

Rare-event simulation for stochastic recurrence equations with heavy-tailed innovations

JOSE BLANCHET, Columbia University
 HENRIK HULT, Royal Institute of Technology
 KEVIN LEDER, University of Minnesota

In this paper rare event simulation for stochastic recurrence equations of the form

$$X_{n+1} = A_{n+1}X_n + B_{n+1}, \quad X_0 = 0,$$

is studied, where $\{A_n; n \geq 1\}$ and $\{B_n; n \geq 1\}$ are independent sequences consisting of independent and identically distributed real-valued random variables. It is assumed that the tail of the distribution of B_1 is regularly varying whereas the distribution of A_1 has a suitably light tail. The problem of efficient estimation, via simulation, of quantities such as $P\{X_n > b\}$ and $P\{\sup_{k \leq n} X_k > b\}$ for large b and n is studied. Importance sampling strategies are investigated, that provide unbiased estimators with bounded relative error as b, n tend to infinity.

Categories and Subject Descriptors: I.6.1 [Simulation and Modeling]: Simulation Theory; G.3 [Mathematics of Computing]: Probability and Statistics *Probabilistic algorithms (including Monte Carlo)*; I.6.0 [Simulation and Modeling]: General

General Terms: Algorithms, Performance, Theory

Additional Key Words and Phrases: Importance sampling, stochastic recurrence equations, heavy-tails

ACM Reference Format:

Jose Blanchet, Henrik Hult, and Kevin Leder, 2013. Rare-event simulation for stochastic recurrence equations with heavy-tailed innovations. *ACM Trans. Embedd. Comput. Syst.* 9, 4, Article 39 (March 2010), 25 pages.

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

1. INTRODUCTION

A goal of rare-event simulation is to design so-called strongly efficient estimators for probabilities that become increasingly rare in a meaningful asymptotic regime. An estimator is strongly efficient if a single replication of such estimator is unbiased and its coefficient of variation remains uniformly bounded even if the probability of interest becomes increasingly rare; see [Blanchet and Lam 2011] and references therein for additional notions on rare-event simulation.

Our focus here, as we shall precisely explain, is on rare event simulation for stochastic recurrence equations with heavy-tailed stochastic recursive equations. The paper by Asmussen et al. [Asmussen et al. 2000] articulates nicely the problems that arise

Blanchet's research was supported by NSF grants CMMI-0846816, CMMI 1069064. Hult's research was supported by Göran Gustafsson Foundation. Leder's research was supported by CNS-0435060, CCR-0325197, and EN-CS-0329609.

Author's addresses: J. Blanchet, Department of Industrial Engineering and Operations Research, Columbia University; H. Hult, Department of Mathematics, Royal Institute of Technology; K. Leder, Department of Industrial and Systems Engineering, University of Minnesota

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2010 ACM 1539-9087/2010/03-ART39 \$15.00

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

in trying to apply importance sampling in heavy-tailed settings. To quickly get to the heart of the problem, recall that if one wishes to estimate the probability of some event A , namely $P(A) > 0$, the conditional distribution $P(\cdot | A)$ coincides with the zero-variance importance sampling estimator of $P(A)$. Therefore, a natural approach when designing good importance sampling estimators is to approximate $P(\cdot | A)$ in some asymptotic regime. The main message of Asmussen et al. [Asmussen et al. 2000] is that importance sampling is difficult to apply in the context of heavy-tailed systems because the conditional distribution of such systems given a rare event of interest typically becomes singular with respect to the original distribution as the probability of interest tends to zero. In contrast, in light-tailed systems, such asymptotic conditional distribution can often be described by an exponential change of measure and therefore absolute continuity is guaranteed, leading to a good starting point for the design of importance sampling estimators.

Our goal in this paper is the theoretical study of sampling strategies that serve as a remedy to the situation discussed by Asmussen et al. for a class of Markov chains of the form

$$X_{n+1} = A_{n+1}X_n + B_{n+1}, \quad X_0 = 0, \quad (1)$$

where $\{A_n; n \geq 1\}$ and $\{B_n; n \geq 1\}$ are independent sequences consisting of independent and identically distributed random variables. It is assumed that the tail of the distribution of B_1 is regularly varying tail with index $-\alpha \leq 0$, that is,

$$\lim_{t \rightarrow \infty} \frac{P\{B_1 > tx\}}{P\{B_1 > t\}} = x^{-\alpha}, \quad x > 0,$$

and that $A_1 \geq 0$ satisfies the moment condition

$$E[A_1^{2\alpha+\epsilon}] < \infty, \quad (2)$$

for some $\epsilon > 0$.

The properties of the solution to a stochastic recurrence equation (1) have been thoroughly studied, see e.g. [Kesten 1973; Goldie 1991; Roitershtein 2007]. They have been used in a number of applications such as financial time series, in particular in the analysis of GARCH processes, see [Basrak et al. 2002], and population biology, see e.g. [Lewontin and Cohen 1969; Tuljapurkar 1990]. They also appear in the context of actuarial risk theory with investments. In the actuarial context, let U_k denote the reserve of an insurance company at time k and R_k and Z_k denote the return on investments and the net payments, respectively, from time $k-1$ to k . Then, $\{U_k\}$ satisfies the recursion

$$U_k = R_k(U_{k-1} - Z_k), \quad U_0 = b.$$

With $B_k = Z_{n-k+1}$ and $A_k = 1/R_{n-k+1}$ it follows that the value X_k of the reserve at time n discounted to time $n-k$ satisfies the recursion

$$\begin{aligned} X_0 &= U_n, \\ X_k &= \frac{1}{R_{n-k+1}}X_{k-1} + Z_{n-k+1} = A_kX_{k-1} + B_k, \quad k = 1, \dots, n. \end{aligned}$$

It follows that the ruin event $\{U_n < 0\}$ when starting the initial capital $U_0 = b$ is equal to the event $\{X_n > b\}$ when starting from $X_0 = 0$. The assumptions made in this paper, where the innovations B_1, \dots, B_n are heavy-tailed and the multiplicative factors A_n are sufficiently light-tailed, corresponds to the case where the insurance risk is heavy-tailed and dominates the asymptotic behavior over the financial risk. We refer to [Nyrhinen 1999; Gaier and Grandits 2002] and the overview by Paulsen [Paulsen

2008] for more details on the connection between stochastic recurrence equations and actuarial risk processes. Under our assumptions, large deviation asymptotics of rare-event probabilities have been studied in [Kostantinides and Mikosch 2005] and [Hult and Samorodnitsky 2010]. In this paper the goal is to obtain more precise estimates of rare-event probabilities by constructing efficient simulation algorithms to compute them.

The two problems considered here are the computation of the exceedance probabilities of the form:

$$p_b = P \{X_n > b | X_0 = 0\} \quad (3)$$

$$q_b = P \left\{ \sup_{k \leq n(b)} X_k > b | X_0 = 0 \right\}. \quad (4)$$

Efficiency is quantified in an asymptotic sense as $b \rightarrow \infty$ and $n(b)$ might increase to ∞ with b as long as $n(b) P(B_1 > b) = o(1)$ as $b \rightarrow \infty$. In particular, if $\alpha > 1$ we can pick $n(b) = O(b)$.¹

A number of approaches that deal with the problem discussed in [Asmussen et al. 2000] have been introduced in recent years; we refer the reader to the survey [Blanchet and Lam 2011] for a discussion of these approaches and focus here on literature that is most related to our problem. The bulk of the literature on efficient heavy-tailed simulation concentrates on regularly-varying random walk problems in a finite horizon. In our context this corresponds to $A_j = 1$ for $1 \leq j \leq n$ and $n = n(b) = O(1)$. The first provably efficient estimator of p_n , based on conditional Monte Carlo, was introduced by Asmussen and Binswanger [Asmussen and Binswanger 1997]. Conditional Monte Carlo has the advantage that it guarantees variance reduction, even if the probability of interest is not small. Nevertheless, since Conditional Monte Carlo requires the explicit evaluation of conditional expectations it is difficult to apply if one is interested in, say, estimating conditional expectations given events such as those indicated in (3) and (4) above. In contrast, other techniques such as importance sampling can be immediately applied to computing such conditional expectations. In addition, to guarantee efficiency in the estimator studied in [Asmussen and Binswanger 1997] it is important to have a finite horizon, that is $n(b) = O(1)$. The paper by Asmussen and Kroese [Asmussen and Kroese 2006] proposes refined conditional Monte Carlo algorithms which have been shown to achieve asymptotically zero relative error, see [Hartinger and Kortschak 2009]. Other algorithms, for instance those that appear in [Juneja and Shahabuddin 2002] and [Boots and Shahabuddin 2001] are based on importance sampling, applying hazard rate tilting. These algorithms also require n to be bounded and, although they are provably efficient beyond regularly varying distributions, only weak efficiency can be guaranteed; that is, the coefficient of variation grows at an subexponential rate relative to that of crude Monte Carlo. More recently, interesting methodology based on truncation and sequential importance resampling strategies have been proposed for standard random walk problems, see [Chan et al. 2012].

A useful family of sampling distributions based on mixtures and sequential sampling was introduced by Dupuis, Leder and Wang [Dupuis et al. 2008]. Their sequential sampling scheme applies only to regularly varying increments but it has a very

¹Given two non-negative functions $f(\cdot)$ and $g(\cdot)$, we say $f(n) = O(g(n))$ if there exists $c, n_0 \in (0, \infty)$ such that $f(n) \leq cg(n)$ all $n \geq n_0$, we say $f(n) = o(g(n))$ if given $\varepsilon > 0$ there exists $n_0 > 0$ such that $f(n)/g(n) < \varepsilon$ for all $n \geq n_0$, and we say $f(n) = \Omega(g(n))$ if there exists c, n_0 such that $cg(n) \leq f(n)$ for all $n \geq n_0$.

intuitive mixture, fully in agreement with the way in which extremes occur for heavy-tailed sums: due to the contribution of a single large increment that causes the rare event. At time $k \in \{0, 1, \dots, n-1\}$ one samples the next increment, B_{k+1} given that $B_{k+1} > a(b - X_k)$ for some $a \in (0, 1)$ with some probability $p(k, X_k) \in (0, 1)$, and with probability $1 - p(k, X_k)$ no importance sampling is applied. The selection of $a \in (0, 1)$ is needed in their analysis, which is based on a weak convergence argument, in order to guarantee that a certain likelihood ratio remains bounded. Intuitively, one needs such $a \in (0, 1)$ to control the likelihood ratio on paths that take more than one large jump to reach the level b . Those paths are negligible in the large deviations analysis of regularly varying sums, but they turn out to be important for the variance control of importance sampling estimators.

In the analysis of [Dupuis et al. 2008] it is important to let $n(b) = O(1)$ because this allows one to obtain a bounded likelihood ratio as $b \rightarrow \infty$. A verification technique, based on Lyapunov inequalities introduced in [Blanchet and Glynn 2008], allows the estimation of rare-event probabilities that involve an unbounded number of increments and also rare events that involve more than one large jump for their occurrence [Blanchet and Liu 2010; Blanchet et al. 2011]. As it turns out, this technique is closely related to the so-called subsolutions approach introduced by Dupuis and Wang [Dupuis and Wang. 2004] in the setting of light-tailed random variables. A thorough discussion on the relationship between Lyapunov inequalities and subsolutions is given in [Blanchet et al. 2012].

Our contributions can be put in perspective as follows:

(i) We consider an extension of the classical random walk model, which requires us to extend and adapt the existing algorithms for level crossings of heavy-tailed random walks to the more complicated setting of stochastic recurrence equations described by (1) and (2). In this model the rare event is most likely to be caused by a large B_j , so the idea is to proceed by simulating the A_j 's first. The first strategy is a sequential importance sampling algorithm based on conditional mixtures, similar to that of [Dupuis et al. 2009] for the random walk. At time k , one samples B_{k+1} according to a mixture involving a big jump and a regular increment. In contrast to [Dupuis et al. 2009] the correct mixture for B_{k+1} must not only involve the remaining distance $(b - X_k)$, as in the standard random walk case, but also the future impact that B_{k+1} has when combined with A_j for $j > k$. This is because the increment B_{k+1} can make the rare event occur at future times other than $k + 1$, leading to a strategy that turns off importance sampling at some specific times. This situation manifests itself not only in the selection of the correct mixture distribution, but also on the construction of the Lyapunov function, which is used to decide precisely when to turn off importance sampling. This contribution is summarized in Theorem 3.1, which shows that the importance sampling strategy achieves arbitrarily small relative error for weighted random walks (defined in Section 3), and Theorem 4.1 which shows that the strategy is strongly efficient for estimating p_b in (3).

(ii) A mixture-based importance sampling procedure is presented which yields a distribution for the X_n 's that is not Markovian. The new procedure is based on a variation of a mixture sampling strategy called Target Bridge Sampling (TBS) which has previously been applied to light-tailed settings such as Gaussian processes (see [Blanchet and Li 2011; Adler et al. 2012]). This approach allows us to construct strongly efficient estimators in situations where $n(b) \rightarrow \infty$ as $b \rightarrow \infty$. It has the advantage that it is easy to analyze as it bypasses the problem of finding a suitable Lyapunov function to turn off sequential importance sampling. It also provides a method for the generation of exact conditional samples given the event of interest. The disadvantage is that the cost per replication, in terms of function evaluations of the distribution of B_n , is quadratic in $n(b)$ as opposed to linear as is typically the case in sequential procedures. This con-

tribution is summarized in Theorem 5.1, which shows that the variation of TBS yields a strongly efficient estimator for q_b in (4).

(iii) As an alternative to the conditional mixture algorithm, in which the B_k 's must be sampled conditional on being large, Hult and Svensson [Hult and Svensson 2012] introduced a scaling technique that avoids such conditional sampling. We show in the on-line supplement that this technique can also be rigorously proven to be strongly efficient using the Lyapunov inequalities developed in (i). Thus, a new tool is provided, based on the combination of the scaling technique and Lyapunov inequalities, that avoids the need for sampling from the conditional distribution of B and yet preserves strong efficiency for p_b in 3 above. This contribution is summarized in Theorem 1 of the on-line supplement.

The rest of the paper is organized as follows. In Section 2 the connection between importance sampling and dynamic programming is reviewed. This connection is important to understand the nature of our algorithm which requires the use of Lyapunov functions which, in the end, are bounds on the value function of a dynamic programming recursion. In Section 3 the finite horizon rare-event estimation problem is studied for a modified random walk model; this step is necessary as it serves as an intermediate step between the standard random walk model and the affine Markov chain. Section 4 builds on the ideas explained in Section 3 and provides results on the construction of efficient simulation algorithms for estimating $p_b = P\{X_n > b\}$. The concluding Section 5 presents an algorithm based on target bridge sampling for the efficient estimation of $q_b = P\{\sup_{k \leq n(b)} X_k > b\}$. The concluding Section 6 contains the results on numerical experiments.

2. IMPORTANCE SAMPLING AND DYNAMIC PROGRAMMING

Let $\{X_k; k \geq 1\}$, $X_0 = 0$, be a Markov chain with transition density $f_{k|k-1}(x_k | x_{k-1})$. The objective is to compute a probability of the form $p_b = P\{X_n \in A_b\}$ for some $b > 0$ under the assumption that $\lim_{b \rightarrow \infty} p_b = 0$.

An importance sampling estimator is constructed by sampling N independent and identically distributed copies of $\mathbf{X} = (X_1, \dots, X_n)$, which will be denoted by $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)}$, using transition densities $g_{k|k-1}(x_k | x_{k-1})$. For each copy the quantity

$$\hat{p}_b^{(j)} = I\{X_n^{(j)} \in A_b\} \prod_{k=1}^n \frac{f_{k|k-1}(X_k^{(j)} | X_{k-1}^{(j)})}{g_{k|k-1}(X_k^{(j)} | X_{k-1}^{(j)})} \quad (5)$$

is computed. An unbiased estimator \hat{p}_b for p_b is then given by averaging N independent copies of $\hat{p}_b^{(1)}$:

$$\hat{p}_b = \frac{1}{N} \sum_{j=1}^N \hat{p}_b^{(j)}.$$

A common performance demand of a Monte Carlo estimate is that for $\varepsilon > 0, \delta > 0$ one wants to choose an N such that $P\{|\hat{p}_b - p_b| > p_b \varepsilon\} \leq \delta$. An analysis based on Chebyshev's inequality would imply that one then needs to choose

$$N > \frac{1}{\varepsilon^2 \delta} \left(\frac{\tilde{E}[(\hat{p}_b^{(1)})^2]}{p_b^2} - 1 \right) = \frac{1}{\varepsilon^2 \delta} RE(\hat{p}_b^{(1)}),$$

where RE denotes the relative error of the estimator. Therefore, if we establish that

$$\sup_b \frac{\tilde{E}[(\hat{p}_b^{(1)})^2]}{p_b^2} < \infty, \quad (6)$$

then it follows that the number of Monte Carlo replications, N , needed to achieve a given level of accuracy is bounded as $b \rightarrow \infty$. Here \tilde{E} refers to the expectation under the law induced by the transition densities $g_{k|k-1}$. An importance sampling estimator that satisfies (6) is said to have bounded relative error, this property is also known as strong efficiency. While strong efficiency is a useful criteria it should be noted that there are alternative metrics (often based on higher moments) by which to judge the performance of importance sampling estimators, see e.g. [L'ecuyer et al. 2010].

Using the identity

$$\tilde{E}[(\hat{p}_b^{(1)})^2] = E[\hat{p}_b^{(1)}] = E\left[\prod_{k=1}^n \frac{f_{k|k-1}(X_k | X_{k-1})}{g_{k|k-1}(X_k | X_{k-1})} I\{X_n \in A_b\} \mid X_0 = 0\right],$$

it is clear that if we wish to minimize the variance of our importance sampling estimator then we must choose the transition densities that return the minimum

$$V_b(0, 0) = \inf_{\{g_{k|k-1}; k \geq 1\}} E\left[\prod_{k=1}^n \frac{f_{k|k-1}(X_k | X_{k-1})}{g_{k|k-1}(X_k | X_{k-1})} I\{X_n \in A_b\} \mid X_0 = 0\right].$$

The value $V_b(0, 0)$ can be interpreted as the initial value of a control problem where the control variables are the sampling densities $\{g_{k|k-1}; k = 1, \dots, n\}$. The associated value function is given by

$$V_b(j, x) = \inf_{\{g_{k|k-1}; k \geq j+1\}} E\left[\prod_{k=j+1}^n \frac{f_{k|k-1}(X_k | X_{k-1})}{g_{k|k-1}(X_k | X_{k-1})} I\{X_n \in A_b\} \mid X_j = x\right].$$

Then $V_b(j, x)$ satisfies the dynamic programming equation

$$V_b(j, x) = \inf_{g_{j+1|j}} E\left[\frac{f_{j+1|j}(X_{j+1} | X_j)}{g_{j+1|j}(X_{j+1} | X_j)} V_b(j+1, X_{j+1}) I\{X_n > b\} \mid X_j = x\right]$$

and the optimum $V_b(0, 0) = p_b^2$ is achieved by selecting $g_{k|k-1}(\cdot | X_{k-1})$ as the conditional density of X_k given X_{k-1} and $X_n \in A_b$. Since this choice requires either knowledge of the probability of the conditioning event or that the conditioning event is not rare it is not feasible from a simulation standpoint. Instead of searching for a solution to the dynamic programming equation it suffices to find (parameterized) sampling densities $g_{k|k-1}(x_k | x_{k-1})$ and a function $\hat{V}_b(j, x)$ such that

$$\hat{V}_b(j, x) \geq E\left[\frac{f_{j+1|j}(X_{j+1} | X_j)}{g_{j+1|j}(X_{j+1} | X_j)} \hat{V}_b(j+1, X_{j+1}) I\{X_n \in A_b\} \mid X_j = x\right], \quad (7)$$

$$\hat{V}_b(n, x) \geq I\{x \in A_b\}. \quad (8)$$

Then, by a standard verification argument, see Lemma 2.1 below, $\hat{V}_b(0, 0)$ is an upper bound of the second moment $\tilde{E}[(\hat{p}_b^{(1)})^2]$ of the importance sampling estimator. The goal is then to find parameterized sampling densities $g_{k|k-1}(x_k | x_{k-1})$ and $\hat{V}_b(j, x)$ satisfying (7) and (8) such that

$$\sup_{b>0} \frac{\hat{V}_b(0, 0)}{p_b^2} < \infty. \quad (9)$$

In that case it follows that $\hat{\rho}_b$ has bounded relative error in the sense of 6.

Note that our discussion does not make any specific mention of a parametric class where to search for appropriate transition densities $g_{k|k-1}(x_k | x_{k-1})$ and the function $\hat{V}_b(j, x)$ satisfying the functional inequalities (7) and (8). The construction of such a class is part of the problem and is guided by knowledge of how the rare event is most likely to occur. As discussed earlier, for heavy-tailed random walks it has been demonstrated that a useful parametric family is a mixture between the original dynamics and some version of the original dynamics forced to be near the threshold. This parametric family is studied in Section 3.

The inequalities (7) and (8) are commonly used in the rare-event simulation community to establish performance bounds. In the pre-limit they are usually referred to as Lyapunov inequalities and have been used in a wide variety of settings [Blanchet and Glynn 2008; Blanchet et al. 2011]. In this work we will follow this notation and call the functions \hat{V}_b Lyapunov functions and the inequalities (7, 8) Lyapunov inequalities. In the light-tailed setting, by properly scaling the inequalities, it is also possible to obtain a partial differential equation for the large b limit of a suitably normalized version of the value function, see [Dupuis and Wang. 2004; Dupuis et al. 2007; Dupuis et al. 2009]. Below we state the verification argument on the consequences of a relaxed version of the inequalities (7). For completeness we have included a proof.

LEMMA 2.1. *Suppose there are constants $\{\beta_j\}_{j=1}^n$ such that $\beta_j \geq 1$, and a function $\hat{V}_b(j, x)$ such that $\hat{V}_b(n, x) \geq I\{x \in A_b\}$ and for $j = 0, \dots, n-1$*

$$E \left[\frac{\hat{V}_b(j+1, X_{j+1})}{\hat{V}_b(j, x)} \frac{f_{j+1|j}(X_{j+1} | x)}{g_{j+1|j}(X_{j+1} | x)} \mid X_j = x \right] \leq \beta_{j+1}. \quad (10)$$

Then,

$$E \left[\prod_{j=0}^{n-1} \frac{f_{j+1|j}(X_{j+1} | X_j)}{g_{j+1|j}(X_{j+1} | X_j)} I\{X_n \in A_b\} \right] \leq \left(\prod_{j=0}^{n-1} \beta_{j+1} \right) \hat{V}_b(0, 0).$$

This is a standard result, see e.g. Lemma 3.1 of [Blanchet et al. 2012]

Note that in the lemma we consider loose Lyapunov inequalities. That is, we allow constants $\beta_j \geq 1$. As seen in the statement of the lemma, in the setting that $n < \infty$ the estimator based on the sampling densities $\{g_{j+1|j}\}$ will still have bounded relative error with loose Lyapunov inequalities. However when we consider the setting of n increasing to ∞ with b we will have to pay careful attention to the looseness of inequalities, i.e. the size of the $\{\beta_j\}$.

3. A WEIGHTED RANDOM WALK MODEL

Before studying the rare event simulation problem for the solution to a stochastic recurrence equation we consider rare events in a related and simpler model. Consider a random walk model where $\{B_k; k = 1, \dots, n\}$ are independent and identically distributed random variables and

$$X_{n+1} = X_n + c_{n+1}B_{n+1}, \quad X_0 = 0, \quad (11)$$

for some positive constants c_1, \dots, c_n . Suppose that, for some $\alpha > 0$, $P\{B_1 > b\}$ is regularly varying with index $-\alpha$, i.e., for all $t > 0$,

$$\lim_{b \rightarrow \infty} \frac{P\{B_1 > tb\}}{P\{B_1 > b\}} = t^{-\alpha}.$$

In addition denote the density of B_1 by f_B . Ultimately, as we will explain in the next section, in the setting of the stochastic recursion (1), the c_j 's will be chosen based on the value of the A_j 's.

It follows from the regularly varying property that

$$p_b = P\{X_n > b\} \sim \sum_{k=1}^n P\{B_k > b/c_k\}.$$

where $f(b) \sim g(b)$ is shorthand for $f(b)/g(b) \rightarrow 1$, as $b \rightarrow \infty$. The asymptotic approximation is based on the heavy-tailed heuristics that the most likely way that $X_n > b$ is that precisely one of the innovations B_1, \dots, B_n is large.

To design an importance sampling algorithm with bounded relative error it is sufficient to construct appropriate sampling densities $g_{k+1|k}$ as well as a Lyapunov function \hat{V}_b satisfying (10) with β_j bounded in b and $\sup_{b>0} \hat{V}_b(0,0)/p_b^2 < \infty$. To come up with a candidate Lyapunov function $\hat{V}_b(j,x)$ the basic intuition is to take it roughly proportional to an asymptotic equivalent of $P\{X_n > b \mid X_j = x\}^2$. For fixed $x < b$,

$$P\{X_n > b \mid X_j = x\}^2 \sim \left(\sum_{k=j+1}^n P\{B_k > (b-x)/c_k\} \right)^2,$$

as $b \rightarrow \infty$. This indicates that an appropriate choice may be to take

$$\hat{V}_b(j,x) = \begin{cases} \min \left\{ d_j \left(\sum_{k=j+1}^n P\{B > (b-x)/c_k\} \right)^2, 1 \right\}, & j \in \{1, \dots, n-1\}, \\ I\{x \geq b\}, & j = n, \end{cases} \quad (12)$$

for some constants d_j .

Assumption 1. In what follows it will be assumed that the constants d_j are chosen sufficiently large so that $\hat{V}_b(j,x) < 1$ implies $b-x > 0$.

Notice that this assumption is trivially satisfied when $j = n$, so we will only discuss it for $j \leq n-1$. Notice then that we may take $d_j > 1/P\{B_1 > 0\}^2$. To see this, note that if $V_b(j,x) < 1$ then

$$d_j \left(\sum_{k=j+1}^n P\{B > (b-x)/c_k\} \right)^2 < 1,$$

which is equivalent to

$$\sum_{k=j+1}^n P\{B > (b-x)/c_k\} < 1/\sqrt{d_j},$$

and by our choice of d_j this implies that

$$P\{B > (b-x)/c_k\} < 1/\sqrt{d_j} < P\{B_1 > 0\}, \quad k = j+1, \dots, n.$$

Since the constants c_k are positive it follows that $b-x > 0$.

With a candidate for the Lyapunov function in place, let us consider the choice of the sampling densities. In an attempt to mimic the heavy-tailed heuristic of one large shock it is quite natural to sample B_1, \dots, B_n , using mixtures. The mixture is such that with some probability a large shock is sampled whereas otherwise an average outcome is sampled. One method to generate this large value is to sample from the original distribution of the increments conditioned on being sufficiently large. Another

method is to sample a variate from the original distribution of the increments and then multiply it by a fixed large number. We will consider both possibilities.

The analysis and design of both algorithms will be guided by the Lyapunov function \hat{V}_b in (12). In particular we will establish the Lyapunov inequality of Lemma 2.1 using this function for both algorithms (the conditional mixture sampler and the scaling algorithm). In addition, the decision of how to sample on step $j + 1$ given $X_j = x$ (i.e. deciding if a big jump or a regular jump will be applied) will be based on the value $\hat{V}_b(j, x)$.

3.1. A conditional mixture importance sampling estimator

Sampling from the distribution of the increments conditioned on being large is a natural choice for inducing a large shock in the sampling distribution. As will be seen in this section, the analysis of such an algorithm follows quite naturally. We now describe the proposed family of sampling distributions for the conditional mixture algorithm.

Proposed family of importance sampling distributions for the conditional mixture algorithm. Consider the sampling density of B_{j+1} given $X_j = x$ as

$$\begin{aligned} g_{B_{j+1}}^C(y | x) &= f_B(y)I\{x \in \Gamma_{j+1}^c\} \\ &+ \frac{p(j+1, x)f_B(y)}{P\{B > a(b-x)/c_{j+1}\}}I\{y > a(b-x)/c_{j+1}\}I\{x \in \Gamma_{j+1}\} \\ &+ \frac{(1-p(j+1, x))f_B(y)}{P\{B \leq a(b-x)/c_{j+1}\}}I\{y \leq a(b-x)/c_{j+1}\}I\{x \in \Gamma_{j+1}\}. \end{aligned}$$

Here Γ_{j+1} is interpreted as the set of x -values where importance sampling is performed. It is defined by

$$\Gamma_j = \{x : \hat{V}_b(j-1, x) < 1\}. \quad (13)$$

If $X_j = x$ is already close to b , then the event $X_n > b$ is not considered to be rare and importance sampling is not needed. Therefore B_{j+1} is sampled according to its original density f_B if $x \in \Gamma_{j+1}^c$. If $x \in \Gamma_{j+1}$ then B_{j+1} is sampled according to the mixture distribution, which is the original distribution conditioned on not being too large (with probability $1 - p(j+1, x)$) and the original distribution conditioned on being large enough to nearly bring the sum near the threshold b (with probability $p(j+1, x)$). The phrase ‘nearly’ is present because of the parameter $a \in (0, 1)$ that serves as a cushion.

The efficiency analysis of the conditional mixture sampling estimator. We now establish that the importance sampling estimator based on the conditional mixture algorithm estimates the probability $P\{X_n > b\}$ with bounded relative error.

THEOREM 3.1. *Let $\{X_k; k = 0, \dots, n\}$ be the random walk model given in (11). The estimator $\hat{p}_b^{(1)}$ in (5), obtained from the conditional mixture algorithm with sampling densities given by $g_{B_{j+1}}^C$, has bounded relative error for computing $p_b = P\{X_n > b\}$. More precisely,*

$$\limsup_{b \rightarrow \infty} \frac{\tilde{E}[(\hat{p}_b^{(1)})^2]}{p_b^2} \leq \frac{1}{a^{2\alpha}P\{B_1 > 0\}}, \quad (14)$$

where the upper bound is obtained by taking $d_j = a^{-2\alpha}/P\{B_1 > 0\}$ and

$$p(j, x) = \frac{\sqrt{v_1(j)}}{\sqrt{v_1(j)} + \sqrt{v_2(j)}},$$

where

$$v_1(j) = P\{B_1 > 0\}c_j^{2\alpha} \left(\sum_{k=j}^n c_k^\alpha \right)^{-2} \text{ and } v_2(j) = \left(\frac{\sum_{k=j+1}^n c_k^\alpha}{\sum_{k=j}^n c_k^\alpha} \right)^2.$$

From (14) it is clear that if the random variables B_i are non-negative, i.e. $P\{B_1 > 0\} = 1$, then it is possible for the algorithm to have relative error arbitrarily close to 0, by choosing a sufficiently close to 1.

The proof is given in Section 3.2.

3.2. Proof of Theorem 3.1

For the proof of Theorem 3.1 we will need Potter's bound (Theorem 1.5.6 in [Bingham et al. 1987]), which we reproduce here for convenience.

THEOREM 3.2. *If f is regularly varying of index ρ and in addition f is bounded away from 0 and ∞ on compact subsets of $[0, \infty)$ then for any $\delta > 0$ there exists $A = A(\delta) > 1$ such that*

$$f(y)/f(x) \leq A \max\{(y/x)^{\rho+\delta}, (y/x)^{\rho-\delta}\}, \quad x, y > 0.$$

We now provide the proof of Theorem 3.1. We begin by showing that each independent copy of $\hat{p}_b^{(1)}$ of the estimator \hat{p}_b has bounded relative error. Let $\hat{V}_b(j, x)$ be as in (12). The second moment satisfies

$$\tilde{E}[(\hat{p}_b^{(j)})^2] = E \left[\prod_{k=1}^n \frac{f_{k|k-1}(X_k | X_{k-1})}{g_{k|k-1}(X_k | X_{k-1})} I\{X_n > b\} \mid X_0 = 0 \right].$$

Since $\limsup_{b \rightarrow \infty} \hat{V}_b(0, 0)/p_b^2 = d_0$, the bounded relative error follows from Lemma 2.1 if we can show that the Lyapunov inequality (10) holds with coefficients $\beta_j = \beta_j^C(b)$ that are bounded in b .

For the conditional mixture algorithm, the likelihood ratio appearing in (10) can be written as

$$\frac{f_{j+1|j}(X_{j+1} | x)}{g_{j+1|j}^C(X_{j+1} | x)} = \frac{f_B((X_{j+1} - x)/c_{j+1})}{g_{B_{j+1}}^C((X_{j+1} - x)/c_{j+1} | x)}.$$

Therefore it suffices to establish the following Lyapunov inequality.

PROPOSITION 3.3. *If*

$$\begin{aligned} \sup\{p(j, x) : x \in \mathbb{R}, j < n\} &< 1 \\ \inf\{p(j, x) : x \in \mathbb{R}, j \leq n\} &> 0 \end{aligned} \tag{15}$$

then there exists constants $\{\gamma_j^C\}_{j=1}^n$, greater or equal to 1 and bounded in b , such that, for $j = 0, \dots, n-1$ and $x \in \mathbb{R}$,

$$E \left[\frac{\hat{V}_b(j+1, X_{j+1})}{\hat{V}_b(j, x)} \frac{f_B((X_{j+1} - x)/c_{j+1})}{g_{B_{j+1}}^C((X_{j+1} - x)/c_{j+1} | x)} \mid X_j = x \right] \leq \gamma_{j+1}^C.$$

In addition for $0 \leq j \leq n-2$ and $x \in \Gamma_{j+1}$ it holds that

$$\limsup_{b \rightarrow \infty} E \left[\frac{\hat{V}_b(j+1, X_{j+1})}{\hat{V}_b(j, x)} \frac{f_{j+1|j}(X_{j+1} | x)}{g_{j+1|j}^C(X_{j+1} | x)} \mid X_j = x \right] \leq \tilde{\beta}_{j+1}^C,$$

where

$$\tilde{\beta}_{j+1}^C = \frac{d_j^{-1}}{p(j+1)} \left(\sum_{k=j+1}^n \left(\frac{ac_k}{c_{j+1}} \right)^\alpha \right)^{-2} + \frac{d_{j+1}d_j^{-1}}{(1-p(j+1))} \frac{\left(\sum_{k=j+1}^n \left(\frac{c_k}{c_{j+1}} \right)^\alpha - 1 \right)^2}{\left(\sum_{k=j+1}^n \left(\frac{c_k}{c_{j+1}} \right)^\alpha \right)^2}. \quad (16)$$

Proposition 3.3 completes the proof that the estimator has bounded relative error.

PROOF. To test the Lyapunov inequality, first consider the case $\hat{V}_b(j, x) = 1$, which, recalling the definition (13), corresponds to $x \in \Gamma_{j+1}^c$. Then, $g_{B_{j+1}}^C(y | x) = f_B(y | x)$ which implies that

$$E \left[\frac{\hat{V}_b(j+1, X_{j+1})}{\hat{V}_b(j, x)} \frac{f_B((X_{j+1} - x)/c_{j+1})}{g_{B_{j+1}}^C((X_{j+1} - x)/c_{j+1} | x)} \mid X_j = x \right] \leq 1,$$

since $\hat{V}_b(j+1, X_{j+1}) \leq 1$.

Consider now the case $\hat{V}_b(j, x) < 1$, which corresponds to $x \in \Gamma_{j+1}$. Put

$$\begin{aligned} J_1(j, x) &= E \left[\frac{\hat{V}_b(j+1, x + c_{j+1}B_{j+1})}{\hat{V}_b(j, x)p(j+1, x)} I\{c_{j+1}B_{j+1} > a(b-x)\} \right] \\ &\quad \times P\{c_{j+1}B_{j+1} > a(b-x)\}, \\ J_2(j, x) &= E \left[\frac{\hat{V}_b(j+1, x + c_{j+1}B_{j+1})}{\hat{V}_b(j, x)(1-p(j+1, x))} I\{c_{j+1}B_{j+1} \leq a(b-x)\} \right] \\ &\quad \times P\{c_{j+1}B_{j+1} \leq a(b-x)\}. \end{aligned}$$

Since $\hat{V}_b(j+1, X_{j+1}) \leq 1$ it follows that, for $j = 0, \dots, n-1$,

$$\begin{aligned} J_1(j, x) &\leq \frac{P\{c_{j+1}B_{j+1} > a(b-x)\}^2}{\hat{V}_b(j, x)p(j+1, x)} \\ &\leq \frac{P\{c_{j+1}B_{j+1} > a(b-x)\}^2}{d_j p(j+1, x) \left(\sum_{k=j+1}^n P\{c_k B_{j+1} > b-x\} \right)^2} \\ &= \frac{1}{d_j p(j+1, x)} \left(\sum_{k=j+1}^n \frac{P\{c_k B_{j+1} > b-x\}}{P\{c_{j+1}B_{j+1} > a(b-x)\}} \right)^{-2}. \end{aligned}$$

Since $\hat{V}_b(j, x) < 1$ it follows from the discussion after Assumption 1 that for $k = j+1, \dots, n$,

$$\frac{b-x}{c_k} \geq \bar{F}^{-1}(1/\sqrt{d_j}) > 0.$$

Potter's bound (Theorem 3.2), implies that, for each $\varepsilon > 0$ and $k = j+1, \dots, n$, there is a constant $K_{j+1, k}$ such that

$$\frac{P\{B > (b-x)/c_k\}}{P\{B > a(b-x)/c_{j+1}\}} \geq K_{j+1, k}^{-1} \min \left\{ \left(\frac{ac_k}{c_{j+1}} \right)^{\alpha+\varepsilon}, \left(\frac{ac_k}{c_{j+1}} \right)^{\alpha-\varepsilon} \right\}.$$

It follows that

$$\begin{aligned} J_1(j, x) &\leq \frac{1}{d_j p(j+1, x)} \left(\sum_{k=j+1}^n K_{j+1, k}^{-1} \min \left\{ \left(\frac{ac_k}{c_{j+1}} \right)^{\alpha+\epsilon}, \left(\frac{ac_k}{c_{j+1}} \right)^{\alpha-\epsilon} \right\} \right)^{-2} \\ &\leq \frac{1}{d_j p(j+1, x)} K_1^2 a^{-2(\alpha+\epsilon)}, \end{aligned} \quad (17)$$

where the second inequality follows because we drop non-negative summands from the summation and for ease of notation we write $K_{j+1, j+1} = K_1$. The second inequality is obviously less sharp, but has the advantage that it is independent of the constants $\{c_k\}$.

Now consider the term $J_2(j, x)$ for $j \in \{0, \dots, n-2\}$. Observe that

$$J_2(j, x) \leq \frac{1}{1-p(j+1, x)} E \left[\frac{\hat{V}_b(j+1, x + c_{j+1} B_{j+1})}{\hat{V}_b(j, x)} I_{\{c_{j+1} B_{j+1} \leq a(b-x)\}} \right].$$

For any number b_{j+1} such that $c_{j+1} b_{j+1} \leq a(b-x)$ it holds that $b-x-c_{j+1} b_{j+1} \geq (b-x)(1-a)$, which leads to

$$\begin{aligned} \hat{V}_b(j+1, x + c_{j+1} b_{j+1}) &\leq d_{j+1} \left(\sum_{k=j+2}^n P\{B > (b-x-c_{j+1} b_{j+1})/c_k\} \right)^2 \\ &\leq d_{j+1} \left(\sum_{k=j+2}^n P\{B > (b-x)(1-a)/c_k\} \right)^2. \end{aligned}$$

Recalling that $\hat{V}_b(j, x) < 1$ the previous display shows that on the event $c_{j+1} B_{j+1} \leq a(b-x)$,

$$\begin{aligned} \frac{\hat{V}_b(j+1, x + c_{j+1} B_{j+1})}{\hat{V}_b(j, x)} &\leq \frac{d_{j+1} \left(\sum_{k=j+2}^n P\{B > (b-x)(1-a)/c_k\} \right)^2}{d_j \left(\sum_{k=j+1}^n P\{B > (b-x)/c_k\} \right)^2} \\ &\leq \frac{d_{j+1} \left(\sum_{k=j+2}^n P\{B > (b-x)(1-a)/c_k\} \right)^2}{d_j \left(\sum_{k=j+2}^n P\{B > (b-x)/c_k\} \right)^2}. \end{aligned}$$

Since $\hat{V}_b(j, x) < 1$ it follows, as above, that $(b-x)/c_k \geq \bar{F}^{-1}(1/\sqrt{d_j})$. Then, Potter's bound implies that, for each $\epsilon > 0$, there is a constant K_2 such that

$$\frac{P\{B > (b-x)(1-a)/c_k\}}{P\{B > (b-x)/c_k\}} \leq K_2 (1-a)^{-\alpha-\epsilon}.$$

We conclude that

$$\frac{d_{j+1} \left(\sum_{k=j+2}^n P\{B > (b-x)(1-a)/c_k\} \right)^2}{d_j \left(\sum_{k=j+2}^n P\{B > (b-x)/c_k\} \right)^2} \leq \frac{d_{j+1}}{d_j} K_2^2 (1-a)^{-2(\alpha+\epsilon)} \quad (18)$$

and

$$J_2(j, x) \leq \frac{d_{j+1}}{d_j (1-p(j+1, x))} K_2^2 (1-a)^{-2(\alpha+\epsilon)}.$$

It remains to consider the case $j = n - 1$ and $x \in \Gamma_n$. Recall that if $x \in \Gamma_n$ then $b > x$, and observe that

$$\begin{aligned} J_2(n-1, x) &\leq \frac{1}{1-p(n, x)} E \left[\frac{\hat{V}_b(n, x + c_{j+1}B_{j+1})}{\hat{V}_b(n-1, x)} I\{c_{j+1}B_{j+1} \leq a(b-x)\} \right] \\ &= \frac{1}{1-p(n, x)} E \left[\frac{I\{x + c_{j+1}B_{j+1} \geq b, c_{j+1}B_{j+1} \leq a(b-x)\}}{\hat{V}_b(n-1, x)} \right]. \end{aligned}$$

The last expression is non-zero only if $x > b$. Thus, for $x \in \Gamma_n$, it follows that $J_2(n-1, x) = 0$. To summarize we have that, for $x \in \Gamma_{j+1}^c$,

$$E \left[\frac{\hat{V}_b(j+1, X_{j+1})}{\hat{V}_b(j, x)} \frac{f_{j+1|j}(X_{j+1} | x)}{g_{j+1|j}^C(X_{j+1} | x)} \middle| X_j = x \right] \leq 1,$$

whereas, for $x \in \Gamma_{j+1}$,

$$\begin{aligned} &E \left[\frac{\hat{V}_b(j+1, X_{j+1})}{\hat{V}_b(j, x)} \frac{f_{j+1|j}(X_{j+1} | x)}{g_{j+1|j}^C(X_{j+1} | x)} \middle| X_j = x \right] \\ &\leq J_1(j, x) + J_2(j, x) \\ &\leq \frac{1}{d_j p(j+1, x)} K_1^2 a^{-2(\alpha+\epsilon)} + \frac{d_{j+1}}{d_j(1-p(j+1, x))} K_2^2 (1-a)^{-2(\alpha+\epsilon)} I(j < n-1). \end{aligned}$$

Since we have assumed the uniform bound condition (15) we have that we can define the following constant

$$\tilde{\gamma}_{j+1}^C \doteq \frac{1}{d_j p(j+1, x)} K_1^2 a^{-2(\alpha+\epsilon)} + \frac{d_{j+1}}{d_j(1-p(j+1, x))} K_2^2 (1-a)^{-2(\alpha+\epsilon)} I(j < n-1).$$

Since the previous expression is independent of b we see that its supremum over b is trivially finite. The proof is completed by putting $\gamma_j^C = \max(\tilde{\gamma}_j^C, 1)$.

It remains to prove the asymptotic upper bound (16).

The expression for $\hat{\beta}_{j+1}^C$ in (16) consists of two terms where the first term is an asymptotic upper bound of the J_1 -term and the second term is an asymptotic upper bound of the J_2 -term, as $b \rightarrow \infty$, where the J_1 and J_2 terms are the expressions that appeared in the proof of Proposition 3.3.

The limit for the J_1 -term follows by inspection and application of the regularly variation property of the tail probabilities. The limit for the J_2 term is a bit more complicated. By conditioning on B_{j+1} it follows from the dominated convergence theorem, via (18), that it suffices to find the following limit for $j \leq n-2$,

$$\begin{aligned} &\lim_{b \rightarrow \infty} \frac{\hat{V}_b(j+1, x + c_{j+1}b_{j+1})}{\hat{V}_b(j, x)} I\{c_{j+1}b_{j+1} \leq a(b-x)\} \\ &= \lim_{b \rightarrow \infty} \frac{d_{j+1}}{d_j} \left(\frac{\sum_{k=j+2}^n P\{c_k B > b-x-c_{j+1}b_{j+1}\}}{\sum_{k=j+1}^n P\{c_k B > b-x\}} \right)^2 I\left\{b_{j+1} \leq \frac{a(b-x)}{c_{j+1}}\right\}. \end{aligned}$$

The term inside the parenthesis in the previous display can be rewritten as

$$\frac{\sum_{k=j+1}^n P\{c_k B > b-x-c_{j+1}b_{j+1}\}}{\sum_{k=j+1}^n P\{c_k B > b-x\}} - \frac{P\{c_{j+1}B > b-x-c_{j+1}b_{j+1}\}}{\sum_{k=j+1}^n P\{c_k B > b-x\}}.$$

Since the distribution of B is assumed to have regularly varying tails, it also has the long tail property, i.e.,

$$\lim_{b \rightarrow \infty} \frac{P\{c_k B > b - x - c_{j+1} b_{j+1}\}}{P\{c_k B > b - x\}} = 1,$$

and therefore

$$\lim_{b \rightarrow \infty} \frac{\sum_{k=j+1}^n P\{c_k B > b - x - c_{j+1} b_{j+1}\}}{\sum_{k=j+1}^n P\{c_k B > b - x\}} = 1.$$

In addition we can use the regularly varying tails property to see

$$\lim_{b \rightarrow \infty} \frac{P\{c_{j+1} B > b - x - c_{j+1} b_{j+1}\}}{\sum_{k=j+1}^n P\{c_k B > b - x\}} = \left(\sum_{k=j+1}^n \left(\frac{c_k}{c_{j+1}} \right)^\alpha \right)^{-1}.$$

Combining the previous two displays we see that

$$\begin{aligned} & \lim_{b \rightarrow \infty} \frac{\hat{V}_b(j+1, x + c_{j+1} b_{j+1})}{\hat{V}_b(j, x)} I\{c_{j+1} b_{j+1} \leq a(b-x)\} \\ &= \frac{d_{j+1}}{d_j} \left(\frac{\sum_{k=j+1}^n \left(\frac{c_k}{c_{j+1}} \right)^\alpha - 1}{\sum_{k=j+1}^n \left(\frac{c_k}{c_{j+1}} \right)^\alpha} \right)^2. \end{aligned}$$

□

We continue with the proof of the asymptotic upper bound (14). From Proposition 3.3 and Lemma 2.1 we have seen that the second moment of $\hat{p}_b^{(1)}$ satisfies

$$E \left[I\{X_n > b\} \prod_{j=0}^{n-1} \frac{f_B((X_{j+1} - X_j)/c_{j+1})}{g_{B_{j+1}}^C((X_{j+1} - X_j)/c_{j+1}|X_j)} \right] \leq \hat{V}_b(0, 0) \prod_{j=0}^{n-1} \gamma_j^C,$$

where $\beta_j^C(b) = \max(\tilde{\beta}_j^C(b), 1)$. Let us study the limit of $\beta_j^C(b)$ when b is large.

From (16) we observe that the optimal probabilities $p(j, x)$ do not need to depend on x . Therefore we call them p_j and minimize $\tilde{\beta}_{j+1}^C$ over p_{j+1} and obtain

$$p_{j+1} = \frac{\sqrt{v_1(j+1)}}{\sqrt{v_1(j+1)} + \sqrt{v_2(j+1)}}, \quad (19)$$

for $j \leq n-2$, where

$$v_1(j+1) = \frac{c_{j+1}^{2\alpha}}{a^{2\alpha} d_j} \left(\sum_{k=j+1}^n c_k^\alpha \right)^{-2},$$

and

$$v_2(j+1) = \frac{d_{j+1}}{d_j} \left(\frac{\sum_{k=j+1}^n \left(\frac{c_k}{c_{j+1}} \right)^\alpha - 1}{\sum_{k=j+1}^n \left(\frac{c_k}{c_{j+1}} \right)^\alpha} \right)^2 = \frac{d_{j+1}}{d_j} \left(\frac{\sum_{k=j+2}^n c_k^\alpha}{\sum_{k=j+1}^n c_k^\alpha} \right)^2.$$

The optimum value of p_n is trivially 1. Plugging in the optimal probabilities and $d_j = a^{-2\alpha}/p_0$ for all j into (16) gives

$$\begin{aligned}\tilde{\beta}_{j+1}^C &= \left(\frac{c_{j+1}^\alpha}{a^\alpha \sqrt{d_j}} \left(\sum_{k=j+1}^n c_k^\alpha \right)^{-1} + \left(\frac{d_{j+1}}{d_j} \right)^{1/2} \frac{\sum_{k=j+2}^n c_k^\alpha}{\sum_{k=j+1}^n c_k^\alpha} \right)^2 \\ &= \left(\sum_{k=j+1}^n c_k^\alpha \right)^{-2} \left(\sqrt{p_0} c_{j+1}^\alpha + \sum_{k=j+2}^n c_k^\alpha \right)^2,\end{aligned}\quad (20)$$

for $j = 0, \dots, n-2$, and $\tilde{\beta}_n^C = \max(1, a^{-2\alpha}/d_{n-1}) = 1$. Combining the asymptotic bound (16) and Proposition 3.3 implies that

$$\limsup_{b \rightarrow \infty} \frac{1}{\tilde{V}_b(0,0)} E \left[I\{X_n > b\} \prod_{j=0}^{n-1} \frac{f_B((X_{j+1} - X_j)/c_{j+1})}{g_{B_{j+1}}^C((X_{j+1} - X_j)/c_{j+1}|X_j)} \right] \leq \prod_{j=0}^{n-1} \tilde{\beta}_{j+1}^C,$$

which, for the normalized second moment, gives

$$\begin{aligned}\limsup_{b \rightarrow \infty} \frac{1}{p_b^2} E \left[I\{X_n > b\} \prod_{j=0}^{n-1} \frac{f_B((X_{j+1} - X_j)/c_{j+1})}{g_{B_{j+1}}^C((X_{j+1} - X_j)/c_{j+1}|X_j)} \right] \\ \leq d_0 \prod_{j=0}^{n-1} \tilde{\beta}_{j+1}^C \leq d_0 = \frac{a^{-2\alpha}}{p_0}.\end{aligned}$$

Thus establishing Theorem 3.1.

4. THE SOLUTION TO THE STOCHASTIC RECURRENCE EQUATION

In this section the main problem is studied. Consider the solution $\{X_k; k = 0, \dots, n\}$ to a stochastic recurrence equation of the form

$$X_{n+1} = A_{n+1}X_n + B_{n+1}, \quad X_0 = 0, \quad (21)$$

where $\{A_k\}$ and $\{B_k\}$ form two independent sequences consisting of independent and identically distributed random variables. It is assumed that the tail of the distribution of B_1 is regularly varying with index $-\alpha$ and that A_1 satisfies

$$E[A_1^{2\alpha+\epsilon}] < \infty,$$

for some $\epsilon > 0$. The solution to the stochastic recurrence equation (21) can be written as

$$X_n = B_n + A_n B_{n-1} + \dots + A_n \dots A_2 B_1. \quad (22)$$

The moment condition on A_1 implies that any finite product of $A_n \dots A_{k+1}$, $k = 0, \dots, n-1$ has lighter tails than B_k and it follows from Breiman's lemma [Breiman 1965] that each term in the representation of X_n satisfies

$$P\{A_n \dots A_{k+1} B_k > b\} \sim E[(A_n \dots A_{k+1})^\alpha] P\{B_1 > b\}.$$

Since the terms are asymptotically independent it is most likely that the sum is large because one of the terms is large and it is possible to show that, see e.g. [Kostantinides and Mikosch 2005] for details,

$$p_b = P\{X_n > b\} \sim P\{B_1 > b\} \sum_{k=0}^{n-1} E[(A_n \dots A_{n-k})^\alpha]. \quad (23)$$

When it comes to efficient importance sampling for computing p_b we will make use of the following observation. Let $\{Y_k; k = 1, \dots, n\}$, $Y_0 = 0$ be the process defined by

$$\begin{aligned} Y_{k+1} &= Y_k + C_{k+1}B_{k+1}, \quad k = 0, \dots, n-1, \\ C_{k+1} &= A_n \cdots A_{k+2}, \quad k = 0, \dots, n-2, \\ C_n &= 1. \end{aligned}$$

By (22) it is immediately clear that $Y_n = X_n$ and therefore $p_b = P\{Y_n > b\}$. To construct the importance sampler the basic idea is to mimic the most likely way that the rare event occurs, which indicates that we do not need to pay special attention to the multiplicative factors A_1, \dots, A_n . Because of the independence between the sequences $\{A_k; k = 1, \dots, n\}$ and $\{B_k; k = 1, \dots, n\}$ there is no problem to simulate A_1, \dots, A_n first and then sample $\{B_k; k = 1, \dots, n\}$ conditional on the A 's.

We propose simulating A_1, \dots, A_n from their original distribution and then, given their outcomes, use one of the (weighted) random walk algorithms, described in Section 3, to sample B_1, \dots, B_n . In what follows the conditional mixture algorithm will be used. The estimate \hat{p}_b of p_b is the arithmetic mean of N independent copies of $\hat{p}_b^{(1)}$, where the algorithm for computing $\hat{p}_b^{(1)}$ can then be described as follows.

Our main result states that the estimator $\hat{p}_b^{(1)}$ obtained by Algorithm 1 has bounded relative error.

THEOREM 4.1. *Let $\{X_k; k = 0, \dots, n\}$ be the solution to the stochastic recurrence equation given in (21). The estimator $\hat{p}_b^{(1)}$ obtained from Algorithm 1 just described has bounded relative error for computing $p_b = P\{X_n > b\}$. That is,*

$$\sup_{b>0} \frac{\tilde{E}((\hat{p}_b^{(1)})^2)}{P\{X_n > b\}^2} < \infty.$$

4.1. Proof of Theorem 4.1

In analogy to (12) we write

$$\hat{V}_b(j, y; A_1, \dots, A_n) = \min \left\{ d_j \left(\sum_{k=j+1}^n P \left\{ B > \frac{b-y}{A_n \cdots A_{k+1}} \mid A_1, \dots, A_n \right\} \right)^2, 1 \right\}, \quad (25)$$

where an empty product is interpreted as unity. By definition of the estimator $\hat{p}_b^{(1)}$ in (24) we have

$$\tilde{E}[(\hat{p}_b^{(1)})^2] = E \left[E \left[I\{Y_n > b\} \prod_{k=0}^{n-1} L_{k+1}(B_{k+1} \mid A_1, \dots, A_n, Y_k) \mid A_1, \dots, A_n \right] \right],$$

and it follows from Proposition 3.3 that

$$\begin{aligned} & E \left[I\{Y_n > b\} \prod_{k=0}^{n-1} L_{k+1}(B_{k+1} \mid A_1, \dots, A_n, Y_k) \mid A_1, \dots, A_n \right] \\ & \leq \hat{V}_b(0, 0; A_1, \dots, A_n) \prod_{j=0}^{n-1} \beta_j^C(b) \quad \text{a.s.} \end{aligned}$$

TOMACS 1: State dependent mixture algorithm for $p_b = P_0\{X_n > b\}$

Input: Sampling densities $g_i(x | s)$, $i = 1, \dots, n$ and numbers $p_i + q_i = 1$, for $i = 1, \dots, n - 1$, with $p_i \in (0, 1)$.

Output: A sample X_1, \dots, X_n of a stochastic recurrence equation and a single replication of an estimator for (3).

Sample A_1, \dots, A_n from their original distribution;

Put $C_{k+1} = A_n \cdots A_{k+2}$ for $k = 0, \dots, n - 2$, and $C_n = 1$;

Let $\hat{V}_b, \Gamma_1, \dots, \Gamma_n, p_1, \dots, p_n$, and $a \in (0, 1)$ be as in the statement of Theorem 3.1;

$X_0 = 0$; $k = 0$;

while $k \leq n - 1$ **do**

if $\hat{V}_b(k, X_k) \geq 1$ **then**

 Sample B_{k+1} from original distribution

else

 Draw U from uniform on $[0, 1]$;

if $U \leq p_{k+1}$ **then**

 Sample B_{k+1} from the density

$$\frac{f_B(y)}{P\{B > a(b - X_k)/C_{k+1}\}} I(y > a(b - X_k)/C_{k+1});$$

else

 Sample B_{k+1} from density

$$\frac{f_B(y)}{P\{B \leq a(b - X_k)/C_{k+1}\}} I(y \leq a(b - X_k)/C_{k+1});$$

end

end

$k = k + 1$ and $X_{k+1} \rightarrow X_k + C_{k+1}B_{k+1}$;

end

Form the likelihood

$$\begin{aligned} L_{k+1}(B_{k+1} | A_1, \dots, A_n, X_k) &= I\{X_k \notin \Gamma_{k+1}\} \\ &+ \frac{\bar{F}(a(b - X_k)/C_{k+1})}{p(k+1, X_k)} I\{B_{k+1} > a(b - X_k)/C_{k+1}\} I\{X_k \in \Gamma_{k+1}\} \\ &+ \frac{1 - \bar{F}(a(b - X_k)/C_{k+1})}{1 - p(k+1, X_k)} I\{B_{k+1} \leq a(b - X_k)/C_{k+1}\} I\{X_k \in \Gamma_{k+1}\}; \end{aligned}$$

Return the quantity

$$\hat{p}_b^{(1)} = I\{X_n > b\} \prod_{k=0}^{n-1} L_{k+1}(B_{k+1} | A_1, \dots, A_n, X_k); \quad (24)$$

Here $\beta_j^C(b)$, $j = 1, \dots, b$, are bounded in b and independent of A_1, \dots, A_n . In particular,

$$\tilde{E}[(\hat{p}_b^{(1)})^2] \leq E\left[\hat{V}_b(0, 0; A_1, \dots, A_n)\right] \prod_{j=0}^{n-1} \beta_j^C(b).$$

It remains to show that

$$\limsup_{b \rightarrow \infty} \frac{E[\hat{V}_b(0, 0; A_1, \dots, A_n)]}{p_b^2} < \infty. \quad (26)$$

To prove (26), take an arbitrary $v > 0$ and let

$$G_{b/v} = \{\max\{A_n, A_n A_{n-1}, \dots, A_n \cdots A_2\} \leq b/v\}.$$

Consider the following decomposition:

$$E[\hat{V}_b(0, 0; A_1, \dots, A_n)] = E[\hat{V}_b(0, 0; A_1, \dots, A_n)I_{G_{b/v}}] + E[\hat{V}_b(0, 0; A_1, \dots, A_n)I_{G_{b/v}^c}].$$

For the first term we have, for b sufficiently large

$$\begin{aligned} & E[\hat{V}_b(0, 0; A_1, \dots, A_n)I_{G_{b/v}}] \\ &= E\left[d_0 \left(\sum_{k=1}^n P\left\{B > \frac{b}{A_n \cdots A_{k+1}} \mid A_1, \dots, A_n\right\} \right)^2 I_{G_{b/v}}\right] \end{aligned}$$

Take $\delta > 0$ such that $E[A_1^{2\alpha+2\delta}] < \infty$. For v and b sufficiently large, Potter's bound (see Theorem 3.2) implies that on $G_{b/v}$

$$\frac{P\left\{B > \frac{b}{A_n \cdots A_{k+1}} \mid A_1, \dots, A_n\right\}}{P\{B > b\}} \leq c_1 (A_n \cdots A_{k+1})^{\alpha+\delta},$$

for some constant c_1 . Therefore,

$$\limsup_{b \rightarrow \infty} \frac{E[\hat{V}_b(0, 0; A_1, \dots, A_n)I_{G_{b/v}}]}{P\{B > b\}^2} \leq E\left[d_0 \left(\sum_{k=1}^n c_1 (A_n \cdots A_{k+1})^{\alpha+\delta} \right)^2\right] < \infty.$$

For the second term it holds that

$$\begin{aligned} & \limsup_{b \rightarrow \infty} \frac{E[\hat{V}_b(0, 0; A_1, \dots, A_n)I_{G_{b/v}^c}]}{P\{B > b\}^2} \\ & \leq \frac{P\{G_{b/v}^c\}}{P\{B > b\}^2} \leq \sum_{k=1}^n \frac{P\{A_n \cdots A_{k+1} > b/v\}}{P\{B > b\}^2}. \end{aligned}$$

By Chebyshev's inequality it follows that each term in the sum satisfies

$$\limsup_{b \rightarrow \infty} \frac{P\{A_n \cdots A_{k+1} > b/v\}}{P\{B > b\}^2} \leq \limsup_{b \rightarrow \infty} \frac{E(A_n \cdots A_{k+1})^{2\alpha+2\delta}}{(b/v)^{2\alpha+2\delta} P\{B > b\}^2} = 0.$$

The claim follows because, by (23),

$$\lim_{b \rightarrow \infty} \frac{P\{X_n > b\}}{P\{B > b\}} = \sum_{k=1}^n E[(A_n \cdots A_{n-k})^\alpha].$$

This completes the proof.

5. A FIRST PASSAGE TIME PROBLEM

We continue within the context of (21) but now we study the problem of designing an importance sampling estimator for

$$q_b = P_0 \left\{ \sup_{k \leq n(b)} X_k > b \right\}$$

with bounded relative error when $b, n(b) \nearrow \infty$. We assume that $n(b)P\{B_1 > b\} = o(1)$ as $b \nearrow \infty$.

Assumption 2. We assume that there exists $\theta_* > 2\alpha$ such that $EA_j^{\theta_*} = 1$.

Note that if such $\theta_* > 0$ exists then by continuity one can guarantee that there exists $\rho \in (0, 1)$ and $\theta' \in (2\alpha, \theta_*)$ such that $E[A_j^{\theta'}] = \rho^{\theta'}$. We note that this selection implies that $E[\sup\{m \geq 0 : \prod_{j=1}^m (\rho^{-1} A_j)\}^{2\alpha+\varepsilon}] < \infty$ for some $\varepsilon > 0$.

We will take advantage of a different technique, other than sequential importance sampling, based on a variation of a mixture procedure called Target Bridge Sampling (TBS) (see [Blanchet and Li 2011; Adler et al. 2012]).

First we simulate A_1, \dots, A_n . Given these values, standard TBS requires efficient sampling of X_1, \dots, X_l given $X_l > b$ and explicit evaluation of $p_{b,l} = P_0\{X_l > b\}$ for each $l \in \{1, \dots, n(b)\}$. However, since computing $p_{b,l}$ explicitly is not feasible in our setting, we will replace the event $\{X_l > b\}$ by a slightly larger event, denoted by $\mathcal{A}_{b,l}$, so that $P_0(\mathcal{A}_{b,l})$ is easy to compute.

The procedure TBS constructs a change of measure, Q_n , by forming a mixture distribution for (X_1, \dots, X_n) which is not Markovian but basically attempts to sample a point (T, X_T) in the set $\{1, 2, \dots, n\} \times [b, \infty]$; this is called the ‘‘Target’’ step. Then, conditional on this point, one samples X_1, \dots, X_T – this is the ‘‘Bridge Sampling’’ step. The value of $T = l$ is selected according to a probability mass function based on $P_0(\mathcal{A}_{b,l})$, and the value X_l given $T = l$ is sampled from the conditional distribution of X_l given the event $\mathcal{A}_{b,l}$.

Let us now construct our sampling procedure explicitly. Throughout the following we shall assume that $A_1 = a_1, \dots, A_n = a_n$ have been simulated.

Defining an appropriate enlarged set $\mathcal{A}_{b,l}$. In order to implement TBS first we note that

$$X_l = A_2 \dots A_l B_1 + A_3 \dots A_l B_2 + \dots + A_l B_{l-1} + B_l.$$

Therefore, $X_l > b$ implies that for each sequence of numbers $\gamma_{k,l} > 0$ satisfying $\sum_{k=1}^l \gamma_{k,l} \leq 1$ it holds that there exists $k \in \{1, \dots, l\}$ for which $B_k a_{k+1} \dots a_l \geq b \gamma_{k,l}$. We select $\gamma_{k,l} = \rho^{l-k}(1 - \rho)$ for $\rho \in (0, 1)$ indicated in Assumption 2. The event

$$\mathcal{A}_{b,l} \doteq \bigcup_{k=1}^l \{B_k a_{k+1} \dots a_l \geq b \gamma_{k,l}\}$$

will be used instead of $\{X_l > b\}$ as discussed earlier. Note that $\{X_l > b\} \subseteq \mathcal{A}_{b,l}$ and that

$$\beta_{b,l}(a_1, \dots, a_l) \doteq P\{\mathcal{A}_{b,l}\} = 1 - \prod_{k=1}^l P\{B_k a_{k+1} \dots a_l < b \gamma_{k,l}\}$$

can be computed in closed form by performing $O(l)$ function evaluations.

Sampling conditional on the set $\mathcal{A}_{b,l}$. We now explain how to sample X_1, \dots, X_l given $\mathcal{A}_{b,l}$. Let us write

$$\begin{aligned} \eta_{l,k}(a_1, \dots, a_l) &= P(B_k a_{k+1} \dots a_l \geq b \gamma_{k,l}), \\ w_{l,k}(a_1, \dots, a_l) &= \frac{\eta_{l,k}(a_{k+1}, \dots, a_l)}{\sum_{k=1}^l \eta_{l,k}(a_{k+1}, \dots, a_l)}, \end{aligned}$$

and define a probability measure Q_l^* as follows. For arbitrary Borel subsets of the real line, $\mathcal{B}_1, \dots, \mathcal{B}_l$, put

$$Q_l^* \{B_1 \in \mathcal{B}_1, \dots, B_l \in \mathcal{B}_l\} = \sum_{k=1}^l w_{l,k}(a_1, \dots, a_l) P\{B_1 \in \mathcal{B}_1, \dots, B_l \in \mathcal{B}_l \mid B_k a_{k+1} \dots a_l \geq b \gamma_{k,l}\}.$$

TOMACS 2: Target Bridge Sampling**Input:** Distribution of A and B , n , b and ρ .**Output:** Single replication of importance sampling estimator: $\hat{q}_b^{(1)}$.Sample n i.i.d. copies from distribution of $A \rightarrow a_1, \dots, a_n$;Form vector $\beta_{b,\ell}(a_1, \dots, a_\ell)$ $1 \leq \ell \leq n$;Generate integer $j \in \{1, \dots, n\}$ according to probability distribution formed by normalizing $\{\beta_{b,\ell}\}_{\ell=1}^n$;Sample B_1, \dots, B_j given $\mathcal{A}_{b,j}$ via acceptance-rejection algorithm with Q_j^* as proposal distribution.;

Return

$$\frac{\sum_{\ell=1}^n \beta_{b,\ell}(a_1, \dots, a_\ell) I\{\max_{k \leq n(b)} X_k > b\}}{\sum_{\ell=1}^n I(\mathcal{A}_{b,\ell})}.$$

Note that simulation from Q_l^* is straightforward; simply select an index k according to the probabilities $w_{l,k}$, then simulate B_k given that $B_k a_{k+1} \dots a_l \geq b \gamma_{k,l}$ and finally sample the rest of the B_j 's independently.

Now set $P_l^* \{\cdot\} = P\{\cdot \mid \mathcal{A}_{b,l}\}$ and observe that

$$\frac{dP_l^*}{dQ_l^*} = \frac{\sum_{k=1}^l \eta_{l,k}(a_{k+1}, \dots, a_l)}{\sum_{k=1}^l I\{B_k a_{k+1} \dots a_l \geq b \gamma_{k,l}\}} \times \frac{I\{\mathcal{A}_{b,l}\}}{\beta_{b,l}(a_1, \dots, a_\ell)} \leq \frac{\sum_{k=1}^l \eta_{l,k}(a_{k+1}, \dots, a_l)}{\beta_{b,l}(a_1, \dots, a_l)}.$$

The right hand side is a constant, so acceptance/rejection can be applied in order to simulate from P_l^* . The expected number of proposals from Q_l^* required to sample from P_l^* is precisely the ratio

$$R_{b,l}(a_1, \dots, a_l) = \frac{\sum_{k=1}^l \eta_{l,k}(a_{k+1}, \dots, a_l)}{\beta_{b,l}(a_1, \dots, a_l)}. \quad (27)$$

The importance sampling distribution for Q_n TBS and estimator for $q_b = P_0\{\sup_{k \leq n(b)} X_k > b\}$. We proceed to construct Q_n via the mixture distribution

$$Q_n \{\cdot\} = \frac{\sum_{l=1}^n P_l^* \{\cdot\} \beta_{b,l}(a_1, \dots, a_l)}{\sum_{l=1}^n \beta_{b,l}(a_1, \dots, a_l)}.$$

Note that it is straightforward to sample from Q_n . First one selects an index l with probability $\beta_{b,l} / \sum_{l=1}^n \beta_{b,l}$ and then one samples B_1, \dots, B_l given $\mathcal{A}_{b,l}$ – as explained earlier.

We then can output a single replication of the importance sampling estimator

$$\hat{q}_b^{(1)}(a_1, \dots, a_n) \doteq \frac{dP}{dQ_n} I\{\max_{k \leq n(b)} X_k > b\} = \frac{\sum_{l=1}^n \beta_{b,l}(a_1, \dots, a_l) I\{\max_{k \leq n(b)} X_k > b\}}{\sum_{l=1}^n I\{\mathcal{A}_{b,l}\}}. \quad (28)$$

It is important to observe that the denominator in (28) is at least one on the set $\{\max_{k \leq n(b)} X_k > b\}$ (this follows from the fact that $\{X_l > b\} \subseteq \mathcal{A}_{b,l}$). We then conclude that

$$\hat{q}_b^{(1)}(a_1, \dots, a_n) \leq \sum_{l=1}^n \beta_{b,l}(a_1, \dots, a_l). \quad (29)$$

Efficiency analysis. The following is the main result of this section. It summarizes the efficiency properties of the estimator (28).

THEOREM 5.1. *As long as $n(b)P\{B_1 > b\} = o(1)$ as $b \rightarrow \infty$, the estimator generated by Q_n in (28) possesses the bounded relative error property, namely*

$$\sup_{b>0} \frac{E[E^{Q_n}[(\hat{q}_b^{(1)}(A_1, \dots, A_{n(b)}))^2 | A_1, \dots, A_{n(b)}]]}{q_b^2} < \infty.$$

In addition, the expected number of function evaluations required to generate $\hat{q}_b^{(1)}(A_1, \dots, A_{n(b)})$ is $O(n(b)^2)$.

PROOF. First we will derive a simple lower bound on q_b . Note that $X_l > B_l$ and therefore

$$q_b \geq P \left\{ \bigcup_{k=1}^n \{B_k > b\} \right\}.$$

Via the Bonferroni inequality we have that

$$P \left\{ \bigcup_{k=1}^n \{B_k > b\} \right\} \geq n(b)P\{B_k > b\} - n(b)^2 P\{B_k > b\}^2 / 2,$$

and since we assumed that $n(b)P\{B_1 > b\} = o(1)$ as $b \rightarrow \infty$, it follows that $q_b = \Omega(n(b)P\{B_1 > b\})$ as $b \rightarrow \infty$. So, in view of (29) we need to verify that

$$E \left[\left(\sum_{l=1}^n \beta_{b,l}(A_1, \dots, A_l) \right)^2 \right] = O \left(n(b)^2 P\{B_1 > b\}^2 \right).$$

We directly compute

$$\left(\sum_{l=1}^n \beta_{b,l}(A_1, \dots, A_l) \right)^2 = \sum_{l=1}^n \beta_{b,l}(A_1, \dots, A_l)^2 + 2 \sum_{1 \leq m < l \leq n} \beta_{b,m}(A_1, \dots, A_m) \beta_{b,l}(A_1, \dots, A_l).$$

Now, note that

$$\beta_{b,l}(A_1, \dots, A_l)^2 \leq \sum_{k=1}^l \eta_{l,k}(A_{k+1}, \dots, A_l)^2 + 2 \sum_{1 \leq j < k \leq l} \eta_{l,j}(A_{j+1}, \dots, A_l) \eta_{l,k}(A_{k+1}, \dots, A_l).$$

Observe that there exists a slowly varying function L such that $P(B > x) = x^{-\alpha} L(x)$ and therefore

$$\begin{aligned} & \eta_{l,k}(a_{k+1}, \dots, a_l)^2 \\ &= P \left\{ B_1 \times \left(\frac{a_{k+1}}{\rho} \right) \times \dots \times \left(\frac{a_l}{\rho} \right) > b(1-\rho) \right\}^2 \\ &= \left(\frac{a_{k+1}}{\rho} \right)^{2\alpha} \times \dots \times \left(\frac{a_l}{\rho} \right)^{2\alpha} \frac{L^2(b(1-\rho))}{(b(1-\rho))^{-2\alpha}} \frac{L^2((\rho/a_{k+1}) \times \dots \times (\rho/a_l) b(1-\rho))}{L^2(b(1-\rho))}. \end{aligned}$$

By Potter's bound (Theorem 3.2) we have that for every $\varepsilon > 0$ there exists $K(\varepsilon)$ such that

$$\begin{aligned} & \frac{L^2 (a_{k+1}^{-1} \rho \times \dots \times a_l^{-1} \rho b(1-\rho))}{L^2 (b(1-\rho))} \\ & \leq K(\varepsilon) \max \left(\left(\frac{a_{k+1}}{\rho} \right)^\varepsilon, \left(\frac{a_{k+1}}{\rho} \right)^{-\varepsilon} \right) \times \dots \times \max \left(\left(\frac{a_l}{\rho} \right)^\varepsilon, \left(\frac{a_l}{\rho} \right)^{-\varepsilon} \right). \end{aligned}$$

On the other hand, by Assumption 2 we know that $E(\rho^{-1}A_1)^{\theta'} = 1$ for $\theta' \in (2\alpha, \theta_*)$ and therefore it follows that we can find $\varepsilon > 0$ and $\phi \in (0, 1)$ such that

$$E[(A_j/\rho)^{2\alpha} \max((A_j/\rho)^\varepsilon, (A_j/\rho)^{-\varepsilon})] \leq \phi.$$

Note that we can apply Potter's bound to the potentially unbounded A_i via the same argument used in the proof of the Theorem 4.1. Thus we conclude that

$$E[\eta_{l,k}(A_{k+1}, \dots, A_l)^2] \leq KP \{B > (1-\rho)b\}^2 \phi^{k-l}$$

for some $K > 0$ uniformly over $l \leq n(b)$. Consequently we can conclude that we can select a constant $\phi \in (0, 1)$ such that

$$E\left[\sum_{k=1}^l \eta_{l,k}(A_{k+1}, \dots, A_l)^2\right] = O(P\{B > b(1-\rho)\}^2 \sum_{k=0}^{l-1} \phi^k) = O(P\{B_1 > b\}^2).$$

Similarly, we have that for $j < k$

$$E[\eta_{l,j}(A_{j+1}, \dots, A_l)\eta_{l,k}(A_{k+1}, \dots, A_l)] = O\left(P\{B_1 > b\}^2 \phi^{l-j}\right)$$

and therefore

$$\begin{aligned} & 2E \left[\sum_{1 \leq j < k \leq l} \eta_{l,j}(A_{j+1}, \dots, A_l)\eta_{l,k}(A_{k+1}, \dots, A_l) \right] \\ & = O\left(P\{B_1 > b\}^2 \sum_{1 \leq j < k \leq l} \phi^{l-j}\right) \\ & = O\left(P\{B_1 > b\}^2 \sum_{1 \leq j \leq l} (l-j)\phi^{l-j}\right) = O\left(P\{B_1 > b\}^2\right). \end{aligned}$$

We then conclude that

$$E \left[\sum_{l=1}^n \beta_{b,l}(A_1, \dots, A_l)^2 \right] = O\left(n(b) P\{B_1 > b\}^2\right).$$

Now we study, for $m < l$,

$$\begin{aligned} & \beta_{b,m}(A_1, \dots, A_m)\beta_{b,l}(A_1, \dots, A_l) \leq \left[\sum_{k=1}^m \eta_{m,k}(A_{k+1}, \dots, A_m) \right] \times \left[\sum_{k'=1}^l \eta_{l,k'}(A_{k'+1}, \dots, A_l) \right] \\ & = \sum_{k=1}^m \sum_{k'=1}^l \eta_{m,k}(A_{k+1}, \dots, A_m)\eta_{l,k'}(A_{k'+1}, \dots, A_l) \end{aligned}$$

Working with each of the terms inside the summations above we obtain, as before, that one can pick $\phi \in (0, 1)$ such that

$$E[\eta_{m,k}(A_{k+1}, \dots, A_l)\eta_{l,k'}(A_{k'+1}, \dots, A_l)] = O\left(P\{B > b\}^2 \phi^{m-k} \phi^{l-k'}\right).$$

Consequently,

$$\begin{aligned} E[\beta_{b,m}(A_1, \dots, A_m)\beta_{b,l}(A_1, \dots, A_l)] \\ = O\left(P\{B > b\}^2\right) \sum_{k=1}^m \sum_{k'=1}^l \phi^{m-k} \phi^{l-k'} = O\left(P\{B > b\}^2\right). \end{aligned}$$

In turn, we obtain that

$$\sum_{1 \leq m < l \leq n(b)} E[\beta_{b,m}(A_1, \dots, A_m)\beta_{b,l}(A_1, \dots, A_l)] = O\left(n(b)^2 P\{B > b\}^2\right),$$

thereby concluding the first part of the theorem. For the complexity count of the cost per replication, we have noted that sampling from Q_l^* and computing $\beta_{b,l}$ requires $O(l)$ function evaluations. The only remaining part to show is that (recall the definition of $R_{b,l}(A_1, \dots, A_l)$ in equation (27)),

$$E[R_{b,l}(A_1, \dots, A_l)] = O(1)$$

uniformly over $l \leq n(b)$ as $b \rightarrow \infty$. However, this follows just as in our previous analysis because

$$\beta_{b,l}(A_1, \dots, A_l) \geq P\{B > \gamma_{l,b}\} = P\{B > (1 - \rho)b\} = \Omega(P\{B > b\}),$$

which implies that

$$E[R_{b,l}(A_1, \dots, A_l)] = O\left(\frac{E\left[\sum_{k=1}^l \eta_{l,k}(A_{k+1}, \dots, A_l)\right]}{P\{B > b\}}\right) = O\left(\sum_{k=1}^l \phi^{k-l}\right) = O(1).$$

Therefore, evaluating the likelihood ratio for $\hat{q}_b^{(1)}(A_1, \dots, A_{n(b)})$ takes $O(\sum_{l=1}^{n(b)} l) = O(n(b)^2)$ function evaluations in expectation. \square

6. NUMERICAL EXAMPLES

All simulations in this section are based on 500000 Monte Carlo replications.

6.1. Estimation of p_b

In this section we utilize Algorithm 1 to estimate p_b . In all tables of this section we assume that $n = 50$ is fixed, we always set the parameter $a = 0.95$, and we choose the mixture probabilities according to the rule specified in Theorem (3.1). Furthermore we assume that B_1, \dots, B_n are Pareto random variables symmetric about zero with right hand tail probabilities, $\bar{F}(b) = 0.5(1 + b)^{-2}$. In the table I we assume that the random variables A_1, \dots, A_n are log-normally distributed with $\sigma = 0.1$ and $\mu = \log(1.05) - \sigma^2/2$. In table II we assume that the random variables A_1, \dots, A_n are non-negative Pareto random variables with $P(A_1 > t) = (1 + t)^{-5}$. Lastly we consider the setting where the random variables A_i have an exponential distribution with mean $1/4$, the estimation in this setting is given in III. From these tables we see that the relative error of the algorithm 1 clearly stays bounded, and furthermore that the computational cost seems independent of b . Lastly we see that the algorithm has slightly better performance in the light tailed setting of Table III versus the heavy tailed settings of Tables I and II.

Table I. Estimation of p_b in Log-Normal setting

b	Estimate	Std. Error	Rel. Error	Comp. Time
25	0.0145	6.808e-05	0.004696	33.29
250	0.0001184	2.271e-07	0.001918	32.35
2500	1.181e-06	1.527e-09	0.001292	31.24
2.5e+04	1.182e-08	1.538e-11	0.0013	31.24

Table II. Estimation of p_b in the Pareto setting

b	Estimate	Std. Error	Rel. Error	Comp. Time
25	0.0008859	1.346e-06	0.00152	33.53
250	9.521e-06	1.589e-08	0.001669	32.89
2500	9.612e-08	1.627e-10	0.001692	34.06
2.5e+04	9.591e-10	1.385e-12	0.001444	33.75

Table III. Estimation of p_b in the Exponential setting

b	Estimate	Std. Error	Rel. Error	Comp. Time
25	0.0008509	9.152e-07	0.001076	31.31
250	9.08e-06	6.415e-09	0.0007065	31.17
2500	9.136e-08	6.464e-11	0.0007076	31.44
2.5e+04	9.138e-10	6.549e-13	0.0007167	31.73

6.2. Estimation of q_b

In this section we utilize Algorithm 2 to estimate q_b . In all tables of this section we assume that $n(b) = \text{ceil}(\sqrt{b})$. Furthermore we assume that B_1, \dots, B_n are non-negative Pareto random variables with tail probabilities, $\bar{F}(b) = (1+b)^{-2}$. We assume that the random variables A_1, \dots, A_n are log-normally distributed with $\sigma = 0.1$ and $\mu = \log(1.05) - \sigma^2/2$. In Table IV we see that the relative error does not grow with b , this is consistent with Theorem 5.1. However we do see that computational cost does grow with b . This is due to the fact that we are now setting $n(b) \approx \sqrt{b}$, and therefore the computational cost associated with each replication is $\approx b$.

ACKNOWLEDGMENTS

The authors would like to thank the editorial staff of ACM TOMACS and two anonymous referees for their helpful comments.

REFERENCES

- R. Adler, J. Blanchet, and J. C. Liu. 2012. Efficient Simulation of High Excursions of Gaussian Random fields. *Ann. Appl. Probab.* 22 (2012), 1167–1214.
- S. Asmussen and K. Binswanger. 1997. Simulation of ruin probabilities for subexponential claims. *Astin Bull.* 27 (1997), 297–318.
- S. Asmussen, K. Binswanger, and B. Højgaard. 2000. Rare events simulation for heavy-tailed distributions. *Bernoulli* 6 (2000), 303–322.
- S. Asmussen and D. Kroese. 2006. Improved algorithms for rare event simulation with heavy tails. *Adv. in Appl. Probab.* 38 (2006), 545–558.
- B. Basrak, R. Davis, and T. Mikosch. 2002. Regular variation of GARCH processes. *Stochastic Process. Appl.* 99 (2002), 95–115.
- N.H. Bingham, C.M. Goldie, and I.L. Teugels. 1987. *Regular Variation*. Cambridge University Press, Cambridge, United Kingdom.
- J. Blanchet and P. Glynn. 2008. Efficient rare-event simulation for the maximum of a heavy-tailed random walk. *Ann. of Appl. Probab.* 18 (2008), 1351–1378.
- J. Blanchet, P. Glynn, and K. Leder. 2012. On Lyapunov Inequalities and Subsolutions for Efficient Importance Sampling. *ACM TOMACS* 22, 3 (2012).
- J. Blanchet and H. Lam. 2011. State-dependent importance sampling for rare-event simulation: An overview and recent advances. *Surveys in Operations Research and Management Science* 17 (2011), 3824–3831.

Table IV. Estimation of q_b via the TBS algorithm

b	Estimate	Std. Error	Rel. Error	Comp. Time
25	0.01023	0.0001372	0.01341	0.7479
250	0.0003091	1.855e-05	0.06001	3.852
2500	7.998e-06	7.591e-07	0.09491	28.45
2.5e+04	3.168e-07	2.185e-08	0.06896	268.9

- J. Blanchet and C. Li. 2011. Efficient Simulation for the Maximum of Infinite Horizon Gaussian Processes. *J. Appl. Probab.* 48 (2011), 467–489.
- J. Blanchet and J.C. Liu. 2010. Efficient Importance Sampling in Ruin Problems for Multidimensional Regularly Varying Random Walks. *J. Appl. Probab.* 47 (2010), 301–322.
- J. Blanchet, J.C. Liu, and P. Glynn. 2011. Efficient Rare Event Simulation for Regularly Varying Multi-server Queues. *Submitted* (2011).
- N. Boots and P. Shahabuddin. 2001. Statistical tools for simulation design and analysis: simulating ruin probabilities in insurance risk processes with subexponential claims. In *Winter Simulation Conference*. 468–476.
- L. Breiman. 1965. On some limit theorems similar to the arcsine law. *Theory Probab. Appl.* 10 (1965), 323–331.
- H.P. Chan, S. Deng, and T.L. Lai. 2012. Rare-event simulation of heavy-tailed random walks by sequential importance sampling and resampling. *Adv. in Appl. Probab.* 44 (2012), 1173–1196.
- P. Dupuis, K. Leder, and H. Wang. 2008. Importance sampling for sums of random variables with regularly varying tails. *ACM TOMACS* 17 (2008).
- P. Dupuis, K. Leder, and H. Wang. 2009. Importance Sampling for Weighted-Serve-the-Longest-Queue. *Math. Oper. Res.* 34 (2009), 642–660.
- P. Dupuis, A. Sezer, and H. Wang. 2007. Dynamic importance sampling for queueing networks. *Ann. Appl. Probab.* 17 (2007), 1306–1346.
- P. Dupuis and H. Wang. 2004. Importance sampling, large deviations, and differential games. *Stoch. and Stoch. Reports* 76 (2004), 481–508.
- J. Gaier and P. Grandits. 2002. Ruin probabilities in the presence of regularly varying tails and optimal investment. *Insurance: Mathematics and Economics* 30 (2002), 211–217.
- C.M. Goldie. 1991. Implicit Renewal Theory and Tails of Solutions of Random Equations. *Ann. Appl. Probab.* 1 (1991), 126–166.
- J. Hartinger and D. Kortschak. 2009. On the efficiency of the Asmussen-Kroese estimators and its application to stop-loss transforms. *Bl. DGVFM* 30 (2009), 363–377.
- H. Hult and G. Samorodnitsky. 2010. Large deviations for point processes based on stationary sequences with heavy tails. *J. Appl. Probab.* 47 (2010), 1–40.
- H. Hult and J. Svensson. 2012. On importance sampling with mixtures for random walks with heavy tails. *ACM TOMACS* 22, 2 (2012).
- S. Juneja and P. Shahabuddin. 2002. Simulating heavy-tailed processes using delayed hazard rate twisting. *ACM TOMACS* 12 (2002), 94–118.
- H. Kesten. 1973. Random difference equations and renewal theory for products of random matrices. *Acta Math.* 131 (1973), 207–248.
- D. Kostantinides and T. Mikosch. 2005. Large deviations and ruin probabilities for solutions to stochastic recurrence equations with heavy tailed innovations. *Ann. Probab.* 33 (2005), 1992–2035.
- P. Lecuyer, J. Blanchet, B. Tuffin, and P. Glynn. 2010. Asymptotic robustness of estimators in rare-event simulation. *ACM TOMACS* 20, 6 (2010).
- R.C. Lewontin and D. Cohen. 1969. On Population Growth in a randomly varying environment. *Proc. Natl. Acad. of Sci. USA* 62 (1969), 1056–1060.
- H. Nyrhinen. 1999. On the ruin probabilities in a general economic environment. *Stochastic Process. Appl.* 83 (1999), 319–330.
- J. Paulsen. 2008. Ruin models with investment income. *Probability Surveys* 5 (2008), 416–434.
- A. Roitershtein. 2007. One-dimensional linear recursions with Markov-dependent coefficients. *Ann. Appl. Probab.* 17 (2007), 572–608.
- S. Tuljapurkar. 1990. Delayed reproduction and fitness in variable environments. *Proc. Natl. Acad. of Sci. USA* 87 (1990), 1139–1143.

Received February 2007; revised March 2009; accepted June 2009