

Efficient Simulation and Conditional Functional Limit Theorems for Ruinous Heavy-tailed Random Walks

Jose Blanchet ^{*} and Jingchen Liu [†]

June 14, 2010

Abstract

The contribution of this paper is to introduce change of measure based techniques for the rare-event analysis of heavy-tailed stochastic processes. Our changes-of-measure are parameterized by a family of distributions admitting a mixture form. We exploit our methodology to achieve two types of results. First, we construct Monte Carlo estimators that are strongly efficient (i.e. have bounded relative mean squared error as the event of interest becomes rare). These estimators are used to estimate both rare-event probabilities of interest and associated conditional expectations. We emphasize that our techniques allow us to control the expected termination time of the Monte Carlo algorithm even if the conditional expected stopping time (under the original distribution) given the event of interest is infinite – a situation that sometimes occurs in heavy-tailed settings. Second, the mixture family serves as a good approximation (in total variation) of the conditional distribution of the whole process given the rare event of interest. The convenient form of the mixture family allows us to obtain, as a corollary, functional conditional central limit theorems that extend classical results in the literature. We illustrate our methodology in the context of the ruin probability $P(\sup_n S_n > b)$, where S_n is a random walk with heavy-tailed increments that have negative drift. Our techniques are based on the use of Lyapunov inequalities for variance control and termination time. The conditional limit theorems combine the application of Lyapunov bounds with coupling arguments.

1 Introduction

Change-of-measure techniques constitute a cornerstone in the large deviations analysis of stochastic processes (see for instance [16]). In the light-tailed setting, it is well understood that a specific class of changes-of-measure, namely exponential tilting, provide just the right vehicle to perform not only large deviations analysis but also to design provably efficient importance sampling simulation estimators. There is a wealth of literature on structural

^{*}Research supported in part by DMS-0902075 and CMMI-0846816

[†]Research supported in part by Institute of Education Sciences, U.S. Department of Education, through Grant R305D100017

results, such as conditional limit theorems, that justify the use of exponential changes of measure in these settings (see for instance [1, 7] in the setting of random walks and [19] in the context of networks).

Our contributions in this paper is the introduction of change-of-measure techniques for the rare-event analysis of heavy-tailed stochastic processes. Our general motivation is to put forward tools that allow to perform both, large deviations analysis for heavy-tailed systems and, at the same time, construction of efficient Monte Carlo algorithms for estimation of rare events, in the same spirit as in light-tailed settings. To this end, we introduce a family of changes of measures that are parameterized by a mixture of finitely many distributions and develop mathematical tools for their analyses. We concentrate on a class of problems of interest both in queueing theory and risk theory, namely first passage time probabilities for random walks, which serve as a good stylized model for testing and explaining techniques at the interface of large deviations and simulation. For instance, the first paper ([23]) that introduced the notations of efficiency together with the application of light-tailed large deviations ideas and exponential changes-of-measure, focused on this class of model problems. Such notations are now standard in rare-event simulation. In the heavy-tailed setting, first passage time problems for random walks also serve as an environment for explaining the challenges that arise when trying to develop efficient importance sampling estimators (see [3]). We will provide additional discussion on those challenges and contrast our methods here with recent approaches that have been developed for first passage time problems for heavy-tailed random walks. We will illustrate the flexibility of our method in terms of simulation estimators that have good variance performance and good control on the cost per replication of the simulation estimator. The proposed change of measure also satisfies structural results (in the form of conditional limit theorems) in the spirit of the theory that has been developed in light-tailed environments. Let us introduce the setup that will be the focus of our paper.

Let $S = \{S_n : n \geq 0\}$ be a random walk with independently and identically distributed (i.i.d.) increments, $\{X_n : n \geq 1\}$, that is, $S_{n+1} = S_n + X_{n+1}$ for all $n \geq 0$ and $S_0 = 0$. We assume that $\mu = EX_n < 0$ and that the X_n 's are suitably heavy-tailed (see Section 2). For each $b \in \mathbb{R}^+$, let $\tau_b = \inf\{n \geq 1 : S_n > b\}$. Of interest in this paper is the first passage time probability¹

$$u(b) = P(\tau_b < \infty), \tag{1}$$

and the conditional distribution of the random walk given $\{\tau_b < \infty\}$, namely

$$P(S \in \cdot | \tau_b < \infty). \tag{2}$$

This paper introduces a family of unbiased simulation estimators for $u(b)$ that can be shown to have bounded coefficient of variation uniformly over $b > 0$. The associated sampling distribution approximates (2) in total variation as $b \rightarrow \infty$. Unbiased estimators with bounded coefficient of variation are called *strongly efficient estimators* in rare event simulation.

The construction of provably efficient importance sampling estimators has been the focus of many papers in the applied probability literature. A natural idea behind the construction of efficient importance sampling estimators is that one should mimic the behavior of the zero

¹If $S_0 = 0$ we use $P(\cdot)$ and $E(\cdot)$ to denote the associated probability measure and expectation operators in path space, respectively. If $S_0 = s$, then we write $P_s(\cdot)$ and $E_s(\cdot)$.

variance change of measure, which coincides precisely with the conditional distribution (2). As it is well known, heavy-tailed large deviations are often governed by the “principle of the big jump”, which, qualitatively speaking, indicates that asymptotically as $b \rightarrow \infty$ the event of interest (in our case $\{\tau_b < \infty\}$) occurs due to the contribution of a single large increment of size $\Omega(b)$.² Consequently, the principle of the big jump naturally suggests to mimic the zero variance change of measure by a distribution which assigns zero probability to the event that ruin occurs due to the contribution of more than one large jump of order $\Omega(b)$. However, such an importance sampling strategy is not feasible because it violates absolute continuity requirements to define a likelihood ratio. This is the most obvious problem that arises in the construction of efficient importance sampling schemes for heavy-tailed problems. A more subtle problem discussed in [3] is the fact that the second moment of an importance sampling estimator for heavy-tailed large deviations is often very sensitive to the behavior of the likelihood ratio precisely on paths that exhibit more than one large jump for the occurrence of the rare event in question. We shall refer to those paths that require more than one large jump for the occurrence of the event $\tau_b < \infty$ *rogue paths*.

In the last few years state-dependent importance sampling has been used as a viable way to construct estimators for heavy-tailed rare-event simulation. A natural idea is to exploit the Markovian representation of (2) in terms of the so-called Doob’s h-transform. In particular, it is well known that

$$P(X_{n+1} \in dx | S_n, n < \tau_b < \infty) = \frac{u(b - S_n - x)}{u(b - S_n)} F(dx), \quad (3)$$

where F is the distribution of X_{n+1} . In [9], a state dependent importance sampling estimator based on an approximation to (3) is constructed and a technique based on Lyapunov inequalities was introduced for variance control. In particular, by constructing a suitable Lyapunov function, in [9], it is shown that if $v(b - s)$ is a suitable approximation to $u(b - s)$ as $b - s \nearrow \infty$ and $w(b - s) = Ev(b - s - X)$ then simulating the increment X_{n+1} given S_n and $\tau_b > n$ via the distribution

$$\tilde{P}(X_{n+1} \in dx | S_n) = \frac{v(b - S_n - x)}{w(b - S_n)} F(dx) \quad (4)$$

provides a strongly efficient estimator for $u(b)$. This approach provided the first provably efficient estimator for $u(b)$ in the context of a general class of heavy-tailed increment distributions, the class S^* , which includes in particular Weibull and regularly varying distributions. Despite the fact that the importance sampling strategy induced by (4) has been proved to be efficient in substantial generality, it has a few inconvenient features. First, it typically requires to numerically evaluate $w(b - S_n)$ for each S_n during the course of the algorithm. Although this issue does not appear to be too critical in the one dimensional setting (see the analysis in [11]), for higher dimensional problems, the numerical evaluation of $w(b - S_n)$ could easily require a significant computational overhead. For instance, see the first passage time computations for multiserver queues, which have been studied in the regularly varying

²For $f(\cdot)$ and $g(\cdot)$ non-negative we use the notation $f(b) = O(g(b))$ if $f(b) \leq cg(b)$ for some $c \in (0, \infty)$. Similarly, $f(b) = \Omega(g(b))$ if $f(b) \geq cg(b)$ and we also write $f(b) = o(g(b))$ as $b \rightarrow \infty$ if $f(b)/g(b) \rightarrow 0$ as $b \rightarrow \infty$.

case in [10]. The second inconvenient feature is that if the increments have finite mean but infinite variance we obtain $E(\tau_b | \tau_b < \infty) = \infty$. The strategy of mimicking the conditional distribution without paying attention to the cost per replication of the estimator could yield a poor overall computational complexity. Our proposed approach does not suffer from this drawback because our parametric family of changes-of-measures allows to control both the variance and the termination time.

We now proceed to explicitly summarize the contributions of this paper. Further discussion will be given momentarily and precise mathematical statements are given in Section 2.2.

1. We provide a strongly efficient estimator (i.e. bounded relative mean squared error as $b \nearrow \infty$) to compute the rare event probabilities $u(b)$ and the associated conditional expectations, based on a finite mixture family, for which both the simulation and density evaluation are straightforward to perform (see Theorem 1). Several features of the algorithm include:
 - (a) The results require the distribution to have an eventually concave cumulative hazard function, which includes a large class of distributions including regularly varying, Weibull distribution, log-normal distribution and so forth (see assumptions in Section 2).
 - (b) One feature of the proposed algorithm relates to the termination time. When the increments are regularly varying with tail index $\iota \in (1, 2)$, $E(\tau_b | \tau_b < \infty) = \infty$. This implies that the zero-variance change of measure takes infinity expected time to generate one sample. In contrast, we show that the proposed importance sampling algorithm takes $O(b)$ expected time to generate one sample while still maintaining strong efficiency if $\iota \in (1.5, 2)$ –Theorem 3.
 - (c) For the case that $\iota \in (1, 1.5]$, we show that the $(1 + \gamma)$ -th moment of the estimator is of order $O(u^{1+\gamma}(b))$ with $\gamma > 0$ depending on ι . In addition, the expected termination time of the algorithm is $O(b)$ (Theorem 4). Therefore, to compute $u(b)$ with ε relative error and at least $1 - \delta$ probability, the total computation complexity is $O(b)$.
2. The mixture family approximates the conditional distribution of the random walk given ruin in total variation. Based on this strong approximation and on the simplicity of the mixture family’s form, we derive a conditional functional central limit theorem of random walk given ruin which further extends existing results reported in [8] (compare Theorems 2, 5 and 6 below).

As mentioned earlier, the simulation estimators proposed in this paper are based on importance sampling and they are designed to directly mimic the conditional distribution of S given $\tau_b < \infty$ based on the principle of the big jump. This principle suggests that one should mimic the behavior of such a conditional distribution at each step by a mixture of two components: one involving an increment distribution that is conditioned to reach level b and

a second one corresponding to a nominal (unconditional) increment distribution. This two-mixture sampler, which was introduced by [17] in the context of tail estimation of a fixed sum of heavy-tailed random variables, has been shown to produce strongly efficient estimators for regularly varying distributions [17, 14, 13, 12]. However, two-component mixtures are not suitable for the design of strongly efficient estimators in the context of other types of heavy-tailed distributions. In particular, two-component mixtures are not applicable to general semiexponential distributions (see [15] for the definition) including Weibull and log-normal.

As indicated, one of our main contributions in this paper is to introduce a generalized finite-mixture sampler that can be shown to be suitable for constructing strongly efficient estimators in the context of a general class of heavy-tailed distributions, beyond regularly varying tails and including lognormals and Weibullian-type tails. Our mixture family also mimics the qualitative behavior mentioned above; namely, there is the contribution of a large jump and the contribution of a regular jump. In addition, one needs to control the behavior of the likelihood ratio corresponding to rogue sample paths. Depending on the degree of concavity of the cumulative hazard function (which we assume to be eventually strictly concave) we must interpolate between the large jump component and the nominal component in a suitable way. At the end, the number of mixtures is larger for cumulative hazard functions that are less concave.

Our mixture family and our Lyapunov based analysis allow to obtain an importance sampling scheme that achieves strong efficiency and *controlled expected termination time even if the optimal (in terms of variance minimization) change of measure involves an infinite expected termination time*. More precisely, if the increment distribution is regularly varying with tail index $\iota \in (1, 2)$ it follows using the Pakes-Veraverbeke theorem (see Theorem 7) that

$$\begin{aligned} E(\tau_b | \tau_b < \infty) &= \sum_{n=0}^{\infty} \frac{P(\tau_b > n, \tau_b < \infty)}{P(\tau_b < \infty)} \\ &\geq \sum_{n=1}^{\infty} \frac{P(\tau_{b-\mu n/2} < \infty) P(|S_n + n\mu| \leq n|\mu|/2)}{P(\tau_b < \infty)} = \infty. \end{aligned}$$

Nevertheless, as we will show, if $\iota \in (1.5, 2]$ we can choose the mixture parameters (which are state-dependent) in such a way that (using $E^Q(\cdot)$ to denote the probability measure induced by our importance sampling strategy assuming $S_0 = 0$)

$$E^Q \tau_b = O(b) \tag{5}$$

while maintaining strong efficiency. We believe this feature is surprising! In particular, it implies that one can construct a family of estimators for expectations of the form $E(H(S_k : k \leq \tau_b) | \tau_b < \infty)$ that requires overall $O(b)$ random numbers generated uniformly over a class of functions such that $0 < K_0 \leq H \leq K_1 < \infty$, even if $E(\tau_b | \tau_b < \infty) = \infty$. We shall also informally explain why $\iota > 1.5$ appears to be a necessary condition in order to construct an unbiased estimator satisfying both strong efficiency and (5).

In addition, for the case that $\iota \in (1, 1.5]$, we are able to construct an estimator whose $(1 + \gamma)$ -th moment (for $0 < \gamma < (\iota - 1)/(2 - \iota)$) is of order $O(u^{1+\gamma}(b))$ while the expected termination time is $O(b)$. We will also argue that the bound on γ is essentially optimal.

Consequently, as it is shown in Theorem 4, to compute $u(b)$ with ε relative error and at least $1 - \delta$ probability, the total computation complexity is $O(b)$.

In addition to providing a family of strongly efficient estimators for $u(b)$, our finite-mixture family can approximate the conditional measure (2) in total variation as $b \nearrow \infty$. This approximation step further strengthens our family of samplers as a natural rare-event simulation scheme for heavy-tailed systems. Moreover, given the strong mode of convergence and because the mixture family admits a friendly form, we are able to strengthen classical results in the literature on heavy tailed approximations, see [8]. For instance, if a given increment has second moment, we will derive, as a corollary of our approximations, a conditional functional central limit theorem up to the first passage time τ_b . Thereby, this improves the law of large numbers derived in [8]. Another related result in the setting of high dimension regularly varying random walk is given in [21]. We believe that the proof techniques behind our approximations, which are based on coupling arguments, are of independent interest and that they can be used in other heavy-tailed environments.

A central technique in the analysis of both the computational complexity and our conditional limit theorems is the use of Lyapunov functions. The Lyapunov functions are used for three different purposes: First in showing the strong efficiency of the importance sampling estimator, second in providing a bound on the finite expected termination time of the algorithm, and finally in proving the approximation in total variation of the zero-variance change of measure. The construction of Lyapunov functions follows the so called fluid heuristic, which is well known in the literature of heavy-tailed large deviations and has also been successfully applied in rare event simulation, see [14, 13, 12].

This paper is organized as follows. In Section 2, we introduce our assumptions, our family of changes of measures and we provide precise mathematical statements of our main results. Section 3 discusses some background results on large deviations and Lyapunov inequalities for importance sampling and stability of Markov processes. The variance analysis of our estimators is given in Section 4. The results corresponding to the termination time of our algorithm can be found in Section 5. Then we have our results on strong conditional limit theorems in Section 6. We provide numerical experiments in Section 7. Finally, we added an appendix which contains auxiliary lemmas and technical results.

2 Main Results

We shall use X to denote a generic random variable with the same distribution as any of the X_i 's describing the random walk $S_n = \sum_{i=1}^n X_i$, for $n = 1, 2, \dots$ with $S_0 = 0$. We write $F(x) = P(X \leq x)$, $\bar{F}(x) = P(X > x)$ and $EX = \mu \in (-\infty, 0)$. Further, let $\Lambda(\cdot)$ be the cumulative hazard function and $\lambda(\cdot)$ be the hazard function. Therefore, F has density function, for $x \in (-\infty, \infty)$

$$f(x) = \lambda(x)e^{-\Lambda(x)}, \quad \text{and} \quad \bar{F}(x) = e^{-\Lambda(x)}.$$

Of primary interest to us is the design of efficient importance sampling (change of measure based) estimators for

$$u(b) = P(\max_{n \geq 1} S_n > b) = P(\tau_b < \infty), \tag{6}$$

as $b \rightarrow \infty$ when F is suitably heavy-tailed. In particular, throughout this paper we shall assume either of the following two sets of conditions:

Assumption A: F has a regularly varying right tail with index $\iota > 1$. That is,

$$\bar{F}(x) = 1 - F(x) = L(x)x^{-\iota},$$

where $L(\cdot)$ is a slowly varying function at infinity, that is, $\lim_{x \rightarrow \infty} L(xt)/L(x) = 1$.

Or

Assumption B: There exists $b_0 > 0$ such that for all $x \geq b_0$ the following conditions hold.

B1 Suppose that $\lim_{x \rightarrow \infty} x\lambda(x) = \infty$.

B2 There exists $\beta_0 \in (0, 1)$ such that $\partial \log \Lambda(x) = \lambda(x)/\Lambda(x) \leq \beta_0 x^{-1}$ for $x \geq b_0$.

B3 Assume that $\Lambda(\cdot)$ is concave for all $x \geq b_0$; equivalently, $\lambda(\cdot)$ is assumed to be non increasing for $x \geq b_0$.

B4 Assume that

$$P(X > x + t/\lambda(x) | X > x) = \exp(-t) (1 + o(1))$$

as $x \nearrow \infty$ uniformly over compact sets in $t \geq 0$. In addition, for some $\alpha > 1$, $P(X > x + t/\lambda(x) | X > x) \leq t^{-\alpha}$ for all $t, x > b_0$.

Remark 1 *The analysis requires $\Lambda(\cdot)$ to be differentiable only for $x \geq b_0$. The reason for introducing Assumptions A and B separately is that the analysis for regularly varying distributions is somewhat different from (easier than) the cases under Assumption B. Assumption B1 implies that the tail of X decays faster than any polynomial. Assumptions B2 and B3 basically say that the cumulative hazard function of F is “more concave” than at least some Weibull distribution with shape parameter $\beta_0 < 1$. Typically, the more concave the cumulative hazard function is, the heavier the tail is. Therefore, under Assumption B, F is basically assumed to have a heavier tail than at least some Weibull distribution with shape parameter $\beta_0 < 1$. Assumption B4 is required only in Theorem 6 which states the functional central limit theorem of the conditional random walk given ruin. Note that the Assumptions A and B cover a wide range of heavy-tailed distributions that are popular in practice, for instance, regularly varying, log-normal, Weibull with $\beta_0 \in (0, 1)$ and so forth.*

In our random walk context, state-dependent importance sampling involves studying a family of densities (depending on “current” state s of the random walk) which governs subsequent increments of the random walk. More precisely, we write

$$q_s(x) = r_s(x)^{-1} f(x),$$

where $r_s(\cdot)$ is a non-negative function such that $E r_s(X) = 1$ for a generic family of state-dependent importance sampling increment distributions. If we let $Q(\cdot)$ represent the probability measure in path-space induced by the subsequent generation of increments under $q_s(\cdot)$, then it follows easily that

$$u(b) = E^Q[I(\tau_b < \infty) L_b],$$

with

$$L_b = \sum_{j=1}^{\tau_b} r_{S_{j-1}} (S_j - S_{j-1}). \quad (7)$$

We say that

$$Z_b = I(\tau_b < \infty) L_b \quad (8)$$

is an importance sampling estimator for $u(b)$ and its second moment is simply

$$E^Q[I(\tau_b < \infty) L_b^2] = E[I(\tau_b < \infty) L_b].$$

If we select $Q(\cdot) = P(\cdot | \tau_b < \infty)$, or equivalently we let $r_s(x) = u(b-s-x)/u(b-s)$, then the corresponding importance sampling estimator would yield zero variance. Hence, we call it zero-variance importance sampling estimator; and we call $P(\cdot | \tau_b < \infty)$ the zero-variance change of measure or zero-variance importance sampling distribution.

One of our main goals in this paper is to show that we can approximate the zero-variance change of measure quite accurately using finitely many mixtures whose parameters can be easily computed in advance. As a consequence, we can use Monte Carlo simulation to not only accurately estimate $u(b)$ but also associated conditional expectations of the random walk given $\tau_b < \infty$. In fact, we can improve upon the zero variance change of measure in terms of overall computational cost when it comes to estimating sample-path conditional expectations given $\tau_b < \infty$ in situations where $E(\tau_b | \tau_b < \infty) = \infty$. The precise mathematical statements are given later in this section. Future sections are dedicated to the development and the proofs of these statements.

Before stating the main results, we would first introduce the family of change of measures which is based on a mixture of finitely many computable and simulatable distributions.

2.1 The mixture family

We start by describing the precise form of the mixtures that we will use to construct efficient importance sampling schemes. The family is constructed to consider the contribution of a “large jump” which makes the walk reach level b in the next step, a “regular jump” which allows the random walk to continue under (nearly) its original dynamics, and a number of “interpolating” contributions. This intuition is consistent with the way in which large deviations occur in heavy-tailed environments.

If $b-s > \eta_*$ for $\eta_* > 0$ sufficiently large and to be specified in our analysis, we propose to use a finite mixture family of the form

$$q_s(x) = p_* f_*(x|s) + p_{**} f_{**}(x|s) + \sum_{j=1}^k p_j f_j(x|s), \quad (9)$$

where $p_*, p_{**}, p_j \in [0, 1)$, $p_* + p_{**} + \sum_{j=1}^k p_j = 1$, $k \in \mathbb{N}$, and f_*, f_{**} , and f_j for $j = 1, \dots, k$ are properly normalized density functions, whose supports are disjoint and depend on the “current” position of the walk, s . We will give specific forms momentarily. The choice of k depends on the concavity of the cumulative hazard function, but otherwise is independent of b and s . We will ultimately let p_*, p_{**} and the p_j ’s depend on s . In addition, we will also

choose *not to apply importance sampling* if we are suitably close to the boundary level b . In other words, overall we have that

$$q_s(x) = \left[p_* f_*(x|s) + p_{**} f_{**}(x|s) + \sum_{j=1}^k p_j f_j(x|s) \right] I(b-s > \eta_*) + f(x) I(b-s \leq \eta_*). \quad (10)$$

We next specify the functional forms of each mixture distribution. First,

$$f_*(x|s) = f(x) \frac{I(x \leq b-s - \Lambda^{-1}(\Lambda(b-s) - a_*))}{P(X \leq b-s - \Lambda^{-1}(\Lambda(b-s) - a_*))},$$

where $a_* > 0$. So, f_* represents the mixture component corresponding to a “regular” increment.

Further, for $a_{**} > 0$, let

$$f_{**}(x|s) = f(x) \frac{I(x > \Lambda^{-1}(\Lambda(b-s) - a_{**}))}{P(X > \Lambda^{-1}(\Lambda(b-s) - a_{**}))}.$$

f_{**} represents the mixture component that the rare event occurs because this particular increment is large. Note that

$$P(X > b-s | X > \Lambda^{-1}(\Lambda(b-s) - a_{**})) = \exp(-a_{**}).$$

Therefore, if the “next increment”, X , given the current position, s , is drawn from f_{**} , there is probability $1 - \exp(-a_{**}) > 0$ that the next position of the random walk, namely $s + X$, is below the threshold b . This particular feature is important in the variance control. It is necessary to introduce such a positive a_{**} to achieve strong efficiency if we want to consider the possibility of rogue paths in our sampler.

As we mentioned before, the choice of k depends on the “concavity” of the cumulative hazard function $\Lambda(\cdot)$. The more concave $\Lambda(\cdot)$ is, the smaller k one can usually choose. In the regularly varying case, for example, a two-mixture distribution is sufficient (i.e. $k = 0$). The analysis of importance sampling algorithms in this case has been substantially studied in the literature (see [17, 14, 13, 12]). We can see that this feature is captured in our current formulation because in the regularly varying case one can find $a_*, a_{**} > 0$ such that

$$b-s - \Lambda^{-1}(\Lambda(b-s) - a_*) \geq \Lambda^{-1}(\Lambda(b-s) - a_{**}), \quad (11)$$

for all $b-s$ large enough so that one can choose $k = 0$. Indeed, to see how (11) holds for the regularly varying case, just note that for any $a \in (0, 1)$, for each t , the inequality

$$at \geq \Lambda^{-1}(\Lambda(t) - a_{**})$$

is equivalent to

$$\frac{P(X > at)}{P(X > t)} \leq \exp(a_{**}). \quad (12)$$

Similarly,

$$t - \Lambda^{-1}(\Lambda(t) - a_*) \geq at$$

holds if and only if

$$\frac{P(X > (1-a)t)}{P(X > t)} \leq \exp(a_*). \quad (13)$$

Karamata's theorem for regularly varying distributions ensures that it is always possible to choose $a_*, a_{**} > 0$ given any $a \in (0, 1)$ so that (12) and (13) hold for uniformly in t and therefore we have that (11) holds. If Assumption A holds, we choose a_{**} and then select a_* (possibly depending on $b - s$) such that

$$b - s - \Lambda^{-1}(\Lambda(b - s) - a_*) = \Lambda^{-1}(\Lambda(b - s) - a_{**}). \quad (14)$$

This selection is slightly different from the two-mixture form that has been analyzed in the literature (see [14, 13, 12]) which involves a “regular” component with support on $(-\infty, a(b - s)]$ and a “large jump” component with support on $(a(b - s), \infty)$, for $a \in (0, 1)$. Our analysis here also applied to this parameterization. Nevertheless, to have unified statements in our results, under both Assumptions A and B, we opted for using equation (14).

When (11) does not hold (for instance in the case of Weibull tails with shape parameter $\beta \in (0, 1)$), we will need more mixtures. In particular, we consider a set of cut-off points $c_0 < \dots < c_k$ depending on $b - s$. Ultimately, we will have

$$c_j = a_j(b - s) \quad \text{for } j = 1, 2, \dots, k - 1.$$

where $a_1 < \dots < a_{k-1}$. The a_j 's are precomputed depending on β_0 (from Assumption B3) according to Lemma 9 (Section 4). We let $c_0 = b - s - \Lambda^{-1}(\Lambda(b - s) - a_*)$ and $c_k = \Lambda^{-1}(\Lambda(b - s) - a_{**})$. Given these values we define for $1 \leq j \leq k - 1$,

$$f_j(x) = f(x) \frac{I(x \in (c_{j-1}, c_j])}{P(X \in (c_{j-1}, c_j])}.$$

For $j = k$,

$$f_k(x) = f(b - s - x) \frac{I(x \in (c_{k-1}, c_k])}{P(X \in (b - s - c_k, b - s - c_{k-1}])}.$$

In our previous notation, we then can write

$$\begin{aligned} & r_s(x)^{-1} \\ &= \left(\frac{p_* I(x \leq c_0)}{P(X \leq c_0)} + \frac{p_{**} I(x > c_k)}{P(X > c_k)} + \sum_{j=1}^{k-1} \frac{p_j I(x \in (c_{j-1}, c_j])}{P(X \in (c_{j-1}, c_j])} + \frac{f(b - s - x) p_k I(x \in (c_{k-1}, c_k])}{f(x) P(X \in (b - s - c_k, b - s - c_{k-1}])} \right) \\ & \times I(b - s > \eta_*) + I(b - s \leq \eta_*). \end{aligned}$$

With this family of change of measures, we are ready to present our main results which are based on appropriate choices of the various tuning parameters.

2.2 Summary of the results

Our first result establishes that one can explicitly choose η_* , c_j 's, a_* , a_{**} , p_* , p_{**} and the p_j 's in order to have a strongly efficient (in the terminology of rare-event simulation, see [4]) estimator.

Theorem 1 *Under either Assumptions A or B1-3, there exists an explicit selection of η_* , the c_j 's, a_* , a_{**} , p_* , p_{**} and the p_j 's so that the estimator Z_b (defined as in (8)) is strongly efficient in the sense of being unbiased and having a bounded coefficient of variation. In particular, one can compute $K \in (0, \infty)$ (uniform in $b > 0$) such that*

$$\frac{E^Q Z_b^2}{(EZ_b)^2} = \frac{EL_b I(\tau_b < \infty)}{u(b)^2} < K$$

for $b > 0$.

The proof of this result is given at the end of Section 4. The explicit parameter selection is discussed in items I) to IV) stated in Section 4. A consequence of this result is that, by Chebyshev's inequality, at most $n = O(\varepsilon^{-2}\delta^{-1})$ i.i.d. replications of Z_b are enough in order to estimate $u(b)$ with ε -relative precision and with probability at least $1 - \delta$ uniformly in b . Because the estimator Z_b is based on importance sampling, one can estimate a large class of expectations of the form $u_H(b) = E(H(S_n : n \leq \tau_b) | \tau_b < \infty)$ with roughly the same number of replications in order to achieve ε -relative precision with at least $1 - \delta$ probability (uniformly in b). Indeed, if $K_1 \in (0, \infty)$ is such that $K_1^{-1} \leq H \leq K_1$ then we have that $u_H(b) \geq K_1^{-1}u(b)$. We also have that $L_b I(\tau_b < \infty) H(S_n : n \leq \tau_b)$ is an unbiased estimator for $E(H(S_n : n \leq \tau_b); \tau_b < \infty)$ and its second moment is bounded by $K_1^2 u(b)^2$. Therefore, we can estimate both the numerator and the denominator in the expression

$$u_H(b) = E(H(S_n : n \leq \tau_b) | \tau_b < \infty) = \frac{E(H(S_n : n \leq \tau_b); \tau_b < \infty)}{u(b)}$$

with good relative precision (uniformly in b). Naturally, the condition $K_1^{-1} \leq H \leq K_1$ is just given to quickly explain the significance of the previous observation. More generally, one might expect strong efficiency for $u_H(b)$ using an importance sampling estimator designed to estimate $u(b)$ if $u_H(b) \in (K_1^{-1}u(b), K_1u(b))$ uniformly in b .

Given that nothing has been said about the cost of generating a single replication of Z_b , strong efficiency is clearly not a concept that allows to accurately assess the total computational cost of estimating $u(b)$ or $u_H(b)$. For this reason, we will also provide results that estimate the expected cost required to generate a single replication of Z_b . However, before we state our estimates for the cost per replication, it is worth discussing what is the performance of the zero-variance change of measure for the regularly varying case. The following classical result ([5]) provides a good description of $(S_n : n \geq 0)$ given $\tau_b < \infty$.

Theorem 2 (Asmussen and Kluppelberg) *Suppose that X is regularly varying with index $\iota > 1$ and define $a(b) = \int_b^\infty P(X > u) du / P(X > b)$. Then, conditional on $\tau_b < \infty$ we have that*

$$\left(\frac{\tau_b}{a(b)}, \left(\frac{S_{\lfloor u\tau_b \rfloor}}{\tau_b} : 0 \leq u < 1 \right), \frac{S_{\tau_b} - b}{b} \right) \implies (Y_0 / |\mu|, (u\mu : 0 \leq u < 1), Y_1),$$

where the convergence occurs in the space $R \times D[0, 1) \times R$, $P(Y_i > t) = (1 + t/(\iota - 1))^{-\iota+1}$ for $t \geq 0$ and $i = 0, 1$ and $P(Y_0 > y_0, Y_1 > y_1) = P(Y_0 > y_0 + y_1)$.

Remark 2 *The previous result suggests that if Assumption A holds, the best possible performance that one might realistically expect is $E^Q \tau_b = O(b)$ as long as (very important!) $\iota > 2$. The full statement of Asmussen and Kluppelberg's result (Theorem 1.1 in [5]) also covers other subexponential distributions. For instance, in the case of Weibull-type tails with shape parameter β_0 , their result suggests that $E(\tau_b | \tau_b < \infty) = O(b^{1-\beta_0})$.*

As the next theorem states, for the regularly varying case with $\iota > 1.5$, we can guarantee $E^Q \tau_b = O(b)$ while maintaining strong efficiency as stated in Theorem 1. We will also indicate why we believe that this result is basically the best possible that can be obtained among a reasonable class of importance sampling distributions.

Theorem 3

- *If Assumptions A holds and $\iota > 1.5$, then there exists an explicit selection of η_* , the c_j 's, a_* , a_{**} , p_* , p_{**} such that strong efficiency (as indicated in Theorem 1) holds and*

$$E^Q \tau_b \leq \rho_0 + \rho_1 b$$

for some $\rho_0, \rho_1 > 0$ independent of b .

- *If Assumption B1-3 holds, we assume there exists $\delta > 0$ and $\beta \in [0, \beta_0]$ such that $\lambda(x) \geq \delta x^{\beta-1}$ for x sufficiently large. Then, with the parameters selected in Theorem 1, there exists ρ_0 and ρ_1 independent of b , such that,*

$$E^Q \tau_b \leq \rho_0 + \rho_1 b^{1-\beta}.$$

Remark 3 *The results in this theorem follow directly as a consequence of Propositions 6 and 7 in Section 5. For the regularly varying case (Assumption A), in addition to the explicit parameter selection indicated in items I) to IV) in Section 4, which guarantee strong efficiency, we also add item V) in Section 5, which explicitly indicates how to select the parameters to obtain $O(b)$ expected stopping time while maintaining strong efficiency. We assume that it takes at most a fixed cost c of computer time units to generate a variable from $q_s(\cdot)$ (uniformly in s). The previous result implies that if X is regularly varying with index $\iota > 1.5$, then our importance sampling family estimates $u(b)$ and associated conditional expectations such as $u_H(b)$ in $O(\varepsilon^{-2} \delta^{-1} b)$ units of computer time. This is in some sense (given that we have linear complexity in b even if $\iota \in (1.5, 2)$) better than what one might expect in view of Theorem 2. We will further provide an argument, see Remark 6 in Section 5, for why in the presence of regular variation $\iota > 3/2$ appears to be basically a necessary condition to obtain strongly efficient unbiased estimators with $O(b)$ expected termination time.*

Remark 4 *For the second case in Theorem 3, note that when Assumption B1 holds, one can always choose $\beta = 0$ and δ arbitrarily large. This selection actually results in $o(b)$ estimate for the expected termination time. It is desirable to choose β as large as possible because this yields a smaller (asymptotically) the termination time. However, there is an upper bound, namely β_0 , which can be derived from Assumption B2 (Lemma 1).*

For the regularly varying case, we provide further results for all $\iota > 1$. If $\iota > 1$, we are able to construct an importance sampling estimator Z_b such that for some $\gamma > 0$ we can guarantee $E^Q(Z_b^{1+\gamma}) \leq Ku(b)^{1+\gamma}$ and at the same time $E^Q\tau_b = O(b)$. The next result, whose proof is given at the end of Section 5, allows us to conclude that this can be achieved with our method as well.

Theorem 4 *Suppose that Assumption A is in force and $\iota \in (1, 1.5]$. Then, for each $\gamma \in (0, (\iota - 1)/(2 - \iota))$ we can select $K > 0$, and a member of our family of importance sampling distributions such that*

$$E^Q(Z_b^{1+\gamma}) \leq Ku(b)^{1+\gamma}$$

for all $b > 0$ and $E^Q(\tau_b) \leq \rho_0 + \rho_1 b$ for $\rho_0, \rho_1 \in (0, \infty)$. Consequently, assuming that each increment under $q_s(\cdot)$ takes at most constant units of computer time, then $O(\varepsilon^{-2/\gamma} \delta^{-1/\gamma} b)$ expected total cost is required to obtain an estimate for $u(b)$ with ε -relative error and with probability at least $1 - \delta$.

Remark 5 *Similar to the case of controlling the second moment, we believe that the upper bound $(\iota - 1)/(2 - \iota)$ is optimal within a reasonable class of simulation algorithms. A heuristic argument will be given in Section 5.*

Finally, the proposed family of change of measures and analysis techniques are useful not only for Monte Carlo simulation purposes but also for asymptotic analysis. We provide the following approximation results which improve upon classical results in the literature such as Theorem 2. By appropriately tuning various parameters in our family we can approximate $P(S \in \cdot | \tau_b < \infty)$ by $Q(S \in \cdot)$ asymptotically as $b \nearrow \infty$. We will explicitly indicate how to do so in later analysis.

Theorem 5 *Under either Assumptions A or B1-3, there exists an explicit selection of η_* , the c_j 's, a_* , a_{**} , p_* , p_{**} and the p_j 's so that*

$$\overline{\lim}_{b \rightarrow \infty} \sup_A |P(S \in A | \tau_b < \infty) - Q(S \in A)| = 0.$$

The previous result is an immediate consequence of Lemma 10 combined with Theorem 8. It further shows that our mixture family is an appropriate vehicle to approximate the conditional distribution of the random walk given $\tau_b < \infty$. Moreover, due to the convenience of the mixture form, as a corollary of the previous theorem and using a coupling technique, we can show, without much additional effort, the following theorem which further extends Theorem 1.1 in [5] by adding a central limit theorem correction term. This theorem is proven at the end of Section 6.2.

Theorem 6 *Suppose that either Assumption A or Assumptions B1-4 are in force. Let $\sigma^2 = \text{Var}(X_1) \leq \infty$ and $a(b) = \int_b^\infty P(X > u) du / P(X > b)$. Then*

$$\left(\frac{\tau_b}{a(b)}, \left\{ \frac{S_{[t\tau_b]} - t\mu\tau_b}{\sqrt{\tau_b}} \right\}_{0 \leq t < 1}, \frac{S_{\tau_b} - b}{a(b)} \right) \Rightarrow (Y_0 / |\mu|, \{\sigma B(t)\}_{0 \leq t < 1}, Y_1),$$

in $R \times D[0, 1) \times R$. $\{B(t) : 0 \leq t < 1\}$ is a standard Brownian motion independent of (Y_0, Y_1) . The joint law of Y_0 and Y_1 is defined as follows. First, $P(Y_0 > y_0, Y_1 > y_1) = P(Y_1 > y_0 + y_1)$ with $Y_0 \stackrel{d}{=} Y_1$ and

- If Assumption A holds then

$$P(Y_1 > t) = \frac{1}{(1 + t/(t-1))^{t-1}}.$$

- If Assumptions B1-4 hold, then Y_1 follows exponential distribution with mean 1 and consequently Y_0 and Y_1 are independent.

3 Preliminaries: Heavy tails, importance sampling and Lyapunov inequalities

3.1 Heavy tails

A non-negative random variable Y is said to be heavy-tailed if $E \exp(\theta Y) = \infty$ for every $\theta > 0$. This class is too big to develop a satisfactory asymptotic theory of large deviations and therefore one often considers the subexponential distributions which are defined as follows.

Definition 1 Let Y_1, \dots, Y_n be independent copies of a non-negative random variable Y . The distribution of Y (or Y itself) is said to be subexponential if and only if

$$\lim_{u \rightarrow \infty} \frac{P(Y_1 + \dots + Y_n > u)}{P(Y > u)} = n.$$

Actually it is necessary and sufficient to verify the previous limit for $n = 2$ only.

Examples of distributions that satisfy the subexponential property include Pareto distribution, Lognormal distributions, Weibull distributions, and so forth. A general random variable X is said to have a subexponential right tail if X^+ is subexponential. In such a case, we simply say that X is subexponential.

If X is subexponential, then X satisfies that $P(X > x + h) / P(X > x) \rightarrow 1$ as $x \rightarrow \infty$ for each $h \in (-\infty, \infty)$. A random variable with this property is said to possess a "long tail". It turns out that there are long tailed random variables that do not satisfy the subexponential property (see [20]).

In order to verify the subexponential property in the context of random variables with a density function (as we shall assume here) one often takes advantage of the so-called cumulative hazard function. Indeed, a sufficient condition to guarantee subexponentiality due to Pitman is given next (see [20]).

Proposition 1 A random variable X with concave cumulative hazard function $\Lambda(\cdot)$ and hazard function $\lambda(\cdot)$ is subexponential if

$$\int_0^\infty \exp(x\lambda(x) - \Lambda(x)) dx < \infty.$$

A distinctive feature of heavy-tailed random walks is that the rare event $\{\sup_n S_n > b\}$ is asymptotically (as $b \rightarrow \infty$) caused by a single large increment, while other increments behave like “regular” ones. Therefore, one can obtain the following approximation, often called fluid heuristic, for the probability $u(b)$:

$$\begin{aligned} u(b) &= P(\tau_b < \infty) = \sum_{k=1}^{\infty} P(\tau_b = k) \\ &\approx \sum_{k=1}^{\infty} P(X_k > b - (k-1)\mu) \approx -\frac{1}{\mu} \int_b^{\infty} P(X > s) ds. \end{aligned} \quad (15)$$

For notational convenience, we denote the integrated tail by

$$G(x) = \int_x^{\infty} P(X > s) ds. \quad (16)$$

The previous heuristic can actually be made rigorous under subexponential assumptions. This is the content of the Pakes-Veraverbeke theorem which we state next (see page 296 in [2]).

Theorem 7 (Pakes-Veraverbeke) *If F is long tailed (i.e. $\bar{F}(x+h)/\bar{F}(x) \rightarrow 1$ as $x \nearrow \infty$ for every $h > 0$) and $\int_0^t P(X > s) ds/EX^+$ is subexponential (as a function of t) then*

$$u(b) = -(\mu^{-1} + o(1))G(b), \quad (17)$$

as $b \rightarrow \infty$.

We close this subsection with a series of lemmas involving several properties which will be useful throughout the paper. The proofs of these results are given in Appendix A.

Lemma 1 *If B2 holds then $\lambda(x) = O(x^{\beta_0-1}) \rightarrow 0$ as $x \rightarrow \infty$.*

Lemma 2 *Under Assumption B3 there exists a constant κ_1 (depending on a_*) and b_0 , such that for all $x \leq b - \Lambda^{-1}(\Lambda(b) - a_*)$ and $b > b_0$, the integrated tail satisfies*

$$G(b-x)/G(b) \leq \kappa_1.$$

Lemma 3 *Suppose B1 and B3 are in force. For each $\varepsilon_0 > 0$, there exists $b_0 > 0$ such that*

$$\varepsilon_0^{-1} \bar{F}(b) \leq G(b) \leq \varepsilon_0 b \bar{F}(b),$$

for all $b \geq b_0$. In particular, $\bar{F}(b)/G(b) = o(1)$ as $b \rightarrow \infty$. If Assumption A holds then for each $\delta_0 > 0$ we can select $b_0 > 0$ sufficiently large so that

$$G(b) \leq (\iota - 1)(1 + \delta_0)b\bar{F}(b).$$

for $b \geq b_0$, where ι is the tail index of \bar{F} defined in Assumption A.

Lemma 4 Suppose B2 holds, for all $x \geq b_0$ and $y \geq 0$ we have

$$\frac{\Lambda(x)}{\Lambda(x+y)} \geq \left(\frac{x}{x+y} \right)^{\beta_0}.$$

Lemma 5 Suppose B2 is satisfied. Then, we can choose $b_0 > 0$ sufficiently large such that

$$x - \Lambda^{-1}(\Lambda(x) - a_*) \geq x^{(1-\beta_0)/2},$$

for all $x > b_0$.

The following lemma allows us to conclude that the Pakes-Veraberbeke theorem is applicable in our setting.

Lemma 6 Under either Assumption A or B1-3, both $F(x)$ and $\int_0^x P(X > s) ds / (EX^+)$ are subexponential as a function of x .

3.2 State-dependent importance sampling for the first passage time random walk problem and Lyapunov inequalities

Consider two probability measures P and Q on a given space \mathcal{X} with σ -algebra \mathcal{F} . If the Radon-Nikodym derivative $\frac{dP}{dQ}(\omega)$ is well defined on the set $A \in \mathcal{F}$, then

$$P(A) = \int \frac{dP}{dQ}(\omega) I_A(\omega) Q(d\omega).$$

We say that the random variable $\frac{dP}{dQ}(\omega) I_A(\omega)$ is the importance sampling estimator associated to the change of measure / importance sampling distribution Q . If one chooses Q' such that for each $B \in \mathcal{F}$,

$$Q'(B) = P(B \cap A) / P(A),$$

then, $\frac{dP}{dQ'} \equiv P(A)$ almost surely on the set A and therefore the estimator $\frac{dP}{dQ'}(\omega)$ has zero variance. This implies that the best importance sampling distribution (with zero variance for estimating $P(A)$) is the conditional distribution given the event A occurs.

Certainly, this zero variance estimator is not implementable in practice, because the Radon-Nikodym derivative involves precisely computing $P(A)$, which is the quantity to compute. Nevertheless, it provides a general guideline on how to construct efficient importance sampling estimators: try to mimic the conditional distribution given the event of interest.

In the context of this paper, we consider a random walk $(S_n : n \geq 0)$ with $S_0 = 0$ and therefore

$$P(X_{n+1} \in dx | S_1, \dots, S_n) = F(dx).$$

A state-dependent importance sampling distribution Q is such that

$$Q(X_{n+1} \in dx | S_1, \dots, S_n) = r_{S_n}^{-1}(x) F(dx), \tag{18}$$

where, the function $(r_s(x) : s, x \in R)$ is non-negative and it satisfies

$$\int_{-\infty}^{\infty} r_{S_n}^{-1}(x) F(dx) = 1.$$

Now, consider the stopping time $\tau_b = \inf\{n \geq 0 : S_n > b\}$ and set $A_b = \{\tau_b < \infty\}$, then it follows easily that

$$P(A_b) = E^Q \left\{ I_{A_b} \prod_{i=1}^{\tau_b} r_{S_{i-1}}(S_i - S_{i-1}) \right\}.$$

Notational convention: throughout the paper we shall use $E_s^Q(\cdot)$ to denote the expectation operator induced by (18) assuming that $S_0 = s$. We simply write $E^Q(\cdot)$ whenever $S_0 = 0$.

We will work with the specific parametric selection of $r_s(x)$ introduced in Section 2. In proving some of our main results we will be interested in finding an upper bound for the second moment of our estimator under $E^Q(\cdot)$, namely

$$E^Q \left\{ I_A \prod_{i=1}^{\tau_b} r_{S_{i-1}}^2(S_i - S_{i-1}) \right\} = E \left\{ I_A \prod_{i=1}^{\tau_b} r_{S_{i-1}}(S_i - S_{i-1}) \right\}.$$

In general, the $(1 + \gamma)$ -th moment ($\gamma > 0$) of our estimator satisfies

$$E \left\{ I_A \prod_{i=1}^{\tau_b} r_{S_{i-1}}(S_i - S_{i-1})^\gamma \right\}$$

The next lemma provides the mechanism that we shall use to obtain upper bounds for these quantities. The proof can be found in [9].

Lemma 7 *Assume that there exists a non-negative function $g : \mathbb{R} \rightarrow \mathbb{R}^+$, such that for all $s < b$,*

$$g(s) \geq E(g(s + X)r_s(X)^\gamma),$$

where X is a random variable with density $f(\cdot)$ and suppose that for all $s \geq b$, $g(s) \geq \varepsilon$. Then,

$$g(0) \geq \varepsilon E \left\{ I_A \prod_{i=1}^{\tau_b} r_{S_{i-1}}(S_i - S_{i-1})^\gamma \right\}. \quad (19)$$

Most of the time we will work with $\gamma = 1$ (i.e. we concentrate on the second moment). The inequality (19) is said to be a Lyapunov inequality. The function g is called a Lyapunov function. Lemma 7 provides a handy tool to derive an upper bound of the second moment of the importance sampling estimator. However, the lemma does not provide a recipe on how to construct a suitable Lyapunov function. We will discuss the intuition behind the construction of our Lyapunov function in future sections.

If $r_s(x)$ has been chosen in such a way that the second moment of the importance sampling estimator can be suitably controlled by an appropriate selection of a Lyapunov function g , we still need to make sure that the cost per replication (i.e. $E^Q \tau_b$) is suitably controlled as well. The next lemma, which follows exactly the same steps as in the first part of the proof in Theorem 11.3.4 of [22], establishes a Lyapunov criterion required to control the behavior of $E^Q \tau_b$.

Lemma 8 *Suppose that one can find a non-negative function $h(\cdot)$ and a constant $\rho > 0$ so that*

$$E_s^Q(h(s + X)) \leq h(s) - \rho,$$

for $s < b$. Then, $E^Q(\tau_b | S_0 = s) \leq h(s)/\rho$ for $s < b$.

Most of the results discussed in Section 2 of the paper involve constructing suitable selections of Lyapunov functions g and h appearing in the previous lemmas. The construction of these functions is given in subsequent sections.

4 Lyapunov function for variance control

Our approach to designing efficient importance sampling estimators consist of three steps:

1. Propose a family of change of measures suitably parameterized.
2. Propose candidates of Lyapunov functions using fluid heuristics and also depending on appropriate parameters.
3. Verify the Lyapunov inequality by choosing appropriate parameters for the change of measure and the Lyapunov function.

Our family has been introduced in Section 2. This corresponds to the first step. The second and third steps are done simultaneously. We will choose the parameters η_* , the c_j 's, a_* , a_{**} , p_* , p_{**} and the p_j 's of our change of measure in order to satisfy an appropriate Lyapunov function for variance control by means of Lemma 7. Some of the parameters, in particular the c_j 's, can be set in advance without resorting to the appropriate Lyapunov function. The key element is given in the next lemma, whose proof is given in the appendix.

Lemma 9 *For each $\beta_0 \in (0, 1)$ and $\sigma_1 > 0$ (sufficiently small), there exists $\sigma_2 > 0$ and a sequence, $0 < a_1 < a_2 < \dots < a_{k-1} < 1$ such that $a_{j+1} - a_j \leq \sigma_1/2$ for each $1 \leq j \leq k - 2$,*

$$a_j^{\beta_0} + (1 - a_{j+1})^{\beta_0} \geq 1 + \sigma_2.$$

and $a_{k-1} \geq 1 - \sigma_1$, $a_1 \leq \sigma_1$.

Given β_0 in Assumption B2, from now on, we choose

$$c_0 = b - s - \Lambda^{-1}(\Lambda(b - s) - a_*), \quad c_k = \Lambda^{-1}(\Lambda(b - s) - a_{**}), \quad c_j = a_j(b - s), \quad (20)$$

for $j = 1, \dots, k - 1$, with σ_1 chosen small enough and $a_j = a_{j-1} + \sigma_1/2$ according to the previous lemma.

We continue with the second step of our program. We concentrate on bounding the second moment and discuss the case of $(1 + \gamma)$ -th moment later. The value of the Lyapunov function at the origin, namely, $g(0)$ in Lemma 7 serves as the upper bound of the second

moment of the importance sampling estimator. In order to prove strong efficiency, we aim to show that there exists a constant $c < \infty$ such that

$$E^Q Z_b^2 \leq cu^2(b),$$

where

$$Z_b = I(\tau_b < \infty) \prod_{i=1}^{\tau_b} r_{S_{i-1}}(S_i - S_{i-1}) \quad (21)$$

is the estimator of $u(b)$. Therefore, a useful Lyapunov function for proving strong efficiency must satisfy that

$$g(0) \leq cu^2(b).$$

It is natural to consider using an approximation of $u^2(b-s)$ as the candidate. Exactly the same type of fluid heuristic analysis that we used in (15) suggests

$$g(s) = \min\{\kappa G^2(b-s), 1\}, \quad (22)$$

where G is the integrated tail defined in (16) and κ is a non-negative tuning parameter which will be determined later.

It is important to keep in mind that $g(s)$ certainly depends on b . For notational simplicity, we omit the parameter b . The function $g(s)$ will also dictate when we are close enough to the boundary level b where importance sampling is not required. In particular, using our notation in (10) and (18) we propose choosing $\eta_* = G^{-1}(\kappa^{-1/2})$ which amounts to choosing

$$\begin{aligned} & r_s(x)^{-1} \\ &= \left(\frac{p_* I(x \leq c_0)}{P(X \leq c_0)} + \frac{p_{**} I(x > c_k)}{P(X > c_k)} + \sum_{j=1}^{k-1} \frac{p_j I(x \in (c_{j-1}, c_j])}{P(X \in (c_{j-1}, c_j])} + \frac{f(b-s-x)p_k I(x \in (c_{k-1}, c_k])}{f(x)P(X \in (b-s-c_k, b-s-c_{k-1}])} \right) \\ & \times I(g(s) < 1) + I(g(s) = 1). \end{aligned}$$

Now we proceed to the last step – the verification of the Lyapunov inequality. The Lyapunov inequality in Lemma 7 is equivalent to

$$\frac{E(r_s(X)g(s+X))}{g(s)} \leq 1. \quad (23)$$

The interesting part of the analysis is the case $g(s) < 1$ because whenever $g(s) = 1$ the inequality is trivially satisfied given that $0 \leq g(s+X) \leq 1$. Hereafter, we will focus on the case that $g(s) < 1$.

The left hand side of (23) can be decomposed into the following pieces,

$$\begin{aligned}
\frac{E(r_s(X)g(s+X))}{g(s)} &= \frac{P(X \leq b-s-\Lambda^{-1}(\Lambda(b-s)-a_*))}{p_*} \\
&\times E\left(\frac{g(s+X)}{g(s)}; X \leq b-s-\Lambda^{-1}(\Lambda(b-s)-a_*)\right) \\
&+ \frac{P(X > \Lambda^{-1}(\Lambda(b-s)-a_{**}))}{p_{**}} E\left(\frac{g(s+X)}{g(s)}; X > \Lambda^{-1}(\Lambda(b-s)-a_{**})\right) \\
&+ \sum_{i=1}^{k-1} \frac{P(X \in (c_{i-1}, c_i])}{p_i} E\left(\frac{g(s+X)}{g(s)}; X \in (c_{i-1}, c_i]\right) \\
&+ \frac{P(b-s-X \in (c_{k-1}, c_k])}{p_k} E\left(\frac{g(s+X)f(X)}{g(s)f(b-s-X)}; X \in (c_{k-1}, c_k]\right).
\end{aligned}$$

We adopt the following notation

$$J_* = P(X \leq b-s-\Lambda^{-1}(\Lambda(b-s)-a_*))E\left(\frac{g(s+X)}{g(s)}; X \leq b-s-\Lambda^{-1}(\Lambda(b-s)-a_*)\right) \quad (24)$$

$$J_{**} = P(X > \Lambda^{-1}(\Lambda(b-s)-a_{**}))E\left(\frac{g(s+X)}{g(s)}; X > \Lambda^{-1}(\Lambda(b-s)-a_{**})\right) \quad (25)$$

$$J_i = P(X \in (c_{i-1}, c_i])E\left(\frac{g(s+X)}{g(s)}; X \in (c_{i-1}, c_i]\right), \text{ for } i = 1, \dots, k-1 \quad (26)$$

$$J_k = P(b-s-X \in (c_{k-1}, c_k])E\left(\frac{g(s+X)f(X)}{g(s)f(b-s-X)}; X \in (c_{k-1}, c_k]\right), \quad (27)$$

so that inequality (23) is equivalent to showing that

$$\frac{J_*}{p_*} + \frac{J_{**}}{p_{**}} + \sum_{i=1}^{k-1} \frac{J_i}{p_i} + \frac{J_k}{p_k} \leq 1.$$

We shall study each of these terms separately.

At this point it is useful to provide a summary of all the relevant constants and parameters introduced so far:

- $\iota > 1$ is the regularly varying index under Assumption A.
- $b_0 > 0$ is introduced in Assumption B, Lemmas 3 and 5 to ensure regularity properties.
- $\beta_0 \in (0, 1)$ is introduced in B2 to guarantee that the distribution considered is “heavier” than a Weibull distribution with shape parameter β_0
- $a_*, a_{**} > 0$ are introduced to define the mixture components corresponding to a “regular jump” and a “large jump” respectively.

- $a_1 < \dots < a_{k-1}$ are defined according to Lemma 9.
- c_j for $j = 0, 1, \dots, k$ are defined in (20) and correspond to the end points of the support of the interpolating mixture components.
- κ, η_* are parameters for the Lyapunov function. They are basically equivalent since $\eta_* = G^{-1}(\kappa^{-1/2})$, κ appears in the definition of the Lyapunov function. It is important to keep in mind that by letting κ be large, the condition $g(s) < 1$ implies that $b-s > \eta_*$ is large.
- ε_0, δ_0 are arbitrarily small constants introduced in Lemma 3.
- The parameters p_*, p_{**} and p_i for $i = 1, \dots, k$ are the mixture probabilities and will depend on the current state s .

Other critical constants which will be introduced in the sequel concerning the analysis of J_*, J_{**} , and $J_i, i = 1, \dots, k$ are:

- $\delta_0^* > 0$ is a small parameter which appears in the analysis of J_* . It will be introduced in Proposition 2.
- $\delta_1^* > 0$, a small parameter, appears in the definition of p_i and the overall contribution of the J_i 's. It will be introduced in step III) of the parameter selection process.
- $\delta_2^* > 0$ is introduced to control the termination time of the algorithm. It ultimately provides a link between $a_{**} > 0$ and $\delta_0^* > 0$ in Section 5.
- Parameters $\theta, \tilde{\varepsilon}$ and $\tilde{\varepsilon}_1$ which are introduced to specify the probabilities p_{**} and the p_i 's respectively. Their specific values depending on δ_0^* and δ_1^* will be indicated in steps I) to IV) below.

Throughout the rest of the paper we shall use $\varepsilon, \delta > 0$ to denote arbitrarily small positive constants whose values might even change from line to line. Similarly, $K, c \in (0, \infty)$ are used to denote a positive constants that will be employed as generic upper bounds.

Now, we study the terms J_*, J_{**} , and $J_i, i = 1, \dots, k$.

The term J_{} :**

$$\begin{aligned}
J_{**} &= P(X > \Lambda^{-1}(\Lambda(b-s) - a_{**})) E \left(\frac{g(s+X)}{g(s)}; X > \Lambda^{-1}(\Lambda(b-s) - a_{**}) \right) \\
&\leq \frac{P^2(X > \Lambda^{-1}(\Lambda(b-s) - a_{**}))}{g(s)} = e^{2a_{**}} \frac{\bar{F}^2(b-s)}{g(s)}
\end{aligned} \tag{28}$$

A bound for J_* :

Proposition 2 *Suppose the distribution function F satisfies Assumption A or Assumption B1-3. Then, as $b - s \rightarrow \infty$,*

$$E \left(\frac{g(s+X)}{g(s)}; X \leq b - s - \Lambda^{-1}(\Lambda(b-s) - a_*) \right) \leq 1 + (1 + o(1))\mu \frac{\partial g(s)}{g(s)}.$$

Therefore, for any $\delta_0^* > 0$, we can select $\eta_* > 0$ such that for all $b - s > \eta_*$,

$$E \left(\frac{g(s+X)}{g(s)}; X \leq b - s - \Lambda^{-1}(\Lambda(b-s) - a_*) \right) \leq 1 + \mu(1 - \delta_0^*) \frac{\partial g(s)}{g(s)}.$$

Proof of Proposition 2. By Taylor's expansion,

$$\frac{g(s+X)}{g(s)} = 1 + X \frac{\partial g(s+\xi)}{g(s)},$$

where $\xi \in (0, X)$ (or $(X, 0)$). For all s and X such that $g(s) < 1$ and $g(s+X) < 1$,

$$X \partial g(s+\xi)/g(s) = 2X \bar{F}(b-s-\xi) G(b-s-\xi)/G^2(b-s) = 2X \frac{\bar{F}(b-s-\xi)}{\bar{F}(b-s)} \frac{G(b-s-\xi)}{G(b-s)} \frac{\bar{F}(b-s)}{G(b-s)}.$$

Then,

$$\begin{aligned} & \frac{G(b-s)}{\bar{F}(b-s)} E \left(X \partial g(s+\xi)/g(s); X \leq b - s - \Lambda^{-1}(\Lambda(b-s) - a_*) \right) \\ & \leq 2E \left(X \frac{\bar{F}(b-s-\xi)}{\bar{F}(b-s)} \frac{G(b-s-\xi)}{G(b-s)}; X \leq b - s - \Lambda^{-1}(\Lambda(b-s) - a_*) \right) \end{aligned}$$

Note the following fact,

$$\frac{\bar{F}(b-s-\xi)}{\bar{F}(b-s)} \leq e^{a_*},$$

and by Lemma 2,

$$\frac{G(b-s-\xi)}{G(b-s)} \leq \kappa_1,$$

and by Lemma 3 and the fact that F is subexponential (Lemma 6),

$$X \frac{\bar{F}(b-s-\xi)}{\bar{F}(b-s)} \frac{G(b-s-\xi)}{G(b-s)} \rightarrow X,$$

as $b - s \rightarrow \infty$. By the dominated convergence theorem,

$$\lim_{b-s \rightarrow \infty} \frac{G(b-s)}{\bar{F}(b-s)} E \left(X \partial g(s+\xi)/g(s); X \leq b - s - \Lambda^{-1}(\Lambda(b-s) - a_*) \right) = 2\mu. \quad (29)$$

Therefore, we can always choose the constants appropriately such that the conclusion of the proposition holds. ■

As remarked in equation (11), the terms J_i , $i = 1, \dots, k$, do not appear in the context of Assumption A. We consider them in the context of Assumption B.

Bound for J_i , $2 \leq i \leq k-1$:

Proposition 3 *Suppose that Assumptions B1-3 hold. Then, for each $2 \leq i \leq k-1$, we have that for any $\alpha > 0$*

$$J_i = \int_{c_{i-1}}^{c_i} \frac{f(x)g(s+x)}{f_j(x)g(s)} f(x)dx = o((b-s)^{-\alpha}),$$

as $b-s \rightarrow \infty$.

Proof. Thanks to Lemma 4, for each x, y, z sufficiently large, we have

$$\Lambda(x) + \Lambda(y) - \Lambda(x+y+z) \geq \Lambda(x+y+z) \left(\left(\frac{x}{x+y+z} \right)^{\beta_0} + \left(\frac{y}{x+y+z} \right)^{\beta_0} - 1 \right). \quad (30)$$

We first note that by repeatedly using results in Lemma 3

$$\begin{aligned} & \int_{c_{j-1}}^{c_j} \frac{f(x)g(s+x)}{\kappa f_j(x)G^2(b-s)} f(x)dx \\ &= \frac{P(X \in (c_{j-1}, c_j])}{G^2(b-s)} \int_{c_{j-1}}^{c_j} G^2(b-s-x)f(x)dx \\ &\leq \frac{\varepsilon_0 e^{\Lambda(b-s)-\Lambda(c_{j-1})}}{G(b-s)} \int_{c_{j-1}}^{c_j} G^2(b-s-x)\lambda(x)e^{-\Lambda(x)}dx \\ &\leq \frac{\varepsilon_0 e^{\Lambda(b-s)-\Lambda(c_{j-1})}}{G(b-s)} \bar{F}(c_{j-1})G^2(b-s-c_j) \\ &\leq \varepsilon_0^4 \frac{e^{\Lambda(b-s)-\Lambda(c_{j-1})}}{\bar{F}(b-s)} \bar{F}(c_{j-1})(b-s)^2 \bar{F}^2(b-s-c_j) \\ &= \varepsilon_0^4 (b-s)^2 e^{2\Lambda(b-s)-2\Lambda(c_{j-1})-2\Lambda(b-s-c_j)} \\ &\leq \varepsilon_0^4 (b-s)^2 \exp \left\{ -2\Lambda(b-s) \left(a_{j-1}^{\beta_0} + (1-a_j)^{\beta_0} - 1 \right) \right\} = o(1) (b-s)^{-\alpha}, \end{aligned}$$

as $b-s \rightarrow \infty$ for each $\alpha > 0$. The last inequality is thanks to (20), (30). The last step (equality) follows from Lemma 9 and Assumption B1 which implies that the tail of X decreases faster than any polynomial. ■

A bound for J_1 :

Proposition 4 *Suppose that Assumptions B1-3 hold. Then, for each $\alpha > 0$ we have*

$$J_1 = \int_{b-s-\Lambda^{-1}(\Lambda(b-s)-a_*)}^{c_1} \frac{f(x)g(s+x)}{f_1(x)g(s)} f(x)dx = o((b-s)^{-\alpha}),$$

as $b-s \rightarrow \infty$.

Proof of Proposition 4. Use Lemma 3 and $\lim_{x \rightarrow \infty} \lambda(x) = 0$ and obtain

$$\begin{aligned} & \int_{b-s-\Lambda^{-1}(\Lambda(b-s)-a_*)}^{c_1} \frac{f(x)g(s+x)}{\kappa f_1(x)G^2(b-s)} f(x)dx \\ & \leq \frac{P(X > b-s-\Lambda^{-1}(\Lambda(b-s)-a_*))}{G^2(b-s)} \int_{b-s-\Lambda^{-1}(\Lambda(b-s)-a_*)}^{c_1} G^2(b-s-x)f(x)dx \\ & \leq \varepsilon_0^4(b-s)^2 P(X > b-s-\Lambda^{-1}(\Lambda(b-s)-a_*)) \int_{b-s-\Lambda^{-1}(\Lambda(b-s)-a_*)}^{c_1} e^{2\Lambda(b-s)-2\Lambda(b-s-x)-\Lambda(x)} dx. \end{aligned}$$

Also note that by Lemma 4,

$$\Lambda(x) + \Lambda(b-s-x) - \Lambda(b-s) \geq \Lambda(b-s) \left(\left(\frac{x}{b-s} \right)^{\beta_0} + \left(\frac{b-s-x}{b-s} \right)^{\beta_0} - 1 \right), \quad (31)$$

and,

$$\Lambda(b-s) - \Lambda(b-s-x) \leq \Lambda(b-s)(1 - (1 - x/(b-s))^{\beta_0}).$$

Therefore, for all $x \in [b-s-\Lambda^{-1}(\Lambda(b-s)-a_*), \sigma_1(b-s)]$, with σ_1 small enough,

$$2\Lambda(b-s) - 2\Lambda(b-s-x) - \Lambda(x) \leq \Lambda(b-s) \left(2 - 2 \left(1 - \frac{x}{b-s} \right)^{\beta_0} - \frac{x^{\beta_0}}{(b-s)^{\beta_0}} \right) \leq 0.$$

Together with Lemma 5, $P(X > b-s-\Lambda^{-1}(\Lambda(b-s)-a_*))$ decreases to zero faster than any polynomial rate. The conclusion of the lemma follows. ■

A bound for J_k :

Proposition 5 *If Assumption B holds then for each $\alpha > 0$*

$$J_k = \int_{c_{k-1}}^{c_k} \frac{f(x)g(s+x)}{f_k(x)g(s)} f(x)dx = o((b-s)^{-\alpha}),$$

as $b-s \rightarrow \infty$.

Proof of Proposition 5. Note that

$$\begin{aligned} & \int_{c_{k-1}}^{c_k} \frac{g(s+x)}{\kappa G^2(b-s)} \frac{f^2(x)}{f_k(x)} dx \\ & = P(X \in (b-s-c_k, b-s-c_{k-1}]) \int_{c_{k-1}}^{c_k} \frac{g(s+x)}{\kappa G^2(b-s)} \frac{f^2(x)}{f(b-s-x)} dx \\ & \leq \varepsilon_0^4 \bar{F}(b-s-c_k) \int_{c_{k-1}}^{c_k} (b-s)^2 \lambda^2(x) \lambda^{-1}(b-s-x) e^{2\Lambda(b-s)-2\Lambda(x)-\Lambda(b-s-x)} dx. \end{aligned}$$

We note that σ_1 is small enough and $x > (1-\sigma_1)(b-s)$ so that we can apply Lemma 9 to conclude

$$2\Lambda(b-s) - 2\Lambda(x) - \Lambda(b-s-x) \leq 2 - 2 \left(\frac{x}{b-s} \right)^{\beta_0} - \left(\frac{b-s-x}{b-s} \right)^{\beta_0} \leq 0.$$

By Assumption B1, $1/\lambda(x)$ grows at most linearly in x and also we have that $\bar{F}(b-s-c_k)$ decays faster than any polynomial rate. We then have the conclusion of the proposition. ■

Summary of estimates and implications for the design of the change of measure selection. The previous bounds on J_* , J_{**} , and J_i , $i = 1, \dots, k$ imply that we can choose parameters and setup the algorithm as follows.

I If Assumption A holds, we choose a_* and a_{**} such that (14) holds. If Assumption B holds, given $a_*, a_{**} > 0$, $\sigma_1 > 0$ (sufficiently small), and $a_j = a_{j-1} + \sigma_1/2$, chosen according to Lemma 9, let

$$c_0 = b - s - \Lambda^{-1}(\Lambda(b - s) - a_*), \quad c_k = \Lambda^{-1}(\Lambda(b - s) - a_{**}),$$

$$c_j = a_j(b - s) \text{ for } j = 1, \dots, k - 1.$$

II Select $\delta_0^* \in (0, 1/4)$ and let $\eta_* > 0$ be large enough so that if $b - s > \eta_*$ then

$$\frac{J_*}{p_*} \leq \frac{1}{p_*} + \frac{(1 - \delta_0^*)}{p_*} \mu \frac{\partial g(s)}{g(s)}. \quad (32)$$

III Choose $\delta_1^* \in (0, \delta_0^* \mu^2 (1 - \delta_0^*)^2 (1 + \delta_0^*)^{-10} / (k + 1)^2)$ such that if $b - s > \eta_*$ for η_* large enough

$$J_i \leq \delta_1^* \delta_0^* \left(\frac{\partial g(s)}{g(s)} \right)^2$$

for all $i = 1, \dots, k$. Note that the J_i terms are all zero for the regularly varying case.

The choice in III) is feasible because $\partial g(s)/g(s) = 2\bar{F}(b - s)/G(b - s)$ decreases at most a polynomial rate and J_i terms derived in Propositions 3, 4, and 5 are smaller than any polynomial rate. Both II) and III) can be satisfied simultaneously by choosing η_* sufficiently large. Now, with the selections in II) and III) we have that

$$\begin{aligned} & \frac{J_*}{p_*} + \frac{J_{**}}{p_{**}} + \sum_{i=1}^k \frac{J_k}{p_k} \\ & \leq \frac{1}{p_*} + \frac{(1 - \delta_0^*)}{p_*} \mu \frac{\partial g(s)}{g(s)} + e^{2a_{**}} \frac{\bar{F}^2(b - s)}{p_{**} g(s)} + \delta_1^* \delta_0^* \left(\frac{\partial g(s)}{g(s)} \right)^2 \sum_{i=1}^k \frac{1}{p_i}. \end{aligned} \quad (33)$$

Now we must select p_* , p_{**} and the p_i 's so that (33) is less than unity in order to satisfy (23). Recall that p_{**} represents the mixture probability associated to the occurrence of the rare event in the next step. Therefore, it makes sense to select p_{**} of order $\Theta(\bar{F}(b - s)/G(b - s))$ as $b - s \rightarrow \infty$. Motivated by this observation and given the analytical form of the equation above we write

$$p_{**} = \min\{\theta \partial g(s)/g(s), \tilde{\varepsilon}\} = \min\{2\theta \bar{F}(b - s)/G(b - s), \tilde{\varepsilon}\} \quad (34)$$

for some $\theta, \tilde{\varepsilon} > 0$ (the precise values of θ and $\tilde{\varepsilon}$ will be given momentarily) and let

$$p_i = \tilde{\varepsilon}_1 p_{**} \quad (35)$$

for each $i = 1, \dots, k$ for some $\tilde{\varepsilon}_1 > 0$ small enough to be defined shortly. This selection of p_i 's also makes intuitive sense because the corresponding mixture terms will give rise to

increments that are large, yet not large enough to reach the level b of the random walk and therefore they correspond to “rogue paths” – as we called them in the Introduction. In addition, one can always choose η_* large enough such that $p_{**} < \tilde{\varepsilon}$ for all $b - s > \eta_*$. Given these selections we obtain

$$p_* = 1 - p_{**} - k\tilde{\varepsilon}_1 p_{**}. \quad (36)$$

We then conclude that if $p_{**}(1 + k\tilde{\varepsilon}_1) < \delta_0^*/2 < 1/4$ and $\tilde{\varepsilon} < \delta_0^*/2$, then

$$\begin{aligned} & \frac{J_*}{p_*} + \frac{J_{**}}{p_{**}} + \sum_{i=1}^k \frac{J_k}{p_k} \\ & \leq 1 + p_{**} (1 + k\tilde{\varepsilon}_1) (1 - \delta_0^*)^{-1} + \frac{(1 - \delta_0^*)}{\theta} \mu p_{**} + e^{2a_{**}} \frac{p_{**}}{4\theta^2 \kappa} + k \delta_1^* \delta_0^* \frac{p_{**}}{\theta^2 \tilde{\varepsilon}_1} \\ & = 1 + p_{**} \left[(1 + k\tilde{\varepsilon}_1) (1 - \delta_0^*)^{-1} + \frac{(1 - \delta_0^*)}{\theta} \mu + \frac{e^{2a_{**}}}{4\theta^2 \kappa} + k \frac{\delta_1^* \delta_0^*}{\theta^2 \tilde{\varepsilon}_1} \right]. \end{aligned}$$

Now choose $\tilde{\varepsilon}_1 = \delta_0^*/(k+1)$ and then select $\theta = -\mu(1 - \delta_0^*)/(1 + \delta_0^*)^5$. Then we note that our selection of δ_1^* guarantees $\delta_1^* \leq \theta^2 \tilde{\varepsilon}_1 k^{-1}$. Finally it is required that $\kappa \geq e^{2a_{**}}/[4\theta^2 \delta_0^*]$. Note that the selection of $\delta_0^*, \delta_1^* > 0$ requires that $b - s > \eta_*$ for $\eta_* > 0$ sufficiently large, which is guaranteed whenever $g(s) < 1$ and κ is sufficiently large. So, the selection of κ might possibly need to be increased in order to satisfy all the constraints. All this selections in place yield (using the fact that $\delta_0^* < 1/4$)

$$\frac{J_*}{p_*} + \frac{J_{**}}{p_{**}} + \sum_{i=1}^k \frac{J_k}{p_k} \leq 1 + p_{**} \left((1 + \delta_0^*)^2 - (1 + \delta_0^*)^5 + 2\delta_0^* \right) \leq 1 + p_{**} \delta_0^* (\delta_0^* - 1) \leq 1.$$

The various parameter selections based on the previous discussion are summarized next.

IV Select $\tilde{\varepsilon}_1 = \delta_0^*/(k+1)$, $\tilde{\varepsilon} = (\delta_0^*)^2$ (this guarantees $p_{**}(1 + k\tilde{\varepsilon}_1) < \delta_0^*/2$) and $\theta = -\mu(1 - \delta_0^*)/(1 + \delta_0^*)^5$. Set p_{**}, p_i for $i = 1, \dots, k$ and p_* according to (34), (35) and (36) respectively. Then, choose κ large enough so that $\kappa \geq e^{2a_{**}}/[4\theta^2 \delta_0^*]$ and at the same time $g(s) < 1$ implies $b - s > \eta_*$, with η_* also appearing in II) above.

We now can provide a precise description of the importance sampling scheme. Assume that the selection procedure indicated from I) to IV) above has been performed and let $S_0 = 0$. Suppose that the current position at time k , namely S_k , is equal to s and that $\tau_b > k$. We simulate the increment X_{k+1} according to the following law. If $g(s) < 1$ then we sample X_{k+1} with the mixture density in (9). Otherwise, if $g(s) = 1$ we sample X_{k+1} with density $f(\cdot)$. The corresponding importance sampling estimator is precisely

$$Z_b = I(\tau_b < \infty) \prod_{i=1}^{\tau_b} r_{S_{i-1}}(S_i - S_{i-1}). \quad (37)$$

Note that we have not discussed the termination of the algorithm – the expected value of τ_b under the proposed importance sampling distribution. Indeed, this is an issue that will be studied in the next section. Here we are only interested in the variance analysis of Z_b .

Proof of Theorem 1. We must show that the estimator Z_b defined in (37) is strongly efficient for estimating $u(b)$. Our discussion summarized in the selection process from I) to IV) above indicates that $g(\cdot)$ is a valid Lyapunov function. Therefore we have that

$$E^Q Z_b^2 \leq g(0).$$

Hence, according to (17),

$$\sup_{b>1} \frac{g(0)}{u^2(b)} < \infty.$$

■

5 Controlling the expected termination time

As mentioned previously, if Z_b is a strongly efficient estimator for $u(b)$, in order to compute $u(b)$ with ε relative error with at least $1 - \delta$ probability, one needs to generate $O(\varepsilon^{-2}\delta^{-1})$ (uniformly in b) i.i.d. copies of Z_b . The concept of strong efficiency by itself does not capture the complexity of generating a single replication of Z_b . In this section we will further investigate the computational cost of generating Z_b . We shall assume that sampling from the densities $q_s(\cdot)$ or $f(\cdot)$ takes at most a given constant computational cost, so the analysis reduces to finding a suitable upper bound for $E^Q \tau_b$.

We first assume that F is a regularly varying distribution. We will see that if I) to IV) and also V) below are satisfied then the expected termination time is $O(b)$. The key message is that we can always select $a_{**}, \delta_0^* > 0$ sufficiently small in order to satisfy both Lyapunov functions in Lemmas 7 and 8.

V If Assumption A holds, let η_* be large enough so that if $g(s) < 1$ (i.e. $b - s > \eta_* = G^{-1}(\kappa^{-1/2})$) then

$$\frac{\bar{F}(b-s)}{G(b-s)} \geq \frac{(\iota-1)(1-\delta_0^*)}{b-s}.$$

We also have that $a_{**}, \delta_0^* > 0$ are sufficiently close to zero such that

$$\delta_2^* = 2(\iota-1) \frac{(1-\delta_0^*)^2}{(1+\delta_0^*)^5} e^{-a_{**}} - 1 - 2(1 - e^{-2a_{**}/\iota})(\iota-1) > 0$$

with $\iota > 1.5$.

Proposition 6 *Suppose that Assumption A holds and $\iota > 1.5$. Then, the selection indicated in I) to V) yields both Theorem 1 and*

$$E^Q(\tau_b) < \rho_0 + \rho_1 b,$$

for $\rho_0, \rho_1 \in (0, \infty)$ independent of b .

Proof of Proposition 6. We will use Lemma 8 to finish the proof. We propose

$$h(s) = [\rho + b - s]I(s < b),$$

for some $\rho > 0$. First we note that

$$\begin{aligned} & E^Q(b-s-X; X \in (\Lambda^{-1}(\Lambda(b-s) - a_{**}), b-s]) \\ &= p_{**} \frac{P(X \in (\Lambda^{-1}(\Lambda(b-s) - a_{**}), b-s])}{P(X > \Lambda^{-1}(\Lambda(b-s) - a_{**}))} \times E(b-s-X | X \in (\Lambda^{-1}(\Lambda(b-s) - a_{**}), b-s]). \end{aligned} \quad (38)$$

Recall that

$$p_{**} = \min\{2\theta\bar{F}(b-s)/G(b-s), \tilde{\varepsilon}\} = \frac{2\theta(\iota-1)}{b-s} (1+o(1)) \quad (39)$$

as $b-s \nearrow \infty$, where $\theta = -\mu(1-\delta_0^*)/(1+\delta_0^*)^5$. Therefore, we can select $\eta_* > 0$ large enough so that if $b-s \geq \eta_*$

$$-\frac{2\mu(\iota-1)(1-\delta_0^*)^2}{(b-s)(1+\delta_0^*)^5} \leq p_{**} \leq -\frac{2\mu(\iota-1)}{(b-s)}.$$

Now, note that η_* can be chosen sufficiently large so that if $a = e^{-2a_{**}/\iota}$, then

$$\exp(-\Lambda(b-s) + \Lambda(a(b-s))) = \frac{P(X > b-s)}{P(X > a(b-s))} \leq \exp(-a_{**})$$

as long as $b-s \geq \eta_*$. Therefore,

$$X \geq \Lambda^{-1}(\Lambda(b-s) - a_{**})$$

implies $X \geq a(b-s)$ and we have that

$$E(b-s-X | X \in (\Lambda^{-1}(\Lambda(b-s) - a_{**}), b-s]) \leq (1-a)(b-s). \quad (40)$$

Together with (38), (39), and (40), if $b-s \geq \eta_*$ we obtain

$$E^Q(b-s-X; X \in (\Lambda^{-1}(\Lambda(b-s) - a_{**}), b-s]) \leq 2|\mu|(1-a)(\iota-1).$$

The previous estimates imply that by choosing $\eta_* > 0$ large enough we can guarantee that for all $b-s \geq \eta_*$ we have

$$\begin{aligned} & E^Q(h(s+X)) \\ &= E^Q(\rho + (b-s-X); s+X \leq b) \\ &\leq (1-Q(X > b-s))[\rho_0 + b-s-\mu + 2|\mu|(1-a)(\iota-1)] \\ &= (1-p_{**}e^{-a_{**}})[h(s) - \mu + 2|\mu|(1-a)(\iota-1)]. \end{aligned}$$

By noting that $\theta \leq |\mu|$, if $b-s \geq \eta_*$ and η_* is selected large enough we obtain that

$$\begin{aligned} & E^Q(h(s+X)) \\ &\leq h(s) - \mu - p_{**}e^{-a_{**}}h(s) + 2|\mu|(1-a)(\iota-1) + o(1) \\ &\leq h(s) - \mu + \frac{2\mu(\iota-1)(1-\delta_0^*)^2}{(1+\delta_0^*)^5}e^{-a_{**}} \\ &+ \frac{2\mu(\iota-1)(1-\delta_0^*)^2}{(1+\delta_0^*)^5(b-s)}e^{-a_{**}}\rho - 2\mu(1-a)(\iota-1) + o(1). \end{aligned}$$

The above inequality holds for all $\rho > 0$ provided that $b - s \geq \eta_* = G^{-1}(\kappa^{-1/2})$ so that $b - s > \eta_*$ if and only if $g(s) < 1$. Since $\iota > 1.5$, one can choose a_{**} and δ_0^* sufficiently small such that

$$\delta_2^* = 2(\iota - 1) \frac{(1 - \delta_0^*)^2}{(1 + \delta_0^*)^5} e^{-a_{**}} - 1 - 2(1 - e^{-2a_{**}/\iota})(\iota - 1) > 0$$

we conclude that

$$E^Q(h(s + X)) \leq h(s) + \mu\delta_2^*$$

as long as $g(s) < 1$. Now, if $g(s) = 1$ (i.e. if $0 \leq b - s < \eta_*$) we do not apply the change of measure and therefore

$$\begin{aligned} E^Q(h(s + X)) &= E[\rho + b - s - X; X \leq b - s] \\ &\leq h(s) - E(X|X < 0) - \rho P(X > \eta_*). \end{aligned}$$

Given the selection of κ (and therefore of $\eta_* = G^{-1}(\kappa^{-1/2})$), we can choose ρ large such that

$$-E(X|X < 0) - \rho P(X > \eta_*) \leq \mu\delta_2^* < 0.$$

Hence,

$$E^Q\tau_b < h(0)/|\mu| \delta_2^*.$$

Thereby, the conclusion of Lemma 8 follows by redefining the constants. ■

Remark 6 *The previous result concerning the condition $\iota > 1.5$ raises a couple of natural questions. First, what is special about a tail index $\iota = 1.5$? What would be required in order to obtain both strong efficiency and $E^Q\tau_b = O(b)$ assuming only $\iota > 1$? We believe that the previous result is basically optimal. We do not pursue this claim with full rigor here but provide an argument showing why we expect this to be the case. First, Theorem 2 implies the approximation*

$$P(b\delta n < \tau_b \leq b\delta(n+1)|\tau_b < \infty) = [P(Y_0 > \delta|\mu|n(\iota-1)) - P(Y_0 > \delta|\mu|(n+1)(\iota-1))](1+o(1))$$

as $b \nearrow \infty$ for any $\delta > 0$. Even if we could apply importance sampling directly to τ_b (rather than doing it through the X_j 's) it would be reasonable to select $Q(\cdot)$ so that

$$Q(b\delta n < \tau_b \leq b\delta(n+1)) = c_1(\delta) n^{-\gamma_1} (1+o(1))$$

as $b \nearrow \infty$. Since we wish to have $E^Q\tau_b < \infty$ we should impose the constraint $\gamma_1 > 2$. Now, we have that

$$P(Y_0 > \delta|\mu|n(\iota-1)) - P(Y_0 > \delta|\mu|(n+1)(\iota-1)) = \delta|\mu|(\iota-1) (1 + \delta|\mu|n)^{-\iota} (1+o(1))$$

as $n \nearrow \infty$. On the other hand, strong efficiency imposes the constraint that

$$\sum_{n=1}^{\infty} \left(\frac{P(Y_0 > \delta(n+1)) - P(Y_0 > \delta n)}{Q(b\delta n < \tau_b \leq b\delta(n+1))} \right)^2 Q(b\delta n < \tau_b \leq b\delta(n+1)) < \infty \quad (41)$$

which suggests

$$\sum_{n=1}^{\infty} n^{-2\iota+\gamma_1} < \infty. \quad (42)$$

Consequently, we also must have $2\iota > \gamma_1 + 1$. Combined with the previous constraint (i.e. $\gamma_1 > 2$), it yields $\iota > 3/2$.

We will show that if $\iota > 1$ we can control $1 + \gamma$ relative moments (for γ small enough) and still keep $E^Q \tau_b = O(b)$. However, before we do so, in order to complete the argument for the proof of Theorem 3 we will continue working with $\gamma = 1$ in the context of Assumption B.

Proposition 7 *If Assumptions B1-3 hold, we assume there exists $\delta > 0$ and $\beta \in [0, \beta_0]$ such that $\lambda(x) \geq \delta x^{\beta-1}$ for x sufficiently large. Then, there exist a_* , a_{**} , p_* , p_{**} , p_j , $j = 1, \dots, k$, such that Theorem 1 holds and, in addition,*

$$E^Q \tau_b \leq \rho_0 + \rho_1 b^{1-\beta}.$$

for ρ_0 and ρ_1 sufficiently large.

Proof of Proposition 7. Let $\beta \in (0, \beta_0)$ and consider the Lyapunov function,

$$h(s) = [\rho + (b - s)^{1-\beta}]I(s < b).$$

For all $\varepsilon > 0$,

$$\begin{aligned} & E^Q(h(s + X)) \\ & \leq Q\left(X \leq (1 - \varepsilon)(b - s)\right) E^Q\left(\rho + (b - s - X)^{1-\beta} | X \leq (1 - \varepsilon)(b - s)\right) \\ & \quad + (\rho + \varepsilon^{1-\beta}(b - s)^{1-\beta}) Q\left((1 - \varepsilon)(b - s) \leq X \leq b - s\right). \end{aligned}$$

With Assumptions B1-3, if $\beta = 0$, using L'Hopital rule on a subsequence, we have

$$\underline{\lim}_{x \rightarrow \infty} \frac{x \bar{F}(x)}{G(x)} = \underline{\lim}_{x \rightarrow \infty} \frac{-\bar{F}(x) + x \lambda(x) \bar{F}(x)}{\bar{F}(x)} = \infty;$$

if $\beta \in (0, \beta_0)$,

$$\underline{\lim}_{x \rightarrow \infty} \frac{x^{1-\beta} \bar{F}(x)}{G(x)} = \underline{\lim}_{x \rightarrow \infty} x^{1-\beta} \lambda(x) - (1 - \beta)x^{-\beta} \geq \delta.$$

There exists $\varepsilon, \delta' > 0$ small enough and η_* sufficiently large such that for all $b - s > \eta_*$ and all $\rho > 0$,

$$\begin{aligned} & E^Q(h(s + X)) \\ & \leq (1 - 2\theta\delta(b - s)^{\beta-1}) \left(\rho + (b - s)^{1-\beta} - (1 + \delta')(1 - \beta)(b - s)^{-\beta} \mu \right) \\ & \quad + 2\theta\delta(\rho + \varepsilon^{1-\beta}(b - s)^{1-\beta})(b - s)^{\beta-1} \\ & \leq (1 - 2\theta\delta(b - s)^{\beta-1}) \left(h(s) - (1 + \delta')(1 - \beta)(b - s)^{-\beta} \mu \right) \\ & \quad + 2\theta\delta(\rho + \varepsilon^{1-\beta}(b - s)^{1-\beta})(b - s)^{\beta-1} \\ & \leq h(s) - \theta\delta. \end{aligned}$$

The above derivation is true for all $\beta > 0$ satisfying conditions in the proposition. When $\beta = 0$ due to Assumption B1, one can always choose δ large such that $2\theta\delta > 3|\mu|$. Therefore, this derivation is true for all $\beta \in [0, \beta_0]$.

On the other hand, if $b - s \leq \eta_*$ and we select $\eta_* = G^{-1}(\kappa^{-1/2})$ so that $g(s) < 1$ if and only if $b - s > \eta_*$, we obtain that

$$\begin{aligned} & E^Q h(s + X) \\ &= Eh(s + X) \leq \rho + (b - s)^{1-\beta} - \rho P(X > \eta_*) + E((b - s - X)^{1-\beta} - (b - s)^{1-\beta}; X \leq b - s). \end{aligned}$$

Clearly, once η_* has been selected we can pick ρ large enough so that

$$-\rho P(X > \eta_*) + \sup_{0 \leq b-s \leq \eta_*} E((b - s - X)^{1-\beta} - (b - s)^{1-\beta}; X \leq b - s) \leq -\delta/2.$$

Therefore,

$$E^Q(h(s + X)) \leq h(s) - \delta/2$$

and we conclude the result by applying Lemma 8. ■

Proof of Theorem 3. The conclusion follows immediately from Propositions 6 and 7. ■

Finally, we come back to the problem of controlling $(1 + \gamma)$ -th moments in order to guarantee $E^Q \tau_b = O(b)$ when \bar{F} is regularly varying with $\iota > 1$. This corresponds to Theorem 4. The next proposition is central to the proof.

Proposition 8 *Suppose that Assumption A holds and that $\iota \in (1, 1.5]$. Then, we can choose a_* , a_{**} , p_* , and p_{**} , such that for each $\gamma \in (0, (\iota - 1)/(2 - \iota))$ there exists a $K > 0$,*

$$E^Q Z^{1+\gamma} \leq Ku(b)^{1+\gamma}$$

and $E^Q \tau_b = O(b)$ as $b \rightarrow \infty$.

Remark 7 *With a very similar argument as in Remark 6, we believe that the bound $1 + (\iota - 1)/(2 - \iota)$ is the highest moment that one can control while maintaining $O(b)$ expected termination time. An analogous constraint to (42) is that*

$$\sum_{n=1}^{\infty} n^{-(1+\gamma)(\iota-\gamma_1)-\gamma_1} < \infty.$$

This implies that $\gamma < (\iota - 1)/(\gamma_1 - \iota) \leq (\iota - 1)/(2 - \iota)$. Note that it is necessary to have $\gamma_1 > 2$ to have $O(b)$ expected termination time.

Proof of Proposition 8. The strategy is completely analogous to the case of $\gamma = 1$. We define

$$g_\gamma(s) = \min\{\kappa G(b - s)^{1+\gamma}, 1\}.$$

We need to verify the Lyapunov inequality only on $g_\gamma(s) < 1$ (as before the case $g_\gamma(s) = 1$ is automatic). We select

$$p_{**} = \min\{\theta \partial g_\gamma(s) / g_\gamma(s), \tilde{\varepsilon}\}$$

for $\tilde{\varepsilon}$ sufficiently small. Applying Lemma 7 we need to show that

$$\frac{J_*}{(1-p_{**})^\gamma} + \frac{J_{**}}{p_{**}^\gamma} \leq 1, \quad (43)$$

where J_* and J_{**} are redefined as

$$J_* = P(X \leq b-s - \Lambda^{-1}(\Lambda(b-s) - a_{**}))^\gamma E \left(\frac{g_\gamma(s+X)}{g_\gamma(s)}; X \leq b-s - \Lambda^{-1}(\Lambda(b-s) - a_{**}) \right)$$

$$J_{**} = P(X > \Lambda^{-1}(\Lambda(b-s) - a_{**}))^\gamma E \left(\frac{g_\gamma(s+X)}{g_\gamma(s)}; X > \Lambda^{-1}(\Lambda(b-s) - a_{**}) \right).$$

Note that the J_i terms analogous to (26) and (27) are all zero. At the same time, we need to make sure that we can find $\rho > 0$ such that if

$$h(s) = [\rho + (b-s)] I(b-s > 0)$$

then

$$E^Q h(s+X) \leq h(s) - \varepsilon \quad (44)$$

for some $\varepsilon > 0$ if $b > s$.

Inequality (43) can be obtained following the same steps as we did in I) to IV) in the previous section. First we note that if $\eta_* = G^{-1}(\kappa^{-1/(1+\gamma)})$ is large enough (or equivalently κ is sufficiently large)

$$\frac{J_{**}}{p_{**}^\gamma} \leq \frac{P(X > \Lambda^{-1}(\Lambda(b-s) - a_{**}))^{\gamma+1}}{g(s)p_{**}^\gamma} = \frac{e^{a_{**}(\gamma+1)} \bar{F}(b-s)}{\kappa(1+\gamma)^\gamma \theta^\gamma G(b-s)}.$$

Also, for any $\delta > 0$ we can ensure that if η_* is large enough and if $b-s > \eta_*$ then

$$\frac{\theta(1+\gamma)(\iota-1)(1-\delta)}{b-s} \leq p_{**} = \frac{\theta(1+\gamma)\bar{F}(b-s)}{G(b-s)} \leq \frac{\theta(1+\gamma)(\iota-1)(1+\delta)}{b-s}$$

and we also can ensure that

$$\frac{J_*}{(1-p_{**})^\gamma} \leq (1+\gamma(1+\delta)p_{**}) E \left(\frac{g_\gamma(s+X)}{g_\gamma(s)}; X \leq b-s - \Lambda^{-1}(\Lambda(b-s) - a_{**}) \right).$$

A similar development to that of Proposition 2 yields that η_* can be chosen so that if $b-s > \eta_*$,

$$E \left(\frac{g_\gamma(s+X)}{g_\gamma(s)}; X \leq b-s - \Lambda^{-1}(\Lambda(b-s) - a_{**}) \right) \leq 1 + \mu(1-\delta) \frac{\partial g_\gamma(s)}{g_\gamma(s)}.$$

Therefore,

$$\begin{aligned} \frac{J_*}{(1-p_{**})^\gamma} &\leq \left(1 + \mu(1-\delta) \frac{\partial g_\gamma(s)}{g_\gamma(s)} \right) (1+\gamma(1+\delta)p_{**}) \\ &= \left(1 + \mu(1-\delta) \frac{\partial g_\gamma(s)}{g_\gamma(s)} \right) \left(1 + \gamma(1+\delta) \frac{\theta \partial g_\gamma(s)}{g_\gamma(s)} \right) \end{aligned}$$

and then

$$\begin{aligned} & \frac{J_*}{(1-p_{**})^\gamma} + \frac{J_{**}}{p_{**}^\gamma} \\ & \leq \left(1 + \frac{\mu(1-\delta)(1+\gamma)\overline{F}(b-s)}{G(b-s)}\right) \left(1 + \frac{\theta\gamma(1+\delta)(1+\gamma)\overline{F}(b-s)}{G(b-s)}\right) \\ & \quad + \frac{e^{a_{**}(\gamma+1)}}{\kappa(1+\gamma)\gamma\theta^\gamma} \times \frac{\overline{F}(b-s)}{G(b-s)}. \end{aligned}$$

We then can select $\theta = |\mu|(1-\delta)^2/[\gamma(1+\delta)]$, $a_{**} < \delta$ and κ sufficiently large such that the right hand side the above display is less than one. At the same time, the analysis required to enforce (44) is similar to that of Proposition 6. We, therefore, omit the details. The key fact is now that

$$-\frac{(1+\gamma)\mu(\iota-1)(1-\delta)^3}{\gamma(b-s)(1+\delta)} \leq p_{**}$$

and now we need to enforce

$$\delta_2^* = \frac{(1+\gamma)(\iota-1)(1-\delta)^3}{\gamma(1+\delta)} e^{-a_{**}} - 1 - (1+\gamma)(1-a)(\iota-1) > 0,$$

where $a = e^{-2a_{**}/\iota}$. This can always be done if we choose $\gamma < (\iota-1)/(2-\iota)$ and $\delta, a_{**} > 0$ sufficiently small. ■

Now we provide the proof of Theorem 4.

Proof of Theorem 4. From the result in Proposition 8, the $(1+\gamma)$ -th moment of the estimator and $E^Q\tau_b$ is properly controlled. We need to bound the total computation time to achieve prescribed relative accuracy. Let W_1, W_2, \dots be a sequence of non-negative i.i.d. random variables with unit mean and suppose that $EW_i^{1+\gamma} \leq K$ for $\gamma > 0$. Define $R_n = (W_1 + W_2 + \dots + W_n)/n$ and note that

$$P(|R_n - 1| \geq \varepsilon) \leq P\left(|R_n - 1| \geq \varepsilon, \max_{i \leq n} W_i \leq n\right) + P\left(\max_{i \leq n} W_i > n\right).$$

Now using Chebyshev's inequality we have that

$$P\left(\max_{i \leq n} W_i > n\right) \leq nP(W_1 > n) \leq \frac{K}{n^\gamma}.$$

On the other hand, given $\max_{i \leq n} W_i < n$, W_i 's are still i.i.d. and

$$P\left(|R_n - 1| \geq \varepsilon \mid \max_{i \leq n} W_i \leq n\right) \leq \frac{E(W_i^2 | W_i \leq n) + o(1)}{n\varepsilon^2} = \frac{E(W_i^2 I(W_i \leq n)) + o(1)}{n\varepsilon^2 P(W_i \leq n)}.$$

The $o(1)$ term in the above display is in fact $(E(W_i | W_i \leq n) - 1)^2$. Then, we have that for $\gamma \in (0, 1)$

$$\begin{aligned} E(W_i^2 I(W_i \leq n)) &= 2E\left(I(W_i \leq n) \int_0^{W_i} t dt\right) \\ &\leq 2 \int_0^n tP(W_i > t) dt \leq 2K \int_0^n \frac{1}{t^\gamma} dt = \frac{2K}{1-\gamma} n^{1-\gamma}. \end{aligned}$$

Therefore, for n sufficiently large we have that

$$P\left(|R_n - 1| \geq \varepsilon, \max_{i \leq n} W_i \leq n\right) \leq \frac{3K}{(1 - \gamma)\varepsilon^2 n^\gamma}.$$

Thus, we have that

$$P(|R_n - 1| \geq \varepsilon) \leq \frac{3K}{(1 - \gamma)\varepsilon^2 n^\gamma} + \frac{K}{n^\gamma} \leq \frac{4K}{(1 - \gamma)\varepsilon^2 n^\gamma}.$$

Applying these considerations to $W_n = Z_b/u(b)$ and letting $4K/[(1 - \gamma)\varepsilon^2 n^\gamma] \leq \delta$ we obtain the conclusion of the theorem. ■

6 Approximation in total variation and conditional limit theorems

6.1 Approximation of the random walk up to τ_b

We will need the following lemma for the proof of approximation in total variation.

Lemma 10 *Let Q_0 and Q_1 be probability measures defined on the same σ -field \mathcal{F} such that $dQ_1 = M^{-1}dQ_0$ for a positive r.v. $M > 0$. Suppose that for some $\varepsilon > 0$, $E^{Q_1}(M^2) = E^{Q_0}M \leq 1 + \varepsilon$. Then,*

$$\sup_{A \in \mathcal{F}} |Q_1(A) - Q_0(A)| \leq \varepsilon^{1/2}.$$

Proof of Lemma 10. Note that

$$\begin{aligned} |Q_1(A) - Q_0(A)| &= |E^{Q_1}(1 - M; A)| \\ &\leq E^{Q_1}(|M - 1|) \leq [E^{Q_1}(M - 1)^2]^{1/2} = (E^{Q_1}M^2 - 1)^{1/2} \leq \varepsilon^{1/2}. \end{aligned}$$

■

Also, it is not hard to verify that by letting $P^{(b)}(\cdot) = P(\cdot | \tau_b < \infty)$ we have

$$\frac{dP^{(b)}}{dQ} = \frac{Z_b}{P(\tau_b < \infty)}.$$

Then, it is sufficient to show that for ε arbitrarily small there exists b sufficiently large depending on ε ,

$$E^Q Z_b^2 < (1 + \varepsilon)u^2(b).$$

Theorem 8 *Suppose that Assumption A or B1-B3 hold. For any $\varepsilon > 0$, there exists $\eta_* > 0$ such that for all $b > \eta_*$, there exists a choice of p_* , p_{**} , p_j , $j = 1, \dots, k$ such that the corresponding estimator L_b satisfies,*

$$E^Q Z_b^2 \leq (1 + \varepsilon)u^2(b). \quad (45)$$

Therefore, the importance sampling distribution converges in total variation to the conditional distribution of the random walk given $\{\tau_b < \infty\}$, as $b \rightarrow \infty$.

Proof of Theorem 8. Given $\varepsilon, \varepsilon' > 0$ small, we consider $\kappa > 0$ and functions

$$\gamma(s) = \begin{cases} 1 + 5\varepsilon + \kappa s^{1+\varepsilon'} / b^{1+\varepsilon'}, & s > 0 \\ 1 + 5\varepsilon, & s \leq 0 \end{cases}$$

$$g(s) = \min\{1, \mu^{-2}\gamma(s)G^2(b-s)\}.$$

Let $\eta_* = \sup\{b-s : g(s) = 1\}$. We can easily see that $\eta_* \rightarrow \infty$ as $\kappa \rightarrow \infty$. Also,

$$1 + 5\varepsilon \leq \gamma(s) \leq \kappa + 1 + 5\varepsilon,$$

for all $s \leq b$. We proceed with a similar development as in the previous section. We adopt the same notation as in (24), (25), (26), and (27). Since $\gamma(s)$ is bounded, results as in Propositions 3, 4, and 5 still hold. In addition, we can choose a_{**} small enough such that

$$J_{**} \leq \frac{P^2(X > \Lambda^{-1}(\Lambda(b-s) - a_{**}))}{p_{**}g(s)} \leq (1 + \varepsilon) \frac{\bar{F}^2(b-s)}{p_{**}g(s)}.$$

There is one last term, namely J_* . Note that

$$\frac{g(s+X)}{g(s)} = \frac{G^2(s+X)}{G^2(s)} + \frac{G^2(s+X)}{G^2(s)} \left(\frac{\gamma(s+X)}{\gamma(s)} - 1 \right).$$

According to the proof of Proposition 2 (more specifically (29)),

$$E \left(\frac{G^2(s+X)}{G^2(s)}; X \leq b-s - \Lambda^{-1}(\Lambda(b-s) - a_*) \right) \leq 1 + (2\mu + o(1))\bar{F}(b-s)/G(b-s).$$

as $b-s \rightarrow \infty$. Now, we consider the term

$$E \left(\frac{G^2(s+X)}{G^2(s)} \left(\frac{\gamma(s+X)}{\gamma(s)} - 1 \right); X \leq b-s - \Lambda^{-1}(\Lambda(b-s) - a_*) \right).$$

For all $b \geq s > b^{\varepsilon'}$ and $s+X > 0$,

$$\frac{G^2(s+X)}{G^2(s)} \left(\frac{\gamma(s+X)}{\gamma(s)} - 1 \right) = \kappa \gamma^{-1}(s) s^{1+\varepsilon'} b^{-1-\varepsilon'} \left((1 + X/s)^{1+\varepsilon'} - 1 \right) \frac{G^2(s+X)}{G^2(s)}.$$

Therefore, for $b \geq s > b^{\varepsilon'}$, by dominated convergence,

$$\gamma(s) E \left(\frac{b^{1+\varepsilon'} G^2(s+X)}{s^{\varepsilon'} G^2(s)} \left(\frac{\gamma(s+X)}{\gamma(s)} - 1 \right); X \leq b-s - \Lambda^{-1}(\Lambda(b-s) - a_*) \right) \rightarrow \kappa(1 + \varepsilon')\mu,$$

as $b-s \rightarrow \infty$. For $s \leq b^{\varepsilon'}$,

$$\begin{aligned} & E \left(\frac{G^2(s+X)}{G^2(s)} \left(\frac{\gamma(s+X)}{\gamma(s)} - 1 \right); X \leq b-s - \Lambda^{-1}(\Lambda(b-s) - a_*) \right) \\ &= O(b^{-1-\varepsilon'+\varepsilon'^2}) = o(\bar{F}(b-s)/G(b-s)) \end{aligned}$$

as $b \nearrow \infty$ uniformly over $s \leq b^{\varepsilon'}$. Consequently, it follows that

$$E \left(\frac{g(s+X)}{g(s)}; X \leq b-s - \Lambda^{-1}(\Lambda(b-s) - a_*) \right) \leq 1 + (2\mu + o(1)) \bar{F}(b-s)/G(b-s),$$

as $b-s \rightarrow \infty$. We choose,

$$p_{**} = \min\{\varepsilon, -(1-\varepsilon)\mu\bar{F}(b-s)/G(b-s)\}, \quad p_j = \varepsilon^2 p_{**}$$

Then,

$$\begin{aligned} E \left[\frac{g(s+X)}{g(s)} r_s(X) \right] &\leq \left(1 + (1-\varepsilon + o(\varepsilon))\mu \frac{\bar{F}(b-s)}{G(b-s)} \right)^{-1} \left[1 + (2\mu + o(1)) \frac{\bar{F}(b-s)}{G(b-s)} \right] \\ &\quad + o(1)k\varepsilon^{-2} \bar{F}(b-s)/G(b-s) - (1+\varepsilon) \frac{\mu\bar{F}(b-s)}{\gamma(s)G(b-s)(1-\varepsilon)}. \end{aligned}$$

When $s \leq b/2$,

$$\begin{aligned} E \left[\frac{g(s+X)}{g(s)} r_s(X) \right] &\leq 1 - (1+o(\varepsilon))\mu \frac{\bar{F}(b-s)}{G(b-s)} + (2\mu + o(1)) \frac{\bar{F}(b-s)}{G(b-s)} \\ &\quad + o(1)k\varepsilon^{-2} \frac{\bar{F}(b-s)}{G(b-s)} - (1+3\varepsilon) \frac{\mu\bar{F}(b-s)}{\gamma(s)G(b-s)}. \end{aligned}$$

Because $\gamma(s) \geq 1 + 5\varepsilon$, for b large enough, $E \left[\frac{g(s+X)}{g(s)} r_s(X) \right] \leq 1$, when $s \leq b/2$. For $s \geq b/2$,

$$\gamma(s) \geq \kappa/4.$$

Then

$$\begin{aligned} E \left[\frac{g(s+X)}{g(s)} L(X) \right] &\leq 1 - (1+o(\varepsilon))\mu \frac{\bar{F}(b-s)}{G(b-s)} + (2\mu + o(1)) \frac{\bar{F}(b-s)}{G(b-s)} \\ &\quad + o(1)k\varepsilon^{-2} \frac{\bar{F}(b-s)}{G(b-s)} - \frac{4(1+3\varepsilon)}{\kappa} \frac{\mu\bar{F}(b-s)}{G(b-s)}. \end{aligned}$$

For any $\varepsilon > 0$ one can always choose κ large enough such that $E \left[\frac{g(s+X)}{g(s)} r_s(X) \right] \leq 1$ when $s \geq b/2$ and $g(s) < 1$. Therefore,

$$E^Q L^2 \leq g(0) = (1+5\varepsilon)\mu^{-2}G(b)^2,$$

for b large enough. The conclusion then follows from Lemma 10 and Theorem 7. \blacksquare

Proof of Theorem 5. The conclusion is a direct application of Lemma 10 and Theorem 8. \blacksquare

Here we emphasize that the choices of parameters of the mixture family in the current section are different from those in Section 5. Especially for the regularly varying case with $\iota \in (1.5, 2)$, in order to have finite expected termination, we will have the importance sampling distribution deviate from the zero-variance change of measure.

6.2 Conditional central limit theorem

The goal of this section is to provide a functional approximation to the joint distribution of

$$\{(\tau_b, S_{\lfloor u\tau_b \rfloor}, S_{\tau_b}) : u \in [0, 1]\},$$

conditional on $\{\tau_b < \infty\}$ as $b \rightarrow \infty$. To make the discussion smooth, we postpone some technical proofs to Appendix B.

For all the theorems so far, we assume either Assumption A or Assumptions B1-B3. In this section, in the setting of Assumption B, we will further impose Assumption B4.

The approximation will be obtained based on a coupling of two processes governed according to a probability measure which shall be denoted by Q^* . Our importance sampling distribution induces a process that behaves most of the time like a regular random walk, except that occasional large jumps occur with probability p_{**} . We will couple this process with a regular random walk and argue that with high probability as $b \nearrow \infty$ we have that τ_b coincides precisely with the first of such large jumps.

We now proceed to formalize this intuition. Consider the process $\hat{S} = \{\hat{S}_n : n \geq 0\}$, where $\hat{S}_n = \hat{X}_1 + \dots + \hat{X}_n$, $\hat{S}_0 = 0$, and we have that

$$Q^*(\hat{X}_{n+1} \in dx | \hat{S}_n = s) \triangleq q_s(x) dx = r_s^{-1}(x) f(x) dx. \quad (46)$$

The function $r_s^{-1}(x)$ is chosen to satisfy the conditions of Theorem 8. We shall slightly abuse notation by letting $\tau_b = \inf\{n : \hat{S}_n > b\}$.

We further introduce a random walk $\tilde{S} = \{\tilde{S}_n : n \geq 1\}$ such that $\tilde{S}_n = \tilde{X}_1 + \dots + \tilde{X}_n$ and with the property that the \tilde{X}_i 's are i.i.d. under Q^* and have density

$$Q^*(\tilde{X}_i \in dx) = f(x) dx. \quad (47)$$

The joint law of \hat{S} and \tilde{S} will be described next.

We first define

$$p(s) = \frac{p_* I(b - s > \eta_*)}{P(X \leq b - s - \Lambda^{-1}(\Lambda(b - s) - a_*))} + I(b - s \leq \eta_*). \quad (48)$$

Note that by possibly increasing the selection of κ and $\eta_* = \sup\{b - s : g(s) = 1\}$ in Theorem 8, we can always guarantee that $p(s) \in [0, 1]$. Actually $p(s) \rightarrow 1$ as $b - s \rightarrow \infty$. Next define

$$q_s^*(x) = I(p(s) < 1) (1 - p(s))^{-1} (q_s(x) - p(s) f(x)). \quad (49)$$

The next lemma shows that $q_s^*(\cdot)$ is a density function and provides a decomposition of $q_s(x)$ that will allow us to describe the joint law of \hat{S} and \tilde{S} . The proof of the lemma is given in Appendix B.

Lemma 11 *If $p(s) < 1$ we have that $q_s^*(\cdot)$ is a density function provided that κ (and therefore η_*) are chosen large enough. We thus have the mixture decomposition*

$$q_s(x) = p(s) f(x) + (1 - p(s)) q_s^*(x). \quad (50)$$

The processes \hat{S} and \tilde{S} evolve jointly as follows under Q^* . First simply let \tilde{S} evolve according to (47). Now, at any given time $n + 1$ the evolution of \tilde{S} obeys the following rule. Given that $\hat{S}_n = s$, \hat{X}_{n+1} is constructed as follows. First, we sample a Bernoulli random variable to choose among $f(\cdot)$ and $q_s^*(\cdot)$ according to the probabilities $p(s)$ and $1 - p(s)$ respectively. If $f(\cdot)$ has been chosen, we let $\hat{X}_{n+1} = \tilde{X}_{n+1}$. Otherwise, we construct \hat{X}_{n+1} from the $q_s^*(\cdot)$ and \tilde{X}_{n+1} from $f(x)$ independently. We further let

$$N_b = \inf\{n \geq 1 : \tilde{X}_n \neq \hat{X}_n\},$$

which is the first time that $f(x)$ is not chosen. We intend to show that $P(N_b = \tau_b) \rightarrow 1$ as $b \rightarrow \infty$. The result is summarized in the following lemmas and propositions whose proofs are given in Appendix B.

Lemma 12

$$\lim_{b \rightarrow \infty} Q^*(N_b < \infty) = 1.$$

Lemma 13 *Let ε be chosen as in Theorem 8. There exists $b_0 > 0$ (depending on a_{**} and ε) and $\gamma(a_{**}, \varepsilon) > 0$ such that $\gamma(a_{**}, \varepsilon) \rightarrow 0$ as $a_{**} \rightarrow 0$ and $\varepsilon \rightarrow 0$, satisfying that*

$$Q^*(\tau_b = N_b) \geq 1 - \gamma(a_{**}, \varepsilon),$$

for all $b > b_0$, where $\tau_b = \inf\{n \geq 1 : \hat{S}_n \geq b\}$.

Now, we are ready to present the result which uses \tilde{S} to approximate the process \hat{S} up to time τ_b .

Proposition 9 *There exists a family of sets $(B_b : b > 0)$ such that $P(B_b) \rightarrow 1$ as $b \rightarrow \infty$ and with the property that for all $\tilde{S} \in B_b$*

$$Q^*(N_b > ta(b) | \tilde{S}) = P(Z_\theta > t | \mu)(1 + o(1)),$$

as $b \rightarrow \infty$, where $a(x) = G(x)/\bar{F}(x)$.

- Under Assumption A,

$$P(Z_\theta > t) = \left(1 + \frac{t}{\iota - 1}\right)^{-\frac{2\theta(\iota-1)}{|\mu|}},$$

for all $t \geq 0$.

- Under Assumptions B1-4,

$$P(Z_\theta > t) = e^{-\frac{2\theta t}{|\mu|}}.$$

Proof of Theorem 6. Thanks to Theorem 8, the distribution of $\{\hat{S}_n : 1 \leq n \leq \tau_b\}$ under Q^* converges in total variation to the distribution of $\{S_n : 1 \leq n \leq \tau_b\}$ given $\tau_b < \infty$ under P . It is sufficient to show the limit theorem of $\{\hat{S}_n : 1 \leq n \leq \tau_b\}$ under Q^* .

Thanks to Proposition 9, we are able to construct a random variable Z_θ following the distributions stated in Proposition 9 such that Z_θ is independent of \tilde{S} and

$$\frac{N_b}{a(b)} - \frac{Z_\theta}{|\mu|} \rightarrow 0,$$

almost surely as $b \rightarrow \infty$. Thanks to Lemma 13, we have that

$$\left(\frac{N_b}{a(b)}, \left\{ \frac{\tilde{S}_{tN_b} - t\mu N_b}{\sqrt{N_b}} \right\}_{0 \leq t < 1}, \frac{\hat{S}_{N_b} - b}{a(b)} \right) - \left(\frac{\tau_b}{a(b)}, \left\{ \frac{\hat{S}_{t\tau_b} - t\mu\tau_b}{\sqrt{\tau_b}} \right\}_{0 \leq t < 1}, \frac{\hat{S}_{\tau_b} - b}{a(b)} \right) \rightarrow 0$$

in probability as $b \rightarrow \infty$ (in fact, the convergence holds for almost every \tilde{S} in the sequence B_b). Further, as $b \rightarrow \infty$, we can let $\theta \rightarrow -\mu/2$. So it is possible to construct a random variable Y_0 independent of \tilde{S} and following distribution stated in the theorem such that

$$Z_\theta \rightarrow Y_0,$$

almost surely as $b \rightarrow \infty$. Now, using a standard strong approximation result (see for instance [20]) we can (possibly by further enlarging the probability space) assume that

$$\tilde{S}_{[t]} = \mu t + \sigma B(t) + e(t) \tag{51}$$

where $e(\cdot)$ is a (random) function such that

$$\frac{e(xt)}{t^{1/2}} \rightarrow 0$$

with probability one uniformly on compact sets on $x \geq 0$ as $t \nearrow \infty$. Therefore, we have that

$$\frac{\tilde{S}_{tN_b} - t\mu N_b}{\sqrt{N_b}} = \frac{1}{\sqrt{bY_0 + tb\xi_b}} \sigma B(tbY_0 + tb\xi_b) + e_b(t),$$

where $\xi_b \rightarrow 0$ as $b \rightarrow \infty$. We now verify that for each $z > 0$

$$\sup_{0 \leq u \leq 1} \left| \frac{B(ubz + ub\xi_b) - B(ubz)}{b^{1/2}} \right| \rightarrow 0$$

Given $z > 0$ select b large enough so that $\xi_b \leq \varepsilon$. Then, it suffices to bound the quantity

$$\sup_{u, s \in (0, 1): |u-s| \leq \varepsilon} \left| \frac{B(ubz) - B(sbz)}{b^{1/2}} \right|.$$

However, by the invariance principle the previous quantity equals in distribution

$$z^{1/2} \sup_{u, s \in (0, 1): |u-s| \leq \varepsilon} |B(u) - B(s)|,$$

which is precisely the modulus of continuity of Brownian motion evaluated ε . By continuity of Brownian motion, its modulus of continuity goes to zero almost surely as $\varepsilon \rightarrow 0$. Consequently, we obtain

$$\left(\frac{N_b}{a(b)}, \left\{ \frac{\tilde{S}_{tN_b} - t\mu N_b}{\sqrt{N_b}} \right\}_{0 \leq t < 1}, \frac{\hat{S}_{N_b} - b}{a(b)} \right) - \left(\frac{Y_0}{|\mu|}, \left\{ \frac{\tilde{S}_{ta(b)Y_0/|\mu|} + ta(b)Y_0}{\sqrt{a(b)Y_0/|\mu|}} \right\}_{0 \leq t < 1}, \frac{\hat{S}_{N_b} - b}{a(b)} \right) \Rightarrow 0.$$

Because Y_0 is independent of \tilde{S} , using the invariance principle for Brownian motion, we have that

$$\left(\frac{Y_0}{|\mu|}, \left\{ \frac{\tilde{S}_{ta(b)Y_0/|\mu|} + ta(b)Y_0}{\sqrt{a(b)Y_0/|\mu|}} \right\}_{0 \leq t < 1}, \frac{\hat{S}_{N_b} - b}{a(b)} \right) \Rightarrow \left(\frac{Y_0}{|\mu|}, \{\sigma B(t)\}_{0 \leq t < 1}, Y_1 \right).$$

Now, we figure out the joint distribution between Y_0 and Y_1 . Note that the overshoot $\hat{S}_{\tau_b} - b$ clearly satisfies with high probability as $b \nearrow \infty$

$$\frac{\hat{S}_{N_b} - b}{a(b)} = \frac{\hat{X}_{N(b)} + \tilde{S}_{N(b)-1} - b}{a(b)}.$$

In turn, we have,

$$\frac{\tilde{S}_{N(b)-1}}{a(b)} + Y_0 \rightarrow 0$$

in probability. In addition, the conditional distribution of $\hat{X}_{N(b)}$ given $\tilde{S}_{N(b)-1}$ is asymptotically (as $b \rightarrow \infty$) that of \tilde{X} given that $\tilde{X} > b - \tilde{S}_{N(b)-1}$ and $\tilde{S}_{N(b)-1}$, where \tilde{X} is a random variable with density $f(\cdot)$ independent of $\tilde{S}_{N(b)-1}$. Therefore, the law of $(\hat{X}_{N(b)} + \tilde{S}_{N(b)-1} - b)/a(b)$ given $\tilde{S}_{N(b)-1}$ can be approximated by that of $\tilde{X}/a(b) - Y_0 - b/a(b)$ given Y_0 and $\tilde{X} - Y_0 a(b) > b$.

In the setting of Assumptions B1-B4, we establish in the proof of Proposition 9 that $a(b) = (1 + o(1))/\lambda(b)$ as $b \nearrow \infty$. Because of Assumption B1 we have that $a(b) = o(b)$. Because of Assumption B4 we have that for each $y > 0$

$$P(\tilde{X} > ya(b) + Y_0 a(b) + b | \tilde{X} > b + Y_0 a(b), Y_0) \rightarrow P(Y_1 > y) = \exp(-y) \quad (52)$$

as $b \nearrow \infty$. Hence, Y_1 is an exponential random variable with expectation one and is independent of Y_0 .

Now, suppose that Assumption A holds. We have that $a(b) = b/(\iota - 1) + o(b)$ as $b \nearrow \infty$. Therefore,

$$\begin{aligned} & P(\tilde{X} - (Y_0 a(b) + b) > ya(b) | \tilde{X} > Y_0 a(b) + b, Y_0) \\ &= (1 + o(1)) P(\tilde{X} - (Y_0 + \iota - 1)a(b) > ya(b) | \tilde{X} > (Y_0 + \iota - 1)a(b), Y_0) \\ &\rightarrow P(W > y/(Y_0 + \iota - 1) | Y_0), \end{aligned}$$

where

$$P(W > t) = (1 + t)^{-\iota}$$

for $t \geq 0$. Now we need to verify that the law of (Y_0, Y_1) as stated in the theorem coincides with that of $(Y_0, W[Y_0 + (\iota - 1)])$. First we note that the joint density of (Y_0, Y_1) is given by

$$\frac{\partial^2}{\partial y_0 \partial y_1} P(Y_0 > y_0, Y_1 > y_1) = \frac{\iota}{\iota - 1} (1 + (y_0 + y_1)/(\iota - 1))^{-\iota-1}.$$

Therefore,

$$\frac{P(Y_1 \in dy_1 | Y_0 = y_0)}{dy_1} \propto (\iota - 1 + y_0 + y_1)^{-\iota-1}.$$

On the other hand,

$$\frac{P(W[y_0 + (\iota - 1)] \in dy_1)}{dy_1} = \iota (1 + y_1/[y_0 + (\iota - 1)])^{-\iota-1} \propto (\iota - 1 + y_0 + y_1)^{-\iota-1}.$$

The independence between $B(t)$ and (Y_0, Y_1) are straightforward. This concludes the proof of the theorem. ■

7 Implementation and examples

We implemented the algorithm and compare the performance with other existing algorithms in literature. In particular, we investigated two cases: regularly varying distribution and Weibull like distribution.

Regularly varying distribution. We consider the increment has the following representation.

$$X_i = V_i - T_i,$$

where V_i are i.i.d. with distribution that $P(V_i > v) = (1 + v)^{-2.5}$ for $v > 0$ and T_i 's are i.i.d. exponential random variables with expectation $4/3$. It is not hard to verify that $E(X_i) = -2/3$. In fact, this corresponds to the tail probability of the steady-state waiting time of an $M/G/1$ queue. There are a few provably efficient algorithms in literature including. Asmussen and Kroese (2006) (AK) [6], and Dupuis, Leder and Wang (2006) (DLW) [18] proposed efficient rare-event simulation estimators for geometric sums of regularly varying random variables. Blanchet and Glynn (2008) (BG) [9], and Blanchet, Glynn, and Liu (2007) (BGL) [14] proposed estimators for the tail of the steady state $G/G/1$ waiting time. Table 1 compares the performance of these algorithms. We use BL to denote the algorithm proposed in the current paper, with one cut-off point $c_0 = 0.9(b - s)$.

[Estimation] [Std. Error]	$b = 10^2$	$b = 10^3$	$b = 10^4$
BL	$1.047e - 03$ $3.76e - 05$	$3.175e - 05$ $2.602e - 07$	$9.877e - 07$ $8.187e - 09$
AK	$1.199e - 03$ $1.479e - 05$	$3.145e - 05$ $2.186e - 07$	$9.980e - 07$ $6.945e - 09$
BG	$1.079e - 03$ $5.968e - 06$	$3.146e - 05$ $9.725e - 08$	$9.980e - 07$ $2.073e - 09$
BGL	$1.022e - 03$ $3.835e - 05$	$3.167e - 05$ $1.598e - 06$	$1.128e - 06$ $7.280e - 08$
DLW	$1.046e - 03$ $5.195e - 06$	$3.163e - 05$ $1.694e - 07$	$9.905e - 07$ $2.993e - 09$

Table 1: Estimated tail probabilities of regularly varying random walks

Weibull-type distribution For the Weibull-type case, we consider the increment to have the following distribution,

$$P(X > x) = e^{-2\sqrt{t+1}},$$

for $t \geq -1$ and $EX_i = -\frac{1}{2}$. Table 2 compares the algorithm in this paper (BL) and that of Blanchet and Glynn (2008) (BG). For the implementation, we choose that $c_0 = \sqrt{b-s}$, $c_1 = 0.1(b-s)$, $c_2 = 0.5(b-s)$, $c_3 = 0.9(b-s)$, $c_4 = b-s - \sqrt{b-s}$.

[Estimation] [Std. Error]	$b = 250$	$b = 500$	$b = 650$
BL	$6.985e - 13$ $5.639e - 14$	$1.778e - 18$ $1.936e - 19$	$3.900e - 21$ $5.696e - 22$
BG	$7.076e - 13$ $1.20e - 14$	$1.897e - 18$ $5.083e - 20$	$3.971e - 21$ $7.95e - 23$

Table 2: Estimated tail probabilities of the Weibull-type distribution

A Technical proofs in Sections 3 and 4

Proof of Lemma 1. Observe that B2 implies $\log(\Lambda(x)/\Lambda(b_0)) \leq \log((x/b_0)^{\beta_0})$. In other words, $\Lambda(x) \leq \Lambda(b_0) b_0^{-\beta_0} x^{\beta_0}$. Consequently, substituting into B2 we have that for $x \geq b_0$

$$\lambda(x) \leq \beta_0 \Lambda(x) / x \leq \beta_0 \Lambda(b_0) b_0^{-\beta_0} x^{\beta_0-1} = O(x^{\beta_0-1}).$$

■

Proof of Lemma 2. First, since $G(\cdot)$ is decreasing then for $x \leq b - \Lambda^{-1}(\Lambda(b) - a_*)$

$$\frac{G(b-x)}{G(b)} \leq \frac{G(\Lambda^{-1}(\Lambda(b) - a_*))}{G(b)}.$$

By continuity of $G(\cdot)$ it suffices to show that the right hand side is bounded for all b sufficiently large. Using L'Hopital's rule we conclude that

$$\frac{G(\Lambda^{-1}(\Lambda(b) - a_*))}{G(b)} \sim \frac{\exp(-\Lambda(b) + a_*)}{\exp(-\Lambda(b))} \frac{d}{dx} \Lambda^{-1}(\Lambda(x) - a_*) \Big|_{x=b}.$$

Now, note that for all $x \geq b_0$

$$\frac{d}{dx} \Lambda^{-1}(\Lambda(x) - a_*) = \frac{\lambda(x)}{\lambda(\Lambda^{-1}(\Lambda(x) - a_*))} \leq \frac{\lambda(x)}{\lambda(\Lambda^{-1}(\Lambda(x)))} = 1.$$

The inequality follows from the fact that $\lambda(\cdot)$ is non increasing and $a_* > 0$. This allows to conclude the statement of the lemma. ■

Proof of Lemma 3. The second part assuming that $\bar{F}(\cdot)$ is regularly varying follows from Karamata's theorem. Now, for non-regularly varying part, we simply note using L'Hopital's rule and Lemma 1,

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x)}{G(x)} = \lim_{x \rightarrow \infty} \lambda(x) = 0.$$

The lower bound follows immediately. Again, using L'Hopital's rule, the upper bound then follows from the fact that

$$\lim_{x \rightarrow \infty} \frac{x\bar{F}(x)}{G(x)} = \lim_{x \rightarrow \infty} \frac{x\lambda(x)\bar{F}(x) - \bar{F}(x)}{\bar{F}(x)} = \infty.$$

The last step is thanks to Assumption B1.

■

Proof of Lemma 4. This is a direct application of condition B2. Indeed, if $x \geq b_0 > 0$ and $y \geq 0$

$$\log \Lambda(x+y) - \log \Lambda(x) = \int_x^{x+y} \partial \log \Lambda(t) dt \leq \int_x^{x+y} \beta_0 t^{-1} dt = \beta_0 \log \left(\frac{x+y}{x} \right),$$

which is equivalent to the statement of the lemma. ■

Proof of Lemma 5. Equivalently, we must show that for x sufficiently large

$$a_* \geq \Lambda(x) - \Lambda(x - x^\alpha),$$

where $\alpha = (1 - \beta_0)/2$. Now, note using Lemma 4 that

$$\Lambda(x) - \Lambda(x - x^\alpha) \leq \Lambda(x - x^\alpha) \left(\frac{\Lambda(x)}{\Lambda(x - x^\alpha)} - 1 \right) \leq \Lambda(x - x^\alpha) \left(\left(\frac{x}{x - x^\alpha} \right)^{\beta_0} - 1 \right).$$

For all x sufficiently large, using a Taylor expansion, the right hand side is bounded by $\Lambda(x - x^\alpha) (2\beta_0 x^{\alpha-1})$. Consequently, once again applying Lemma 4 we conclude that

$$\Lambda(x) - \Lambda(x - x^\alpha) \leq \Lambda(x - x^\alpha) (2\beta_0 x^{\alpha-1}) \leq 4\beta_0 \Lambda(b_0) x^{\beta_0-1+\alpha}$$

The right hand side goes to zero as $x \nearrow \infty$ given our selection of α and therefore is less than a_* for all x sufficiently large as required. ■

Proof of Lemma 6. If Assumption A is satisfied then it is well known that both F and G are subexponential. Let us then assume that B2 holds, and then we obtain $x\lambda(x) \leq \beta_0 \Lambda(x)$ for all $x \geq b_0$ and $\beta_0 \in (0, 1)$. Applying Pitman's criterion (Proposition 1) and the fact that (by Lemma 1 in particular $\lambda(x) = O(1)$ for $x \geq b_0$) it suffices to verify that

$$\int_{b_0}^{\infty} \exp(x\lambda(x) - \Lambda(x)) dx < \infty.$$

Nevertheless, combining B1 and B2 we have that there exists $c \in (0, \infty)$ such that

$$\int_{b_0}^{\infty} \exp(x\lambda(x) - \Lambda(x)) dx \leq \int_{b_0}^{\infty} e^{(\beta_0-1)\Lambda(x)} dx \leq c \int_{b_0}^{\infty} x^{-2} dx < \infty$$

and we conclude the lemma.

For the subexponentiality of the integrated tail, it is sufficient to show that

$$\limsup_{x \rightarrow \infty} \frac{x\bar{F}(x)}{-G(x) \log G(x)} < 1,$$

and apply the same analysis for the subexponentiality of \bar{F} . By L'Hopital's rule (possibly on a subsequence),

$$\limsup_{x \rightarrow \infty} \frac{x\bar{F}(x)}{-G(x) \log G(x)} \leq \limsup_{x \rightarrow \infty} -\frac{x\lambda(x) - 1}{1 + \log G(x)} \leq \limsup_{x \rightarrow \infty} -\frac{x\lambda(x) - 1}{\log \varepsilon + \log x - \Lambda(x)} \leq \beta_0$$

The second inequality is due to Lemma 3. The last inequality is from the fact that $\log x = o(\Lambda(x))$ and Assumptions B1 and B2. $\bar{F}(x)/G(x)$ and $-\log G(x)$ are the hazard function and cumulative hazard function of the integrated tail. The proof is completely analogous and therefore is omitted. ■

Proof of Lemma 9. Let $\beta \in (0, 1)$ be given and fix $\sigma_1 > 0$ small. Note that there exists a $\delta > 0$ such that for all $\sigma_1 \leq x \leq 1 - \sigma_1$

$$x^\beta + (1 - x)^\beta \geq 1 + \delta.$$

So, by continuity and with σ_1 small enough, we can find $\sigma_2 > 0$ small enough so that

$$x^\beta + (1 - x - \sigma_1/2)^\beta \geq 1 + \sigma_2.$$

Therefore, we know that we can select

$$a_j = a_{j-1} + \sigma_1/2,$$

as long as $\sigma_1/2 \leq a_{j-1} \leq 1 - \sigma_1/2$. Now select $k = \lceil 2(1 - \sigma_1)/\sigma_1 \rceil$ and we have $a_k \geq 1 - \sigma_1/2$. ■

B Technical proofs in Section 6

Proof of Lemma 11. First it is straightforward to verify (50) out of definition (49). By integrating both sides of (50), it is also immediate to see

$$\int_{-\infty}^{\infty} q_s^*(x) dx = 1.$$

Now, we just need to verify that if $p(s) < 1$ then $(1 - p(s))q_s^*(x) \geq 0$. We concentrate on the case in which Assumption B prevails (if Assumption A is in force the arguments carry

over in very similar forms). When $b - s > \eta_*$, using the definition of $q_s(x)$ given in Section 2.1 we obtain

$$\begin{aligned}
q_s(x) &= p_* f(x) \frac{I(x \leq c_0)}{P(X \leq c_0)} + p_{**} f_{**}(x|s) + \sum_{j=1}^k p_j f_j(x|s) \\
&= \frac{p_* f(x)}{P(X \leq c_0)} \\
&+ \frac{p_{**} f(x) I(x > c_k)}{P(X > c_k)} - \frac{p_* f(x) I(x > c_k)}{P(X \leq c_0)} \\
&+ \frac{p_k f(b-s-x) I(x \in (c_{k-1}, c_k])}{P(X \in (b-s-c_k, b-s-c_{k-1}])} - \frac{p_* f(x) I(x \in (c_{k-1}, c_k])}{P(X \leq c_0)} \\
&+ \sum_{j=1}^{k-1} \left(\frac{p_j f(x) I(x \in (c_{j-1}, c_j])}{P(X \in (c_{j-1}, c_j])} - \frac{p_* f(x) I(x \in (c_{j-1}, c_j])}{P(X \leq c_0)} \right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
(1-p(s))q_s^*(x) &= \frac{p_{**} f(x) I(x > c_k)}{P(X > c_k)} - \frac{p_* f(x) I(x > c_k)}{P(X \leq c_0)} \\
&+ \frac{p_k f(b-s-x) I(x \in (c_{k-1}, c_k])}{P(X \in (b-s-c_k, b-s-c_{k-1}])} - \frac{p_* f(x) I(x \in (c_{k-1}, c_k])}{P(X \leq c_0)} \\
&+ \sum_{j=1}^{k-1} \left(\frac{p_j f(x) I(x \in (c_{j-1}, c_j])}{P(X \in (c_{j-1}, c_j])} - \frac{p_* f(x) I(x \in (c_{j-1}, c_j])}{P(X \leq c_0)} \right). \quad (53)
\end{aligned}$$

To verify that $(1-p(s))q_s^*(x) \geq 0$, the most interesting part involves the second line in the above display corresponding to the interval $x \in (c_{k-1}, c_k]$. The reasoning for the rest of the pieces is similar and therefore is omitted. On the interval $(c_{k-1}, c_k]$ we have that $b-s-x \leq x$ assuming that $b-s \geq \eta_*$ and η_* is sufficiently large. Since $f(\cdot)$ is eventually decreasing (a consequence of Assumption B3), then

$$f(b-s-x) \geq f(x),$$

when $x \in (c_{k-1}, c_k]$. Consequently

$$\begin{aligned}
&\frac{p_k f(b-s-x) I(x \in (c_{k-1}, c_k])}{P(X \in (b-s-c_k, b-s-c_{k-1}])} - \frac{p_* f(x) I(x \in (c_{k-1}, c_k])}{P(X \leq c_0)} \\
&\geq \frac{p_k f(x) I(x \in (c_{k-1}, c_k])}{P(X \in (b-s-c_k, b-s-c_{k-1}])} - \frac{p_* f(x) I(x \in (c_{k-1}, c_k])}{P(X \leq c_0)}.
\end{aligned}$$

Further, we have that $p_k = \varepsilon^2 p_{**}$ decreases to zero at most linearly in $(b-s)^{-1}$, whereas $P(X \in (b-s-c_k, b-s-c_{k-1}])$ goes to zero faster than any linear function of $(b-s)^{-1}$. Therefore, $(1-p(s))q_s^*(x) I(x \in (c_{k-1}, c_k]) \geq 0$. The remaining pieces in (53) are handled similarly. ■

Proof of Lemma 12. Note that

$$Q^*(N_b > kb) = E^{Q^*} \left(\prod_{j=0}^{\lceil kb \rceil} p(\tilde{S}_j) \right), \quad (54)$$

where $p(s)$ is defined in (48). In addition, for some $\varepsilon > 0$,

$$\begin{aligned} E^{Q^*} \left(\prod_{j=0}^{\lceil kb \rceil} p(\tilde{S}_j) \right) &\leq E \left(\prod_{j=0}^{\lceil kb \rceil} p(\tilde{S}_j) I(|\tilde{S}_j - \mu j| \leq \varepsilon \max\{j, b\}) \right) \\ &\quad + Q^* \left(\sup_{j=1}^{\lceil kb \rceil} |\tilde{S}_j - \mu j| - \varepsilon \max\{j, b\} > 0 \right). \end{aligned} \quad (55)$$

Notice that for any $\varepsilon > 0$,

$$\lim_{b \rightarrow \infty} Q^* \left(\sup_{j=1}^{\lceil kb \rceil} [|\tilde{S}_j - \mu j| - \varepsilon \max\{j, b\}] > 0 \right) = 0.$$

Then, for some K sufficiently large (using an argument similar to that given in the proof of Proposition 9) we conclude

$$E \left(\prod_{j=0}^{\lceil kb \rceil} p(\tilde{S}_j) I(|\tilde{S}_j - \mu j| \leq \varepsilon \max\{j, b\}) \right) \leq Kk^{-\varepsilon_0}.$$

for some ε_0 small enough. Thereby, we conclude the proof applying the previous two estimates into (54) and (55). ■

Proof of Lemma 13. Let

$$\int_{b-s}^{\infty} q_s^*(x) dx = R(s).$$

Note that for $b - s > \eta_*$ we have that

$$R(s) = O(\varepsilon) + e^{-a^{**}}. \quad (56)$$

Let

$$\tau'_b = \inf\{n \geq 1 : \tilde{S}_n \geq b\}.$$

Now observe that

$$\begin{aligned} Q^*(\tau_b = N_b) &= \sum_{k=1}^{\infty} Q^*(N_b = k, \hat{S}_k > b, \tau_b > k - 1) \\ &\geq \sum_{k=1}^{\infty} Q^*(N_b = k, \hat{S}_k > b, \tau'_{b-\eta_*} > k - 1). \end{aligned}$$

Because of (56) we obtain that

$$\begin{aligned}
& \sum_{k=1}^{\infty} Q^*(N_b = k, \hat{S}_k > b, \tau'_{b-\eta_*} > k-1) \\
& \geq (O(\varepsilon) + e^{-a_{**}}) \sum_{k=1}^{\infty} Q^*(N_b = k, \tau'_{b-\eta_*} > k-1) \\
& = (O(\varepsilon) + e^{-a_{**}}) Q^*(\tau'_{b-\eta_*} > N_b - 1, N_b < \infty) \\
& \geq (O(\varepsilon) + e^{-a_{**}} + o(1)) Q^*(\tau'_{b-\eta_*} = \infty).
\end{aligned}$$

The term $o(1) \rightarrow 0$ as $b \rightarrow \infty$ comes from Lemma 12 which shows that $Q^*(N_b = \infty) = o(1)$ as $b \rightarrow \infty$. Finally, we observe

$$Q^*(\tau'_{b-\eta_*} = \infty) = 1 - u(b - \eta_*) \rightarrow 1,$$

as $b \rightarrow \infty$. The conclusion of this lemma follows. ■

Proof of Proposition 9. For $\delta_b = 1/\log b$, define

$$B_b = \{\tilde{S} : |\tilde{S}_j - j\mu| \leq \max(\delta_b^{-1}, \delta_b j), 1 \leq j \leq ta(b)\}.$$

It is clear that $\lim_{b \rightarrow \infty} P(B_b) = 1$.

If F is regularly varying, note that for all $\tilde{S} \in B_b$

$$Q^*(N_b > ta(b) | \tilde{S}) = E^{Q^*} \prod_{j=0}^{\lfloor ta(b) \rfloor} \left(1 - 2\theta \frac{\bar{F}(b - \tilde{S}_j)}{G(b - \tilde{S}_j)} \right) = (1+o(1)) \exp \left\{ - \sum_{j=0}^{\lfloor ta(b) \rfloor} 2\theta \frac{\bar{F}(b + j|\mu|)}{G(b + j|\mu|)} \right\}.$$

By Karamata's theorem we have that

$$\sum_{j=0}^{\lfloor ta(b) \rfloor} 2\theta \frac{\bar{F}(b + j|\mu|)}{G(b + j|\mu|)} \rightarrow \frac{2\theta(\iota - 1)}{|\mu|} \log \left(1 + \frac{|\mu|t}{\iota - 1} \right).$$

If Assumptions B1-B4 hold, We clearly have that

$$\begin{aligned}
a(x) &= \frac{G(x)}{\bar{F}(x)} \\
&= \int_0^{\infty} P(X > x + t | X > x) dt \\
&= \frac{1}{\lambda(x)} \int_0^{\infty} P(X > x + t/\lambda(x) | X > x) dt.
\end{aligned} \tag{57}$$

Now we can invoke Assumption B4 together with the dominated convergence theorem to conclude that

$$\int_0^{\infty} P(X > x + t/\lambda(x) | X > x) dt \longrightarrow \int_0^{\infty} \exp(-t) dt = 1$$

as $x \rightarrow \infty$. In addition, by the fundamental theorem of calculus we have that

$$\Lambda(x + y/\lambda(x)) - \Lambda(x) = \frac{y}{\lambda(x)} \int_0^1 \lambda(x + yu/\lambda(x)) du$$

and, in view of this representation, Assumption B4 is equivalent to stating that for each $K \in (0, \infty)$

$$\lim_{x \rightarrow \infty} \sup_{0 \leq y \leq K} \left| \int_0^{y/\lambda(x)} \lambda(x + z) dz - y \right| = \lim_{x \rightarrow \infty} \sup_{0 \leq y \leq K} \left| \int_0^{ya(x)} \frac{1}{a(x+z)} dz - y \right| = 0. \quad (58)$$

Observe that, since $\lambda(\cdot)$ is eventually non-increasing,

$$\sum_{j=0}^{\lfloor t/\lambda(b) \rfloor} \lambda(b + (j+1)|\mu|) \leq \int_0^{t/\lambda(b)} \lambda(b+x|\mu|) dx \leq \sum_{j=0}^{\lfloor t/\lambda(b) \rfloor} \lambda(b+j|\mu|).$$

We then conclude that

$$0 \leq \frac{1}{|\mu|} \int_0^{t/\lambda(b)} \lambda(b+x|\mu|) dx - \sum_{j=0}^{\lfloor t/\lambda(b) \rfloor} \lambda(b+(j+1)|\mu|) \leq \lambda(b) \rightarrow 0$$

as $b \nearrow \infty$. Therefore, applying (57) and (58) we conclude that

$$\lim_{b \rightarrow \infty} \sum_{j=0}^{\lfloor ta(b) \rfloor} \frac{\bar{F}(b+j|\mu|)}{G(b+j|\mu|)} = \lim_{b \rightarrow \infty} \int_0^{t/\lambda(b)} \lambda(b+x|\mu|) dx = t$$

as $b \nearrow \infty$ and consequently we have that for all $\tilde{S} \in B_b$

$$Q^*(N_b > ta(b)|\tilde{S}) = (1 + o(1)) \exp \left\{ - \sum_{j=0}^{\lfloor ta(b) \rfloor} 2\theta \frac{\bar{F}(b+j|\mu|)}{G(b+j|\mu|)} \right\}.$$

We then conclude that

$$\lim_{b \rightarrow \infty} Q^*(N_b > ta(b)|\tilde{S}) = e^{-2\theta t}.$$

■

References

- [1] S. Asmussen. Conditioned limit theorems relating a random walk to its associate, with applications to risk reserve processes and the gi/g/1 queue. *Adv. Appl. Prob.*, 14:143–170, 1982.
- [2] S. Asmussen. *Applied Probability and Queues*. Springer-Verlag, New York, 2003.
- [3] S. Asmussen, K. Binswanger, and B. Hojgaard. Rare events simulation for heavy-tailed distributions. *Bernoulli*, 6:303–322, 2000.

- [4] S. Asmussen and P. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer-Verlag, New York, NY, USA, 2008.
- [5] S. Asmussen and C. Kluppelberg. Large deviations results for subexponential tails, with applications to insurance risk. *Stochastic Processes and their Applications*, 64:103–125, 1996.
- [6] S. Asmussen and D. Kroese. Improved algorithms for rare event simulation with heavy tails. *Advances in Applied Probability*, 38:545–558, 2006.
- [7] S. Asmussen and R. Rubinstein. Steady state rare events simulation in queueing models and its complexity properties. In J. Dshalalow, editor, *Advances in Queueing*, pages 429–461. CRC Press, 1995.
- [8] Soren Asmussen and Claudia Kluppelberg. Large deviations results for subexponential tails, with applications to insurance risk. *Stochastic Processes and their Applications*, 64(1):103 – 125, 1996.
- [9] J. Blanchet and P. Glynn. Efficient rare event simulation for the maximum of heavy-tailed random walks. *to appear in Annals of Applied Probability*, 2007.
- [10] J. Blanchet, P. Glynn, and J.C. Liu. Efficient rare event simulation for multiserver queues. preprint. 2007.
- [11] J. Blanchet and C. Li. Efficient rare-event simulation for heavy-tailed compound sums. *Submitted to ACM Transactions on Computer Simulation*.
- [12] J. H. Blanchet and J. C. Liu. Efficient importance sampling in ruin problems for multidimensional regularly varying random walks. *Journal of Applied Probability*.
- [13] J. H. Blanchet and J. C. Liu. State-dependent importance sampling for regularly varying random walks. *Advances in Applied Probability*, 40:1104–1128, 2008.
- [14] J.H. Blanchet, P. Glynn, and J.C. Liu. Fluid heuristics, lyapunov bounds and efficient importance sampling for a heavy-tailed $g/g/1$ queue. *Queueing Syst.*, 57(2-3):99–113, 2007.
- [15] A. A. Borovkov and A. A. Mogul’skii. Integro-local and integral theorems for sums of random variables with semiexponential distributions. *Siberian Mathematical Journal*, 47(6):990–1026, 2006.
- [16] A. Dembo and O. Zeitouni. *Large deviations techniques and applications, 2nd edition*. Springer, New York, NY, USA, 1998.
- [17] P. Dupuis, K. Leder, and H. Wang. Importance sampling for sums of random variables with regularly varying tails. *ACM TOMACS*, 17, 2006.
- [18] P. Dupuis, K. Leder, and H. Wang. Importance sampling for sums of random variables with regularly varying tails. *ACM Trans. Model. Comput. Simul.*, 17(3):14, 2007.

- [19] P. Dupuis and H. Wang. Importance sampling for jackson networks. *Preprint*, 2008.
- [20] P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, New York, 1997.
- [21] H. Hult, F. Lindskog, T. Mikosch, and G. Samorodnitsky. Functional large deviations for multivariate regularly varying random walks. *Annals of Applied Probability*, 15(4):2651–2680, 2005.
- [22] S.P. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Springer-Verlag, London.
- [23] D. Siegmund. Importance sampling in the Monte Carlo study of sequential tests. *Ann. Stat.*, 4:673–684, 1976.