarrow Z_B \left(\frac{\bar{F}(a(b\delta - s)/\delta) \mathbf{1}(\delta x > \delta b - s)}{p} + \frac{F(a(b\delta - s)/\delta) \mathbf{1}(\delta x \leq \delta b - s)}{1 - p} \right),$$

$$s \leftarrow s + \delta X,$$

$$m \leftarrow m + 1.$$

end for

(iii) Update

$$Z_B \leftarrow Z_B \bar{F} \frac{\delta b - s}{\delta}$$

and **return** Z_B .

We close this section with our main result, namely, the proof of strong efficiency of the estimator Z_B . In contrast to the estimator Z_A , the estimator Z_B requires only $O(n)$ elementary operations for its construction.

Theorem 4. *For any $\tilde{\varepsilon} > 0$, the estimator Z_B , provided by Algorithm 2, is strongly efficient when $b \geq n^{1/2+\tilde{\varepsilon}}$.*

Proof. Let us use $Q(\cdot)$ to denote the distribution induced by the importance sampling scheme given in the previous algorithm. Theorem 3 implies that

$$E_{(0,0)}^Q(Z_B^2) \leq \frac{g_\delta(0, 0)}{\min(\kappa, 1)}.$$

Recall that the choice of γ given ε in Section 3 implies that

$$g_\delta(0, 0) \leq 4\kappa\beta_\delta(0)^2 \bar{F}_1(\lambda_\delta(0))^2 = 4\kappa n^2 \bar{F}(b)^2$$

if $b > n^{1/2+2\varepsilon}$ and $n > \beta_0$. From the estimates of Roszovskii (1989) (note that here we actually need only an asymptotic lower bound, not exact asymptotics),

$$\sup_{\substack{n \geq 1 \\ b \geq n^{1/2+2\varepsilon}}} \frac{g_\delta(0, 0)}{P(S_n > b)^2} < \infty.$$

By letting $\tilde{\varepsilon} = 2\varepsilon$, it yields the conclusion of the theorem.

The pending proof of Lemma 1 is given next.

Proof of Lemma 1. We will start with part (i). Consider

$$E\left(\frac{\partial_t g_\delta(t + \delta U, y + \delta X)}{g_\delta(t, y)} \delta; X \leq a\lambda_\delta(y)\right).$$

Note that, for $u \in (0, 1)$, we have

$$\begin{aligned} \frac{\delta \partial_t g_\delta(t + \delta u, y + \delta x)}{g_\delta(t, y)} &= -\frac{2}{\delta} \frac{h_\delta(t + \delta u, y + \delta x)}{h_\delta(t, y)} \\ &\quad \times \left(\frac{\bar{F}(\lambda_\delta(y + \delta x))}{h_\delta(t, y)} \right. \\ &\quad \left. + \frac{(1/2 + \varepsilon)}{h_\delta(t, y)} \frac{\lambda_\delta(y + \delta x)}{\beta_\delta(t + u\delta)^{1/2+\varepsilon}} d_1\left(\frac{\lambda_\delta(y + \delta x)}{\beta_\delta(t + u\delta)^{1/2+\varepsilon}}\right) \right). \end{aligned}$$

We have $\lambda_\delta(y + \delta x) = \lambda_\delta(y) - x$ and $\beta_\delta(t + \delta u) = \beta_\delta(t) - u$. Also, as we pointed out in Remark 1, $g_\delta(t, y) < 1$ implies that $\lambda_\delta(y) \geq \beta_\delta(t)^{1/2+\varepsilon} m_0(\kappa) \geq m_0(\kappa) \nearrow \infty$ as $\kappa \nearrow \infty$. In particular, it follows from regular variation that if $x \leq a\lambda_\delta(y)$ and $g_\delta(t, y) < 1$ then there exists $\gamma'_{1,1} > 0$ such that, for all $\kappa > 0$ and $u \in (0, 1)$,

$$\inf_{\{(t,y): g_\delta(t,y) < 1, 0 \leq x \leq a\lambda_\delta(y)\}} \frac{h_\delta(t + \delta u, y + \delta x)}{h_\delta(t, y)} \geq \gamma'_{1,1} > 0.$$

Similarly, we have, for some $\gamma'_{1,2}, \gamma'_{1,3} > 0$,

$$\inf_{\{(t,y): g_\delta(t,y) < 1, 0 \leq x \leq a\lambda_\delta(y)\}} \frac{\bar{F}_1(\lambda_\delta(y + \delta x))}{\bar{F}_1(\lambda_\delta(y))} \geq \gamma'_{1,2} > 0$$

and

$$\frac{(\lambda_\delta(y) - x)\beta_\delta(t)^{1/2+\varepsilon}}{(\beta_\delta(t) - u)^{1/2+\varepsilon}\lambda_\delta(y)} \frac{d_1((\lambda_\delta(y) - x)/(\beta_\delta(t) - u)^{1/2+\varepsilon})}{d_1(\lambda_\delta(y)/\beta_\delta(t)^{1/2+\varepsilon})} \geq \gamma'_{1,3} > 0.$$

Then

$$\begin{aligned} J_{2,1} &\leq E\left(\frac{\delta \partial_t g_\delta(t + \delta U, y + \delta X)}{g_\delta(t, y)} \delta; 0 \leq X \leq a\lambda_\delta(y)\right) \\ &\leq -\gamma'_{1,1}\gamma'_{1,2} \frac{\bar{F}(\lambda_\delta)}{h_\delta} - \gamma'_{1,1}\gamma'_{1,3} \frac{d_1(\lambda_\delta/\beta_\delta^{1/2+\varepsilon})\lambda_\delta/\beta_\delta^{3/2+\varepsilon}}{h_\delta}. \end{aligned}$$

Part (i) then follows immediately from these estimates. Part (ii) is immediate by noting that $E(X; X \leq a\lambda_\delta(y)) < 0$. Part (iii) follows along the same lines as part (i).

5. A path-dependent example

In order to illustrate how to apply our method to some path-dependent examples we consider the problem of estimating

$$P\left(\min_{k \leq n} S_k \leq -n\eta_-, S_n > n\eta_+\right) \quad \text{for } \eta_-, \eta_+ \in (0, \infty). \tag{14}$$

We will assume that both the left and right tails of X_1 are regularly varying; in particular, we assume that, for each $x > 0$, $\bar{F}(xt)/\bar{F}(t) \rightarrow x^{-\alpha}$ as $t \nearrow \infty$ and $F(xt)/F(t) \rightarrow x^{-\beta}$ as $t \searrow -\infty$, where $\alpha, \beta > 2$. Moreover, let us assume that $\text{var}(X_1) = 1$. For simplicity, we will suppose that X_1 has a continuously differentiable positive density f ; this assumption can be relaxed using the smoothing technique explained at the end of Section 3.

The probability displayed in (14) can be interpreted as the chance of exercising a call-in financial option. Such an option gives the right to the owner of buying one unit of a stock at a specified price (called the strike price) at a given time (called the exercise time) as long as the price process hits a minimum threshold before the exercise time (i.e. the option is only activated when the price process hits a minimum level). The logarithm of the price process is modeled via the RW ($S_k: k \geq 0$), the strike price is represented by $\exp(n\eta_+)$, the time horizon is n , and the activation level is $\exp(-n\eta_-)$ (assuming that the initial price is equal to unity). A popular model used in financial applications is the so-called Black–Scholes model, which, in discrete time, postulates that the logarithm of price processes follows a Gaussian RW. Efficient simulation algorithms based on importance sampling have been studied for evaluating (14) under various light-tailed assumptions (see Glasserman (2004, Chapter 4) and Asmussen and Glynn (2007, Chapter VI)). A criticism of light-tailed models is that returns are often observed empirically to exhibit heavy tails. Instead of the Gaussian increments for the log-return process, we could consider, for instance, t -distributed increments or other types of heavy-tailed distributions. The method that we propose allows us to deal with general regularly varying log returns (including t -distributions) if the increments have finite second moment.

Contrary to the light-tailed case, the rate of decay to 0 of the probability (14) depends on the drift of the RW. For instance, if $E(X_1) < -\eta_-$, we expect (14) to be of order $O(n\bar{F}(n(\eta_+ + \eta_-)))$

as $n \nearrow \infty$ (this is because the law of large numbers will naturally drive the RW to the lower threshold, at which point there will be $O(n)$ units of time to reach boundary η_+). If $E(X_1) \in (-\eta_-, \eta_+ + \eta_-)$, we expect that (14) decreases to 0 at rate $O(n^2 \bar{F}(n\eta_+)F(-n\eta_-))$ as $n \nearrow \infty$. Similarly, if $E(X_1) > \eta_+ + \eta_-$, (14) is expected to be $O(nF(-n\eta_-))$ as $n \nearrow \infty$. The methods that we propose here can be adapted to deal with all the previous cases.

Throughout the rest of our discussion, we will assume that $\text{var}(X_1) = 1$ and $E(X_1) \in [0, \eta_+ + \eta_-)$ (situations where $E(X_1) \in (-\eta_-, 0)$, $E(X_1) < -\eta_-$, and $E(X_1) > \eta_+ + \eta_-$) can be adapted using the techniques that we explain next. The case that occupies our attention here is particularly interesting because the large deviations behavior dictated by (14) is caused by the influence of two large jumps (in contrast to the example in our previous section, which involves a single large jump). Our objective is to propose a state-depending importance sampling estimator, say Z_n , that can be guaranteed to be strongly efficient as $n \nearrow \infty$. In other words, we must have

$$\sup_{n \geq 1} \frac{E(Z_n^2)}{P(\min_{k \leq n} S_k \leq -n\eta_-, S_n > n\eta_+)^2} < \infty.$$

Once again, let $\delta = 1/n$, and consider $Y_\delta(t + \delta) = Y_\delta(t) + \delta X_{t/\delta+1}$ for $t \in \{0, \delta, 2\delta, \dots, 1\}$. Let $\tau_\delta = \inf\{u \geq 0: Y_\delta(u) \leq -\eta_-\}$. The estimator Z_n that we will construct takes the form

$$Z_n = \prod_{k=t/\delta}^{n-1} r_\delta(Y_\delta(k\delta), Y_\delta((k+1)\delta), j\delta) \mathbf{1}(\tau_\delta < 1-t, Y_\delta(1-t) > \eta_+)$$

with $t = 0$ and $y = 0$. The transitions of $Y_\delta(\cdot)$ are generated according to a kernel $K_\delta(\cdot)$ with form (3). As in Section 3, we use $P_{(t,y)}^K(\cdot)$ and $E_{(t,y)}^K(\cdot)$ to denote the associated probability measure and expectation operator induced by $K_\delta(\cdot)$.

In order to analyze the second moment of our estimator, we need to study

$$E_{(t,y)}^K \left(\prod_{k=t/\delta}^{n-1} r_\delta(Y_\delta(k\delta), Y_\delta((k+1)\delta))^2 \mathbf{1}(\tau_\delta < 1-t, Y_\delta(1-t) > \eta_+) \right).$$

In previous sections we derived a successful importance sampling strategy that we can apply for $t \geq \tau_\delta$. If we apply such a strategy then we can bound the previous expectation by

$$\begin{aligned} & E_{(t,y)}^K \left(\prod_{k=t/\delta}^{\tau_\delta/\delta-1} r_\delta(Y_\delta(k\delta), Y_\delta((k+1)\delta), k\delta)^2 \mathbf{1}(\tau_\delta < 1-t) g_\delta(\tau_\delta, Y_\delta(\tau_\delta)) \right) \\ & \leq E_{(t,y)}^K \left(\prod_{k=t/\delta}^{\tau_\delta/\delta-1} r_\delta(Y_\delta(k\delta), Y_\delta((k+1)\delta), k\delta)^2 \mathbf{1}(\tau_\delta < 1-t) g_\delta(\tau_\delta, -\eta_-) \right) \\ & = E_{(t,y)}^K \left(\prod_{k=t/\delta}^{\tau_\delta/\delta-1} r_\delta(Y_\delta(k\delta), Y_\delta((k+1)\delta), k\delta) \mathbf{1}(\tau_\delta < 1-t) g_\delta(\tau_\delta, -\eta_-) \right). \end{aligned}$$

Just as we did in previous sections, we will construct a Lyapunov function to bound the previous expectation. The next proposition, whose proof is completely analogous to that of Proposition 1, provides the means to develop such a Lyapunov function.

Proposition 2. *Suppose that there exists a function $\tilde{g}_\delta: [0, 1) \times \mathbb{R} \rightarrow (0, \infty)$ such that*

$$E_{(t,x)}(\tilde{g}_\delta(t + \delta, Y_\delta(t + \delta))r(x, Y_\delta(t + \delta), t)) \leq \tilde{g}_\delta(t, x)$$

for $x > -\eta_-$ and $0 \leq t \leq 1 - \delta$. Moreover, assume that $\tilde{g}_\delta(t, x) \geq \rho g_\delta(t, -\eta_-)$ on $x \leq -\eta_-$ and $t \leq 1 - \delta$ for some $\rho \in (0, \infty)$. Then,

$$\frac{\tilde{g}_\delta(t, x)}{\rho} \geq E_{(t,y)}^K \left(\prod_{k=t/\delta}^{\tau_\delta/\delta-1} r_\delta(Y_\delta(k\delta), Y_\delta((k+1)\delta), k\delta)^2 \mathbf{1}(\tau_\delta < 1-t) g_\delta(\tau_\delta, -\eta_-) \right).$$

With this proposition in hand, we can develop a successful importance sampling scheme that we can apply whenever $t < \tau_\delta$. We will develop such a scheme using the three-step program described in the previous section. Namely, (i) propose a parametric family of importance samplers that capture the large deviations behavior of interest, (ii) propose an educated guess for the form of a convenient Lyapunov function \tilde{g}_δ , and (iii) tune the parameters in the proposed parametric family of changes of measure and the form of the candidate function \tilde{g}_δ to force the Lyapunov inequality.

Finding a convenient family of importance sampling distributions when $t < \tau_\delta$ is not difficult. Note that $\tau_\delta < 1 - t$ is a rare event that is caused by a single downwards jump. So, given $Y_\delta(t) = y$ and $\tau_\delta > t$, we propose to sample the next increment according to the density

$$f_{X|y}(x) = \frac{p_1 f(x) \mathbf{1}(x \leq -a\lambda_{\delta,1}(y))}{F(-a\lambda_{\delta,1}(y))} + \frac{(1 - p_1) f(x) \mathbf{1}(x > -a\lambda_{\delta,1}(y))}{\bar{F}(-a\lambda_{\delta,1}(y))},$$

where $a, p_1 \in (0, 1)$ and $\lambda_{\delta,1}(y) = (y + \eta_-)/\delta$.

Our educated guess comes from a crude large deviations estimate. We have $(1 - t)/\delta$ increments and one of them must be negative enough in order to make the event $\{\tau_\delta < 1 - t\}$ occur. On the other hand, as the distance to the lower threshold approaches the central limit theorem scaling, the event $\{\tau_\delta < 1 - t\}$ becomes rarer. As a consequence, we propose to define, for $t \in [0, 1 - \delta]$,

$$\tilde{h}_\delta(t, y) = \frac{1-t}{\delta} F(-\lambda_{\delta,1}(y)) + D_1 \left(-\frac{\lambda_{\delta,1}(y)}{\beta_\delta(t)^{1/2+\varepsilon}} \right),$$

where $\beta_\delta(t) = (1 - t)/\delta$ and $D_1(x) = F(x)^2$. This function is completely analogous to the function $h_\delta(\cdot)$ that we defined in Section 3; this should not be surprising because certainly $P(S_n \leq -n\eta_-) \sim P(\min_{k \leq n} S_k \leq -n\eta_-)$ as $n \nearrow \infty$ (see Borovkov and Borovkov (2001)). We then set

$$\tilde{g}_{\delta,1}(t, y) = \min(\tilde{\kappa} \tilde{h}_\delta(t, y)^2, 1),$$

and propose as our candidate Lyapunov function

$$\tilde{g}_\delta(t, y) = \tilde{g}_{\delta,1}(t, y) g_\delta(t, -\eta_-),$$

where g_δ is defined in (6).

Since $g_\delta(\cdot, -\eta_-)$ is decreasing, in order to verify Proposition 2, we need to show that $\tilde{g}_{\delta,1}(t, y) \geq \rho$ for $y \leq -\eta_-$ and

$$\frac{\tilde{J}_1 F(-a\lambda_{1,\delta}(y))}{p_1} + \frac{\tilde{J}_2 \bar{F}(-a\lambda_{1,\delta}(y))}{1 - p_1} \leq 1, \tag{15}$$

where

$$\begin{aligned} \tilde{J}_1 &= E\left(\frac{\tilde{g}_{\delta,1}(t + \delta, y + \delta X)}{\tilde{g}_{\delta,1}(t, y)}; X \leq -a\lambda_{1,\delta}(y)\right), \\ \tilde{J}_2 &= E\left(\frac{\tilde{g}_{\delta,1}(t + \delta, y + \delta X)}{\tilde{g}_{\delta,1}(t, y)}; X > -a\lambda_{1,\delta}(y)\right). \end{aligned}$$

Note that $\inf\{\tilde{h}_\delta(t, y) : 0 \leq t \leq 1 - \delta, y \leq -\eta_-\} > 0$; so there exists $\rho > 0$ such that $\tilde{g}_{\delta,1}(t, y) \geq \rho$ for $y \leq -\eta_-$. We just need to concentrate on showing that (15) holds whenever $\tilde{g}_{\delta,1}(t, y) < 1$ and for this we need to analyze \tilde{J}_1 and \tilde{J}_2 much in the same manner as we did for J_1 and J_2 . In particular, we select

$$p_1 = \frac{\tilde{\theta} F(-a\lambda_{1,\delta}(y))}{\tilde{h}_\delta(t, y)}$$

for $\tilde{\theta} > 0$ appropriately chosen. The parameters $\tilde{\theta}, \tilde{\kappa} > 0$ are selected in order to force the Lyapunov inequality in a completely analogous way as we did in the previous section. In particular, we can show that there exist a $\tilde{\kappa}_0 > 0$ and a constant $\tilde{\gamma} > 0$ so that if $\tilde{\kappa} \geq \tilde{\kappa}_0$ and $\tilde{g}_{\delta,1}(t, y) < 1$ then

$$\begin{aligned} &\frac{\tilde{J}_1 F(-a\lambda_{1,\delta}(y))}{p_1} + \frac{\tilde{J}_2 \bar{F}(-a\lambda_{1,\delta}(y))}{1 - p_1} \\ &\leq 1 - \left(\frac{\tilde{\gamma}}{2} - 2\tilde{\theta}\tilde{m}_F - \frac{\tilde{m}_F}{\tilde{\theta}\tilde{\kappa}}\right) \frac{F(-a\lambda_{1,\delta}(y))}{\tilde{h}_\delta(t, y)}, \end{aligned} \tag{16}$$

where

$$\sup_{t \geq 1} \frac{\bar{F}(at)}{F_1(t)} \leq \tilde{m}_F < \infty.$$

We then immediately conclude the following result.

Theorem 5. *Given $a \in (0, 1)$, it is possible to select $\tilde{\theta} > 0$ and $\tilde{\kappa} \geq 0$ satisfying (16) whenever $\tilde{g}_{\delta,1}(t, y) < 1$, and so*

$$\left(\frac{\tilde{\gamma}}{2} - 2\tilde{\theta}\tilde{m}_F - \frac{\tilde{m}_F}{\tilde{\theta}\tilde{\kappa}}\right) \geq 0.$$

As a consequence, \tilde{g}_δ satisfies the conditions of Proposition 2.

The description of the algorithm suggested by the previous development is given next.

Algorithm 3. Set η_-, η_+ , and $\delta = 1/n$. Select $a \in (0, 1)$ and $\tilde{\theta}, \tilde{\kappa} > 0$ satisfying the conditions of Theorem 5.

(i) Set $\tilde{s} = 0, \tilde{m} = 0, Z_C = 1$, and $hit = 0$.

(ii) **while** $hit = 0$ **or** $\tilde{m} < n - 2$ **do**

(a) **if** $\tilde{g}_{\delta,1}(\tilde{m}\delta, \tilde{s}\delta) < 1$ then let

$$p_1 = \tilde{\theta} \frac{F(-a(\eta_- + \tilde{s})/\delta)}{\tilde{h}_\delta(\tilde{m}\delta, \delta\tilde{s})}$$

and sample X from the density $f_{\tilde{m}+1}(\cdot)$ defined via

$$f_{\tilde{m}+1}(x) = p_1 \frac{f(x)}{F(-a(\eta_- + \tilde{s})/\delta)} \mathbf{1}(\delta x \leq -a(\eta_- + \tilde{s})) + (1 - p_1) \frac{f(x)}{\bar{F}(-a(\eta_- + \tilde{s})/\delta)} \mathbf{1}(\delta x > -a(\eta_- + \tilde{s})).$$

Update

$$Z_C \leftarrow Z_C \left(\frac{F(-a(\eta_- + \tilde{s})/\delta) \mathbf{1}(\delta X \leq -a(\eta_- + \tilde{s}))}{p_1} + \frac{\bar{F}(-a(\eta_- + \tilde{s})/\delta) \mathbf{1}(\delta X > -a(\eta_- + \tilde{s}))}{1 - p_1} \right)$$

else

Sample X from $f(\cdot)$.

end if

(b) Update

$$\tilde{s} \leftarrow \tilde{s} + \delta X, \quad \tilde{m} \leftarrow \tilde{m} + 1.$$

if $\tilde{s} \leq -\eta_-$

Update $hit \leftarrow 1$.

end if

end while

(iii) **if** $hit = 0$

(a) Sample X with its conditional law given that $\{\delta X \leq -(\eta_- + \tilde{s})\}$. Let

$$Z_C \leftarrow Z_C F\left(-\frac{\eta_- + \tilde{s}}{\delta}\right).$$

(b) Update $\tilde{s} \leftarrow \tilde{s} + \delta X$ and **return**

$$Z_C \leftarrow Z_C \bar{F} \frac{\eta_+ - \tilde{s}}{\delta}$$

else

Run Algorithm 2 with initial parameters $b = n\eta_+ - \tilde{s}$, $Z_B = Z_C$, and $s = \tilde{s}$, and let Z_C be the output of Algorithm 2 (that was denoted by Z_B), given the specified parameters.

(c) **return** Z_C .

We conclude the section with a result that summarizes the efficiency of the previous algorithm's estimator.

Theorem 6. *The estimator given by Algorithm 3 is strongly efficient in the sense that*

$$\sup_{n \geq 1} \frac{E_{(0,0)}^K(Z_C^2)}{P(\min_{k \leq n} S_k \leq -n\eta_-, S_n > n\eta_+)} < \infty.$$

Proof. Sharp asymptotics can be derived for the probability of interest (i.e. the denominator in the previous ratio); however, to keep the discussion self-contained and to show the simplicity of the method, we content ourselves with a crude estimate. In particular, it follows easily that

$$\begin{aligned} & \mathbb{P}\left(\min_{k \leq n} S_k \leq -n\eta_-, S_n > n\eta_+\right) \\ & \geq \mathbb{P}\left(\min_{k \leq n/2} S_k \leq -n\eta_-\right) \inf_{n/2 \leq k \leq n} \mathbb{P}(S_k > n(\eta_- + \eta_+)) \\ & \sim \left(\frac{n}{2}\right)^2 F(-n\eta_-) \bar{F}(n(\eta_+ + \eta_-)). \end{aligned}$$

Using Theorem 5 and defining

$$\rho = \min(\tilde{\kappa} \inf\{\tilde{h}_\delta(t, y) : 0 \leq t \leq 1 - 2\delta, y \leq -\eta_-\}^2, 1) > 0,$$

we have

$$\begin{aligned} \mathbb{E}_{(0,0)}^K(Z_C^2) & \leq \frac{\tilde{g}_\delta(0, 0)}{\rho} \\ & = \frac{\tilde{g}_{\delta,1}(0, 0)}{\rho} g_{\delta,1}(0, -\eta_-) \\ & \sim \frac{\tilde{\kappa}\kappa}{\rho} n^4 F(-n\eta_-)^2 \bar{F}(n\eta_-)^2. \end{aligned}$$

The conclusion of the theorem then follows from regular variation.

6. Numerical experiments

We will discuss the implementation of the importance sampling strategies described in Algorithm 1 and Algorithm 2.

6.1. Implementing Algorithm 1

A couple of issues arise regarding the implementation of Algorithm 1. The first one relates to the path generation under the change of measure and the computation of the $w(n - i - 1, s)$ s. The path generation can be done using acceptance/rejection, as in Blanchet and Glynn (2007), and the estimation of $w(n - i - 1, s)$ can be performed by numerical integration (see also Blanchet and Liu (2007) for a Monte Carlo idea based on path sampling that can be used even in higher dimensions). In this paper we adopt an alternative approach, in which we use some other, easy-to-evaluate change of measure, $\tilde{f}_i(\cdot)$ as an approximation to $f_i(\cdot)$ (defined in (2)). Using this density, we follow the same procedure described in Algorithm 1. The new estimator is

$$\begin{aligned} \tilde{Z}_A & = \mathbf{1}(S_n > b) \prod_{i=1}^{\tau-1} \frac{f(X_i)}{\tilde{f}_i(X_i)} \\ & = \mathbf{1}(S_n > b) \prod_{i=1}^{\tau-1} \frac{f_i(X_i)}{\tilde{f}_i(X_i)} \frac{f(X_i)}{f_i(X_i)}, \end{aligned}$$

where $\tau = \inf\{k \leq n : S_k > b - c(n - k)^{1/2+\varepsilon}/2\}$. If we can control the approximation error

so that $f_i(x)/\bar{f}_i(x) \leq (1 + cn^{-1})$, uniformly over x , then we have

$$E(\tilde{Z}_A^2) \leq e^{2c} E\left(\prod_{i=1}^{\tau-1} \frac{f(X_i)^2}{\bar{f}_i(X_i)^2} \mathbf{1}(S_n > b)\right) \leq C P(S_n > b)^2$$

for some $C > 0$ and, therefore, \tilde{Z}_A^2 is strongly efficient. Note that $f_i(x) = f(x)v(n - i, s + x)/w(n - i + 1, s)$. One choice for $\bar{f}_i(\cdot)$ is

$$\begin{aligned} \bar{f}_i(x) &= \frac{1}{\bar{w}(n - i + 1, s)} \sum_{k=1}^m g_k(n - i, s + x) \mathbf{1}(x_{k-1} < x \leq x_k) \\ &\quad + \frac{1}{\bar{w}(n - i + 1, s)} \mathbf{1}(x > x_m) f(x) + v(n - i, x_1) f(x) \mathbf{1}(x \leq x_1), \end{aligned}$$

where $\{x_i\}_1^m$ is an appropriate partition and

$$\begin{aligned} g_k(n - i, s + x) &= \max_{[x_{k-1}, x_k]} f(x)v(n - i, s + x), \\ \bar{w}(n - i + 1, s) &= v(n - i, x_1)F(x_1) + \bar{F}(x_m) + \sum_{k=1}^m g_k(n - i, s + x)(x_k - x_{k-1}). \end{aligned}$$

Under the assumptions of Theorem 2, it follows that if we choose $x_m = b - s - \sqrt{n}$ and $x_{k+1} - x_k = O(1/n)$ then we can guarantee that, for some $c > 0$,

$$\sup_{i \leq n, s} \frac{\bar{w}(n - i + 1, s)}{w(n - i + 1, s)} \leq \left(1 + \frac{c}{n}\right). \tag{17}$$

As a consequence, the total number of elementary operations required to compute \bar{L} is roughly of order $O(bn^2)$. As a particular example, we consider the following model.

Example 1. Define $X = \Lambda R$, where $P(\Lambda > x) = \min(x^{-4}, 1)$, $R \sim \text{Laplace}(1)$, and Λ is independent of R . As we will see below, we can compute the density function of X and its characteristic function analytically. Therefore, we are able to compare our estimates with an estimate based on numerical transform inversion for $P(S_n > b)$. Note that

$$\begin{aligned} E(\exp(i\theta \Lambda R)) &= E\left(\frac{1}{1 + \Lambda^2 \theta^2}\right) \\ &= \int_1^\infty \frac{1}{1 + x^2 \theta^2} \frac{4}{x^5} dx \\ &= \int_1^\infty \frac{1}{1 + x^2 \theta^2} dx \\ &= 1 - 2\theta^2 + 2\theta^4 \log\left(1 + \frac{1}{\theta^2}\right) \\ &= 1 - 2\theta^2 + 2\theta^4 \sum_{k=1}^\infty \frac{\theta^{-2k}}{k} (-1)^{k+1} \\ &= 2\theta^4 \sum_{k=3}^\infty \frac{\theta^{-2k}}{k} (-1)^{k+1}. \end{aligned} \tag{18}$$

The tail of X is

$$\begin{aligned} \bar{F}(t) &= E(\Lambda R > t) \\ &= \frac{1}{2} E(e^{-t/\Lambda}) \\ &= \int_1^\infty e^{-t/x} \frac{2}{x^5} dx \\ &= 2 \int_0^1 e^{-tx} x^3 dx \\ &= \frac{-2}{t^4} e^{-tx} (6 + 6tx + 3t^2x^2 + t^3x^3) \Big|_0^1 \\ &= \frac{2}{t^4} (6 - e^{-t} (6 + 6t + 3t^2 + t^3)) \end{aligned}$$

for all $t > 0$, and the corresponding density is

$$f(x) = \frac{8}{x^5} (6 - e^{-x} (6 + 6x + 3x^2 + x^3)) - \frac{2}{x} e^{-x}.$$

6.2. Implementing Algorithm 2

In order to apply Algorithm 2 to estimate $P(Y_\delta(1) > 1)$, we choose a change of measure based on the density, given

$$f(x | y, t) = f(x) \left(p_t \frac{\mathbf{1}(x > a(1 - y)/\delta)}{\bar{F}(a(1 - y)/\delta)} + (1 - p_t) \frac{\mathbf{1}(x \leq a(1 - y)/\delta)}{F(a(1 - y)/\delta)} \right).$$

The Lyapunov function takes the form $g_\delta(t, y) = \min(\kappa h_\delta(t, y)^2, 1)$, where

$$h_\delta(t, y) = \beta_\delta(t) \bar{F}(\lambda_\delta(y)) + \bar{F}(\lambda_\delta(y)/\beta_\delta(t))^{0.8}.$$

We have chosen $a = 0.5$ and $p_t = \min(\bar{F}(\lambda_\delta(y))/h_\delta(t, y), 0.3)$ (this selection of p_t corresponds to choosing $\theta = 1$); finally, we select $\kappa = 25$.

The comparisons between the analytic approximation given by Theorem 1 (displayed under LDA for ‘large deviations approximation’), and the estimates under Algorithm 1 and Algorithm 2 are shown in Table 1. The values under ‘true value’ were computed using a numerical Fourier transform. Keep in mind that, despite having smaller relative error, Algorithm 2 is actually preferable when one considers the fact that Algorithm 1 requires about $O(n^2)$ times more operations than Algorithm 1 to produce each replication. We used 10 000 replications for each of the estimators displayed.

TABLE 1: Numerical results for Algorithms 1 and 2, where sd is the standard deviation based on 10 000 independent simulations.

n	$P(S_n > n)$		Algorithm 1			Algorithm 2		
	True value	LDA	Z_A	sd	CV	Z_B	sd	CV
100	2.21×10^{-5}	1.20×10^{-5}	2.37×10^{-5}	2.27×10^{-6}	10.3	2.09×10^{-5}	1.04×10^{-6}	4.7
500	1.04×10^{-7}	9.60×10^{-8}	1.02×10^{-7}	1.00×10^{-9}	1.0	1.11×10^{-7}	4.31×10^{-9}	4.1
1000	1.25×10^{-8}	1.20×10^{-8}	1.23×10^{-8}	1.39×10^{-10}	1.1	1.16×10^{-8}	4.72×10^{-10}	3.8

TABLE 2: Numerical results for Algorithm 3.

n	$P(S_n > n, \min_k S_k < -n)$								
	10 000 replications			100 000 replications			1000 000 replications		
	Z_C	sd	CV	Z_C	sd	CV	Z_C	sd	CV
50	6.82×10^{-10}	9.38×10^{-11}	10.4	8.99×10^{-10}	1.07×10^{-10}	37.4	9.05×10^{-10}	3.65×10^{-11}	40.3
100	6.74×10^{-12}	1.24×10^{-12}	19.2	6.42×10^{-12}	3.29×10^{-13}	16.2	6.42×10^{-12}	1.23×10^{-13}	19.2
500	2.64×10^{-16}	3.94×10^{-17}	16.0	2.55×10^{-16}	1.25×10^{-17}	16.0	2.45×10^{-16}	3.87×10^{-18}	15.7
1000	4.62×10^{-18}	6.65×10^{-19}	18.4	3.90×10^{-18}	1.89×10^{-19}	16.5	3.63×10^{-18}	5.72×10^{-20}	15.8

Example 2. We assume that X follows the same distribution as in Example 1. We want to compute $P(S_n > n, \min_k S_k < -n)$. We let $a = 0.5$,

$$p_t = 0.3 \wedge \begin{cases} \frac{2\bar{F}(-a\lambda_{1,\delta}(y))}{\tilde{h}_\delta(t, y)} & \text{if } \tau_{-n} > t, \\ \frac{\bar{F}(a\lambda_\delta(y))}{h_\delta(t, y)} & \text{otherwise,} \end{cases}$$

where $\tau_{-n} = \inf\{t : y(t) < -n\}$. We select $\kappa = \tilde{\kappa} = 25$. In Table 2 we present the estimates using different numbers of replications, which show that the estimates were stabilized for a million replications. Also, simply as a verification of our code, we compute the probability with $n = 10$, using 100 000 replications. The estimate is 1.95×10^{-5} , with standard error 8.82×10^{-7} . We also use crude Monte Carlo to compute the same probability, based on 10^8 replications, which yields an estimate of 1.94×10^{-5} with standard error 4.41×10^{-7} .

Acknowledgement

This research was partially supported by DMS NSF 0595595 grant.

References

- ADLER, J., FELDMAN, R. AND TAQQU, M. (eds) (1998). *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhäuser, Boston, MA.
- ANANTHARAM, V. (1989). How large delays build up in a GI/G/1 queue. *Queueing Systems* **5**, 345–367.
- ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd edn. Springer, New York.
- ASMUSSEN, S. AND GLYNN, P. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer, New York.
- ASMUSSEN, S. AND KROESE, D. P. (2006). Improved algorithms for rare event simulation with heavy tails. *Adv. Appl. Prob.* **38**, 545–558.
- ASMUSSEN, S., BINSWANGER, K. AND HØJGAARD, B. (2000). Rare event simulation for heavy-tailed distributions. *Bernoulli* **6**, 303–322.
- BASSAMBOO, A., JUNEJA, S. AND ZEEVI, A. (2005). Importance sampling simulation in the presence of heavy tails. In *Proc. 37th Conf. Winter Simulation* (December 2005), IEEE, pp. 664–672.
- BLANCHET, J. AND GLYNN, P. (2008). Efficient rare-event simulation for the maximum of heavy-tailed random walks. *Ann. Appl. Prob.* **18**, 1351–1378.
- BLANCHET, J. AND LIU, J. C. (2006). Efficient simulation for large deviation probabilities of sums of heavy-tailed increments. In *Proc. 38th Conf. Winter Simulation* (December 2006), IEEE, pp. 757–764.
- BLANCHET, J. AND LIU, J. C. (2007). Path-sampling for state-dependent importance sampling. In *Proc. 39th Conf. Winter Simulation* (December 2007), IEEE, pp. 380–388.
- BLANCHET, J., GLYNN, P. AND LIU, J. C. (2007). Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue. *Queueing Systems* **57**, 99–113.
- BLANCHET, J., GLYNN, P. AND LIU, J. C. (2008). Efficient rare event simulation for heavy-tailed multiserver queues. Department of Statistics, Columbia University.

- BOROVKOV, A. A. AND BOROVKOV, K. A. (2001). On large deviations of probabilities for random walks. I. Regularly varying distribution tails. *Theory Prob. Appl.* **49**, 193–213.
- BUCKLEW, J. (2004). *Introduction to Rare Event Simulation*. Springer, New York.
- DEMBO, A. AND ZEITOUNI O. (1998). *Large Deviations Techniques and Applications*, 2nd edn. Springer, New York.
- DONSKER, M. AND VARADHAN, S. (1975). Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Commun. Pure Appl. Math.* **28**, 1–47.
- DUPUIS, P. AND WANG, H. (2004). Importance sampling, large deviations, and differential games. *Stoch. Stoch. Reports* **76**, 481–508.
- DUPUIS, P. AND WANG, H. (2005). On the convergence from discrete to continuous time in an optimal stopping problem. *Ann. Appl. Prob.* **15**, 1339–1366.
- DUPUIS, P., LEDER, K. AND WANG, H. (2007). Importance sampling for sums of random variables with regularly varying tails. *ACM Trans. Model. Comput. Simul.* **17**, 1–21.
- FOSS, S. AND KORSHUNOV, D. (2006). Heavy tails in multi-server queues. *Queueing Systems* **52**, 31–48.
- FOSS, S., KONSTANTOPOULOS, T. AND ZACHARY, S. (2007). Discrete and continuous time modulated random walks with heavy-tailed increments. *J. Theoret. Prob.* **20**, 581–612.
- GLASSERMAN, P. (2004). *Monte Carlo Methods in Financial Engineering* (Appl. Math. **53**). Springer, New York.
- GLASSERMAN, P. AND KOU, S. (1995). Analysis of an importance sampling estimator for tandem queues. *ACM Trans. Model. Comput. Simul.* **4**, 22–42.
- GLASSERMAN, P. AND WANG, Y. (1997). Counterexamples in importance sampling for large deviations probabilities. *Ann. Appl. Prob.* **7**, 731–746.
- HULT, H. AND LINDSKOG, F. (2005). Extremal behavior for regularly varying stochastic processes. *Stoch. Process. Appl.* **115**, 249–274.
- HULT, H., LINDSKOG, F., MIKOSCH, T. AND SAMORODNITSKY, G. (2005). Functional large deviations for multivariate regularly random walks. *Ann. Appl. Prob.* **15**, 2651–2680.
- JUNEJA, S. AND SHAHABUDDIN, P. (2002). Simulating heavy tailed processes using delayed hazard rate twisting. *ACM Trans. Model. Comput. Simul.* **12**, 94–118.
- JUNEJA, S. AND SHAHABUDDIN, P. (2006). Rare event simulation techniques: introduction and recent advances. In *Handbook on Simulation*, eds S. Henderson and B. Nelson, North-Holland, Amsterdam, pp. 291–350.
- NAGAEV, A. V. (1969a). Integral limit theorems for large deviations when Cramér’s condition is not fulfilled. I. *Theory Prob. Appl.* **14**, 51–64.
- NAGAEV, A. V. (1969b). Integral limit theorems for large deviations when Cramér’s condition is not satisfied. II. *Theory Prob. Appl.* **14**, 193–208.
- RESNICK, S. (1987). *Extreme Values, Regular Variation, and Point Processes* (Appl. Prob. **4**). Springer, New York.
- ROZOVSKII, L. V. (1989). Probabilities of large deviations of sums of independent random variables with common distribution function in the domain of attraction of the normal law. *Theory Prob. Appl.* **34**, 625–644.
- SADOWSKY, J. S. AND BUCKLEW, J. (1990). On large deviations theory and asymptotically efficient Monte Carlo estimation. *IEEE Trans. Inf. Theory* **36**, 579–588.
- ZWART, A. (2001). *Queueing systems with heavy tails*. Doctoral Thesis, Technische Universiteit Eindhoven.