Efficient Rare-event Simulation for the Maximum of Heavy-tailed Random Walks

Blanchet, J. and Glynn, P.

December, 2006

Abstract

Let $(X_n : n \geq 0)$ be a sequence of iid rv's with negative mean. Set $S_0 = 0$ and define $S_n = X_1 + ... + X_n$. We propose an importance sampling algorithm to estimate the tail of $M = \max\{S_n : n \geq 0\}$ that is strongly efficient for both light and heavy-tailed increment distributions. A key feature of our algorithm is that it is state-dependent. In the presence of light tails, our procedure leads to Siegmund's (1979) algorithm. The rigorous analysis of efficiency requires new Lyapunov-type inequalities that can be useful in the study of more general importance sampling algorithms..

1 Introduction

In this paper, we consider the problem of efficient simulation of first-passage time probabilities for heavy-tailed random walks (rw's). More precisely, suppose that $(S_n : n \geq 0)$ is the rw generated by the sequence of independent and identically distributed (iid) random variables (rv's) $X = (X_n : n \geq 1)$ (i.e. $S_n = S_{n-1} + X_n$ with $S_0 = 0$). We assume that $EX_n < 0$. Define $M = \max\{S_n : n \geq 0\}$ and $\tau(b) = \inf\{n \geq 0 : S_n > b\}$. We are interested in developing efficient simulation methodology to estimate

$$P(\tau(b) < \infty) = P(M > b) \tag{1}$$

when b is large (i.e. the event $\{M > b\}$ is rare) and X_1 is heavy-tailed.

We say that an unbiased simulation estimator $R(b)$ for $P(M > b)$ is strongly efficient as $b \nearrow \infty$ if

$$\overline{\lim}_{b \nearrow \infty} E\left(R(b)^2\right) / P(M > b)^2 < \infty.$$

Strong efficiency implies that the number of simulation runs required to estimate $P(M > b)$ to a given relative accuracy is bounded in b. A weaker criterion that is logarithmic efficiency, which implies that the number of replications required to estimate $P(M > b)$ with a given relative accuracy grows at rate $o(|\log P(M > b)|)$ see Juneja and Shahabuddin (2006) or Bucklew (2004), Section 5.2 for a discussion of efficiency in rare-event simulation.

In this paper, we develop an implementable state-dependent importance sampling algorithm that can be rigorously proved to be strongly efficient. This is the first efficient algorithm the has been developed for estimating the tail of M in the presence of general heavy-tailed increment distributions. Prior efficient algorithms require the increment distribution to be of $M/G/1$ type with regularly varying or Weibull type tails.

A key idea is that our importance distribution is state-dependent. There is a long history of applications of state-dependent importance sampling to simulation problems. Perhaps the first related contributions are those by Hammersley and Morton (1954) and Rosenbluth and Rosenbluth (1955) in the context of molecular simulation; see also the text by Liu (2001) for applications of sequential importance sampling in various scientific contexts. However, a general framework for rigorous analysis of these types of algorithms is still under development. In a sequence of recent papers, Paul Dupuis and Hui Wang (see, for instance, Dupuis and Wang (2004)) have proposed a general methodology that can be applied in the presence of large deviations theory for light-tailed systems. Our paper contributes to this general literature developing Lyapunov-type inequalities (see Theorem 2) that are particularly useful for the analysis of state-dependent algorithms.

The general theory of importance sampling establishes that the theoretically optimal importance distribution (having zero variance) involves sampling from the conditional distribution of the random walk given $\{\tau(b) < \infty\}$. Under this conditional distribution, the random walk has increment distributions that are state-dependent. However, we cannot implement this zero variance sampling scheme because the state-dependent increment distribution requires explicit knowledge of the function $u^*(\cdot) = P(\tau(\cdot) < \infty)$. Our approach involves using asymptotic approximations for $u^*(\cdot)$ to obtain an implementable state-dependent change-of-measure that closely approximates the true conditional distribution. The most common asymptotic approximation for $u^*(\cdot)$ is

$$u^*(b) = P(M > b) \sim \frac{1}{|EX|} \int_b^\infty P(X > s)\, ds, \tag{2}$$

as $b \nearrow \infty$. An important step in our approach is to use (2) in order to construct a function $v(\cdot)$ such that

$$Ev(b - X) - v(b) = o(P(X > b)) \tag{3}$$

as $b \nearrow \infty$. Note that if $v = u^*$, the previous difference vanishes. The convergence rate associated with (3) is convenient in order to develop our simulation algorithm but is not necessary (see Proposition 3 and Theorem 3). We show that a $v(\cdot)$ satisfying

(3) can be constructed using (2) whenever X belongs to the class S^* of heavy-tailed distributions – which is slightly smaller than the class of subexponential distributions but includes regularly varying, Weibull, Lognormal and many more distributions as special cases; see assumption A) in Section 3 for a precise definition.

The problem that we address here is motivated by applications in queueing and insurance. The distribution of M is of great interest in queueing theory as it coincides with the steady-state waiting time distribution of the single-server $G/G/1$ queue. In addition, the first passage time probability displayed in (1) is of central interest in the context of risk insurance processes. In particular, such a first passage time probability can be interpreted as the probability that an insurer receiving premiums at a constant rate is eventually ruined when subjected to a renewal arrival process of iid claims. When the claim distribution is heavy-tailed, the resulting calculation is exactly of the type discussed in this paper. Statistical evidence suggests that such heavy-tailed distributions frequently arise in practice and are a convenient vehicle for capturing many of the key stylized features that are present in observed claim sizes (see, for example, Embrechts, Kluppelberg, and Mikosch (1997) and Adler et al (1998)).

The first efficient rare event simulation algorithm for the tail of M was suggested by Siegmund (1976), who was motivated by the first passage time interpretation displayed in (1) and its connection to one-sided sequential probability ratio tests in the context of statistical sequential analysis. Siegmund's algorithm applies only to light-tailed rw's and involves an importance distribution corresponding to a rw with state-independent increments. It is worth noting that the increment distribution of the random walk under this algorithm is state-independent. Our proposed efficient algorithm is consistent with recent results of Bassamboo, Juneja and Zeevi (2006), who show that no state independent efficient importance sampling algorithm for computing (1) can exist in the (regularly varying) heavy-tailed setting. Another key feature that is present in the light-tailed context is the ability to fully leverage the existing theory of large deviations. A complicating factor in the heavy-tailed setting is that the large deviations literature is not applicable to such problems. Asmussen, Binswanger and Hojgaard (2000) provide a number of examples and counterexamples to illustrate the additional difficulties that arise in the heavy-tailed environment.

As noted above, rare-event simulation algorithms for heavy-tailed distributions have been previously developed in the context of the $M/G/1$ queue. The first logarithmically efficient simulation algorithm for estimation of (1) was given in Asmussen and Binswanger (1997) and was based on the idea of conditional Monte Carlo (and not importance sampling). Logarithmic efficiency for their algorithm was established for regularly varying tails and was shown to fail for Weibull-type heavy tails. Subsequently, Asmussen et al (2000) developed simulation estimators for the $M/G/1$ queue based on importance sampling ideas that are provably efficient for both regularly varying and Weibull-type tails. Juneja and Shahabuddin (2002) also developed efficient importance sampling schemes based on a suitable twisting of the $M/G/1$ service time distribution's hazard rate. More recently, Asmussen and Kroese (2005) proposed other logarithmically efficient importance sampling algorithms for the $M/G/1$ queue that seem to have excellent performance in practice. In addition, they developed a conditional Monte Carlo estimator that is strongly efficient (in the sense that the coefficient of variation is bounded as a function of x) for both regularly varying tails and certain Weibull type heavy-tails. Dupuis, Leder and Wang (2006) proposed a state-dependent importance sampling algorithm that is strongly efficient for a regularly varying $M/G/1$ queue. All the above algorithms take advantage of the fact that the ladder height distribution for the $M/G/1$ queue is explicitly known. In contrast, no such explicit computations are possible for the class of $G/G/1$ models considered here. This significantly complicates both the development and the theoretical analysis of efficient rare-event algorithms for this class of problems.

The paper is organized as follows. Section 2 introduces a general technique to study efficient state-dependent importance sampling algorithms for computing first passage time probabilities of general state-space Markov chains and recovers Siegmund's algorithm as a direct application of the basic ideas underlying our procedure. Section 3 introduces the precise technical assumptions under which we develop our methodology and provides the proof of strong efficiency for our importance sampling estimator. In Section 4 we estimate the computational complexity, measured by the number of operations, required to generate a sample using the suggested procedure and also discuss important issues arising in its efficient implementation. Additional practical observations, some results on simulation experiments and plots are given in our final section.

2 Efficient Importance Samplers for Exit Probabilities

The problem of computing the level crossing probability (1) can be viewed as a special case of computing an exit probability. To be specific, let $Y = (Y_n : n \geq 0)$ be a $\mathcal{X}$-valued Markov chain (with stationary transition probabilities) and let $P_y(\cdot)$ and $E_y(\cdot)$ be the probability distribution and expectation operator on the path-space of Y, conditional on $Y_0 = y$. For $B \subseteq \mathcal{X}$, let $T = \inf\{n \geq 0 : Y_n \in B\}$ be the exit time from B^c. For $A \subseteq B$, the probability $u^*(y) = P_y(Y_T \in A, T < \infty)$ is called an "exit probability" (All the sets considered here are assumed measurable). Note that the level crossing probability (1) is the special case in which Y is given by the rw $(S_n : n \geq 0)$, $\mathcal{X} = [-\infty, \infty)$, $B = \{-\infty\} \cup (b, \infty)$, $A = (b, \infty)$ and $y = 0$. Because of the translation invariance of rw, studying this problem as $b \nearrow \infty$ is equivalent to fixing $B = \{-\infty\} \cup [0, \infty)$, $A = (0, \infty)$, setting $y = -b$ and letting $b \nearrow \infty$. With B and A fixed in this way, our goal is to efficiently compute $u^*(-b)$ as $b \nearrow \infty$. This reformulation of the problem will form the basis of our analysis in the remainder of the paper.

The following result is easily proved (see, for example, Meyn and Tweedie (1993)).

Proposition 1 *The function* $u^* = (u^*(y) : y \in B^c)$ *is the minimal non-negative solution to*

$$u(y) = \int_{\mathcal{X}} P_y(Y_1 \in dz)\, u(z)\,, \quad y \in B^c$$

subject to the boundary conditions that $u(z) = 1$ *for* $z \in A$ *and* $u(z) = 0$ *for* $z \in B \cap A^c$.

As mentioned in the Introduction, the zero-variance importance distribution for computing $u^*(y)$ is that associated with the conditional distribution $P_y(\cdot|\, Y_T \in A, T < \infty)$. Let $\mathcal{F}_n = \sigma(Y_0, ..., Y_n)$ for $n \geq 0$. Our next result characterizes this conditional distribution.

Theorem 1 *Suppose that* $u^*(y) > 0$ *for* $y \in B^c$. *Then, for each non-negative* $\mathcal{F}_T$ *-measurable rv* Λ,

$$E_y[\Lambda|\, Y_T \in A, T < \infty] = E_y^* \Lambda,$$

where $E_y^*(\cdot)$ *is the expectation operator under which* Y *is a Markov chain having one-step transition kernel*

$$P^*(y, dz) = P_y(Y_1 \in dz)\, \frac{u^*(z)}{u^*(y)}.$$

for $y \in B^c$, $z \in \mathcal{X}$.

Proof. Note that $I(T = n)\Lambda = \lambda_n(Y_0, ..., Y_n)$ for some (measurable) function $\lambda_n : \mathcal{X}^{n+1} \to [0, \infty)$. Therefore,

$$\begin{aligned}
& \frac{E_y[\Lambda; T = n, Y_T \in A, T < \infty]}{u^*(y)} \\
= & \int_{B^c \times ... \times B^c \times A} \frac{\lambda_n(y, z_1, ..., z_n)\, u^*(z_n)\, P(y, dz_1) ... P(z_{n-1}, dz_n)}{u^*(y)} \\
= & \int_{B^c \times ... \times B^c \times A} \lambda_n(y, z_1, ..., z_n)\, \frac{P(y, dz_1)\, u^*(z_1)}{u^*(y)} \frac{P(z_1, dz_2)\, u^*(z_2)}{u^*(z_1)} \\
& ... \frac{P(z_{n-2}, dz_{n-1})\, u^*(z_{n-1})}{u^*(z_{n-2})} \frac{P(z_{n-1}, dz_n)\, u^*(z_n)}{u^*(z_{n-1})} \\
= & E_y^*[\Lambda; T = n].
\end{aligned}$$

Summing over n, we conclude that

$$E[\Lambda|\, Y_T \in A, T < \infty] = E^*[\Lambda; T < \infty].$$

Letting $\Lambda = 1$ establishes that $P_y^*(T < \infty) = 1$, proving the result. ∎

This theorem makes clear that the zero-variance importance sampling distribution for computing (1) corresponds to a random walk in which the increments have a state-dependent distribution. The above result suggests that a good importance sampling distribution can be obtained by simulating Y under transition dynamics that closely approximate those induced by the zero-variance importance distribution's transition kernel P^*.

Suppose that Q is the Markov transition kernel chosen by the simulator to compute the exit probability $u^*(y) = P_y(Y_T \in A, T < \infty)$ via importance sampling. Assume that $(Q(y, dz) : y, z \in B^c \cup A)$ can be represented as

$$\begin{aligned} Q(y, dz) &= r(y,z)^{-1} P_y(Y_1 \in dz) I(y \in B^c, z \in B^c \cup A) \\ &\quad + \delta_y(dz) I(y \in A, z \in A) \end{aligned}$$

for some positive function $r(\cdot)$. Note that

$$\begin{aligned} & P_y(Y_T \in A, T = n) \\ = \;& E_y^Q[I(Y_T \in A, T = n) \prod_{j=1}^{T} r(Y_{j-1}, Y_j)], \end{aligned}$$

where $E_y^Q(\cdot)$ is the expectation operator under which Y evolves according to the transition kernel Q, conditional on $Y_0 = y$. Summing over n, we conclude that u^* can be represented as

$$u^*(y) = E_y^Q[I(T < \infty) \prod_{j=1}^{T} r(Y_{j-1}, Y_j)].$$

An important step in any theoretical analysis of the estimator

$$R = I(T < \infty) \prod_{j=1}^{T} r(Y_{j-1}, Y_j) \tag{4}$$

is to bound its variance. The variance, conditional on $Y_0 = y$, is given by $s^*(y) - u^*(y)^2$, where $s^*(y) = E_y^Q R^2$. Since only $s^*(\cdot)$ depends on the choice of the importance distribution, we focus on bounding this quantity.

Theorem 2

i) The function $s^* = (s^*(y) : y \in B^c)$ is the minimal non-negative solution to

$$s(y) = \eta(y) + \int_{B^c} K(y, dz) s(z),$$

for $y \in B^c$, where

$$\begin{aligned} \eta(y) &= \int_A r(y,z) P_y(Y_1 \in dz), \\ K(y, dz) &= r(y,z) P_y(Y_1 \in dz), \end{aligned}$$

for $y, z \in B^c$

ii) The function s^* *is given by*

$$s^* = \sum_{n=0}^{\infty} K^n \eta,$$

where $K^n(y, dz) = \int_{B^c} K^{n-1}(y, dy_1) K(y_1, dz)$ *for* $n \geq 1$, $K^0(y, dz) = \delta_y(dz)$ *and* $(K^n \eta)(y) = \int_{B^c} K^n(y, dz) \eta(z)$.

iii) Suppose that $h = (h(y) : y \in B^c)$ *is a finite-valued non-negative function for which*

$$(Kh)(y) \leq h(y) - \eta(y) \tag{5}$$

for $y \in B^c$. *Then,* $s^*(y) \leq h(y)$ *for* $y \in B^c$.

Proof. Part ii) follows by expanding $E_y^Q[R^2 I(T = n)]$ and summing over n using Fubini's theorem. Part i) follows easily from ii).

For part iii), first note that Kh must be finite-valued by virtue of (5). Induction based on applying K^n to both sides of (5) establishes that $K^n h$ is finite-valued for $n \geq 1$. By applying $K^n h$ to (5) and using the fact that $K^n h$ is finite-valued for $n \geq 1$, we conclude that $K^n \eta \leq K^n h - K^{n+1} h$ for $n \geq 0$. Summing over $0 \leq n \leq m$ and using the non-negativity of h, we obtain the bound

$$\sum_{n=0}^{m} K^n \eta \leq h - K^{m+1} h \leq h.$$

The result follows by sending $n \nearrow \infty$ and using part iii). ∎

We call the function $h(\cdot)$ a Lyapunov function and refer to bounds based on part iii) of Theorem 2 as Lyapunov bounds on the second moment.

Returning to the exit probability computations, suppose that $v = (v(y) : y \in \mathcal{X})$ is chosen by the simulator to be a good approximation to $u^* = (u^*(y) : y \in \mathcal{X})$. In view of Theorem 2 above, it is then natural to consider simulating Y via the transition kernel

$$Q(y, dz) = P(y, dz) \frac{v(z)}{w(y)}, \tag{6}$$

(for $y \in B^c$, $z \in B^c \cup A$), where $w(y)$ is the normalization constant given by

$$w(y) = \int_{B^c \cup A} P(y, dz) v(z)$$

(assumed to be finite). In this case, $r(y, z) = w(y) / v(z)$. The following result provides a Lyapunov bound on the second moment $s^*(\cdot)$ that is specifically suited to this setting.

Proposition 2 *Assume that $w(y) > 0$ for $y \in B^c$ and suppose that there exists a finite-valued function $h : B^c \cup A \longrightarrow [\varepsilon, \infty)$ satisfying*

$$w(y) \int v(z) h(z) P(y, dz) \leq h(y) v(y)^2, \tag{7}$$

for $y \in B^c$. If $h(z) \geq 1$ for $z \in A$ and $v(z) \geq \kappa > 0$ for $z \in A$, then $s^*(y) \leq \varepsilon^{-1}\kappa^{-2} v(y)^2 h(y)$.

Proof. Put $\widetilde{h}(\cdot) = \kappa^{-2} h(\cdot) v^2(\cdot)$ and note that (7) is equivalent to assuming that

$$\left(K\widetilde{h}\right)(y) \leq \widetilde{h}(y) - \kappa^{-2} w(y) \int_A P(y, dz) w(y) h(z) \tag{8}$$

for $y \in B^c$. But

$$\begin{aligned} \eta(y) &= \int_A P(y, dz) \frac{w(y)}{v(z)} \leq \int_A \kappa^{-2} P(y, dz) w(y) v(z) \\ &\leq \kappa^{-2} w(y) \varepsilon^{-1} \int_A P(y, dz) v(z) h(z), \end{aligned}$$

so that (8) implies that

$$\left(K\widetilde{h}\right)(y) \leq \widetilde{h}(y) - \eta(y)$$

for $y \in B^c$. We now apply part iii) of Theorem 2 to complete the proof. ■

Suppose that $v(\cdot)$ has been chosen by the simulator to be within a constant multiple of $u^*(\cdot)$, as occurs whenever $v(\cdot)$ has the same asymptotic behavior as $u^*(\cdot)$. In this case, it follows that the importance sampling algorithm based on $r(y, z) = w(y)/v(z)$ has bounded relative variance (i.e. the ratio of the variance to the square of $u^*(x)$) across B^c whenever the function h of Proposition 2 can be chosen to be bounded. On the other hand, if h grows at a suitable rate (e.g. $h(y) = |\log(v(y))|^{1/2}$), the logarithmic efficiency of the importance sampler can be assured.

To illustrate, consider the problem of estimating

$$u^*(-b) = P(\tau(0) < \infty | S_0 = -b)$$

for $b > 0$ in the light-tailed setting. In particular, suppose that there exists a positive root θ^* of $E[\exp(\theta^* X_1)] = 1$ for which $E[X_1 \exp(\theta^* X_1)] < \infty$. If X_1 is non-lattice, then it is known that

$$u^*(-b) \sim c \exp(-\theta^* b)$$

for some positive constant c; see, for example, Asmussen (2003), p. 365. The natural choice for v is, of course, $v(z) = \exp(\theta^* z)$, in which case $w(z) = \exp(\theta^* z)$. If we put $h(y) = 1$ for $y \in \mathbb{R}$, Proposition 2 applies, yielding the bound

$$s^*(-b) \leq \exp(-2\theta^* b).$$

Hence, this importance sampling algorithm (which is precisely the one proposed by Siegmund (1979)) is strongly efficient.

3 Elements of Our Algorithm for Heavy-tailed rw's

We shall explore how to adapt the ideas discussed in the previous sections to the case of a random walk with heavy-tailed increment distributions. We need the following definitions. Set $X^{+}=\max(X,0)$ and $X^{-}=\max(-X,0)$.

Definition 1 A non-negative rv Z is said to be subexponential if

$$P(Z_1+Z_2>t)\sim 2P(Z>t),$$

as $t\nearrow\infty$ where Z_1 and Z_2 are independent copies of Z. A rv X is said to be subexponential if X^{+} is subexponential.

Definition 2 A non-negative rv Z belongs to the family S^{*} if

$$2EZP(Z>t)\sim\int_0^t P(Z>t-s)\,P(Z>s)\,ds$$

as $t\nearrow\infty$. In addition, a rv X is in S^{*} if X^{+} is in S^{*}.

Definition 3 A rv X is said to possess a long tail if for every constant $a\in\mathbb{R}$

$$P(X>t+a)\sim P(X>t)$$

as $t\nearrow\infty$.

It can be shown that if Z is in S^{*} then it must be subexponential. Also, any subexponential rv possesses a long tail. The class S^{*} of random variables includes as particular cases regularly varying and Weibull-type distributions among many others. For more on the specific properties of various types of heavy-tailed distributions see Embrechts et al (1997) Section 1.4.

The following assumption will be imposed throughout the rest of the paper.

Assumption A Assume that X_n^{+} belongs to S^{*}, that is

$$2EX_n^{+}P(X_n>t)\sim\int_0^t P(X_n>t-s)\,P(X_n>s)\,ds$$

as $t\nearrow\infty$.

If X belongs to S^{*} then both the distribution of X and its integrated tail

$$\int_x^{\infty}\frac{P(X>s)}{EX^{+}}ds$$

are subexponential (see Asmussen (2003), Section 10.9). Under assumption A, it is known (see, for example, Asmussen (2003), p. 296) that

$$u^*(-b) = P(\tau(0) < \infty | S_0 = -b) \sim \frac{-1}{EX} \int_b^\infty P(X > t)\, dt \tag{9}$$

as $b \nearrow \infty$.

The natural strategy is to use this approximation to construct an appropriate importance sampling transition kernel $Q(x, dy)$ (defined in (6)) by means of a function $v(\cdot)$ that mimics the behavior of $u^*(\cdot)$.

An important estimate in the efficiency analysis of our importance sampling scheme involves the behavior of $v(y) - w(y)$ as $y \searrow -\infty$, where $w(y) = Ev(y + X)$. As we indicated earlier, if one selects $v = u^*$ then the difference $v(y) - w(y)$ vanishes. Thus, it is natural to expect that the asymptotic behavior of this difference will play an important role in the performance of the importance sampling estimator. As we shall see, in order to guarantee strong efficiency of the importance sampling estimator it suffices to select $v(\cdot)$ so that $v(y) - w(y) = o(P(X > -y))$ as $y \searrow -\infty$.

Recent estimates by Borovkov and Borovkov (2001) under regularly varying or semiexponential assumptions provide asymptotes to $u^*(y)$ that hold with an error of order $o(P(X > -y))$ as $y \searrow -\infty$. Under these assumptions, Borovkov and Borovkov (2001) add an additional term to (9) of order $O(P(X > -y))$ to the approximation (9) which yields an error rate $o(P(X > -y))$ as $y \searrow -\infty$.

Given the form of (9), it may be surprising at first sight that making use only of approximation (9) and assuming only that the distribution of X belongs to the class S^* one can easily construct $v(\cdot)$ that actually achieves an error of order $o(P(X > -y))$ for the difference $v(y) - w(y)$ as $y \searrow -\infty$. In fact, as we shall prove in our next proposition, $v(-t)$ can be defined as the tail probability of a non-negative random variable Z such that

$$P(Z > t) = \min[-(EX)^{-1} \int_t^\infty P(X > s)\, ds, 1] \tag{10}$$

for $t > 0$ (this may imply $P(Z = 0) > 0$). Then, we write $v(y) = P(Z > -y)$ for all $y \in \mathbb{R}$. Note that if we could pick $u^* = v$ this would correspond to choosing $Z = M$. Given our representation for $v(\cdot)$ as a tail probability we can write

$$w(y) = E[v(y + X)] = P(X + Z > -y).$$

The next result shows that this choice of $v(\cdot)$ has the indicated convergence rate for the difference $v(y) - w(y)$. However, for the purpose of our efficiency analysis, is the second part of the following result, namely inequality (11), which we shall invoke.

Proposition 3 *Under assumption A,*

$$w(y) - v(y) = o(P(X > -y))$$

as $y \searrow -\infty$. *Consequently, for each* $\gamma \in (0,1)$, *there exists* $a^*(\gamma) \in (-\infty, 0]$ *such that for all* $y \leq a_*(\gamma)$,

$$-\gamma \leq \frac{v(y)^2 - w(y)^2}{P(X > -y)\, w(y)} \tag{11}$$

Proof. We must show that

$$P(X + Z > t) - P(Z > t) = o(P(X > t))$$

as $t \nearrow \infty$. Note that

$$\begin{aligned} P(X + Z > t) &= P(X + Z > t, Z > t) + P(X + Z > t, Z \leq t) \\ &= P(Z > t) - P(X + Z \leq t, Z > t) \\ &\quad + P(X + Z > t, Z \leq t). \end{aligned}$$

First, we will show that as $t \nearrow \infty$,

$$P(X + Z > t, Z \leq t) \sim P(X > t)\, EX^- / (-EX).$$

Let $y_0 = \inf\{t \in \mathbb{R} : P(Z > t) < 1\}$. Then,

$$\begin{aligned} & P(X + Z > t, Z \leq t) \\ = & \frac{-1}{EX} \int_{y_0}^{t} P(X > t - s)\, P(X > s)\, ds + P(X > t - y_0)\, P(Z = y_0). \end{aligned}$$

We now analyze the integral on the right hand side of the previous display,

$$\begin{aligned} & \int_{y_0}^{t} P(X > t - s)\, P(X > s)\, ds \\ = & \int_{0}^{t - y_0} P(X > t - y_0 - s)\, P(X > s + y_0)\, ds \\ = & \int_{0}^{t - y_0} P(X > t - y_0 - s)\, P(X > s)\, ds \\ & + \int_{0}^{t - y_0} P(X > t - y_0 - s)\, [P(X > s + y_0) - P(X > s)] ds. \end{aligned}$$

Let us define by I_1 and I_2 the two last integrals on the right hand side of the display above. Then, assumption A yields

$$\begin{aligned} I_1 &= \int_{0}^{t - y_0} P(X > t - y_0 - s)\, P(X > s)\, ds \\ &\sim 2P(X > t)\, EX^+ \text{ as } t \nearrow \infty. \end{aligned}$$

Now, for the integral I_2, we have

$$\begin{aligned} I_2 &= \int_0^{t-y_0} P(X > t - y_0 - s)\, d\int_s^{s+y_0} P(X > u)\, du \\ &= -\int_0^{t-y_0}\int_{t-y_0-s}^{t-s} P(X > u)\, du P(X \in ds) \\ &\quad + P(X > 0)\int_{t-y_0}^{t} P(X > u)\, du - P(X > t - y_0)\int_0^{y_0} P(X > u)\, du. \end{aligned}$$

Note that

$$\begin{aligned} P(X > t - s)\, y_0 &\leq \\ \int_{t-y_0-s}^{t-s} P(X > u)\, du &= t\int_1^{1+y_0/t} P(X > ut - s - y_0)\, du \\ &\leq P(X > t - s - y_0)\, y_0. \end{aligned}$$

Hence, by virtue of assumption A, we have that as $t \nearrow \infty$,

$$\int_0^{t-y_0}\int_{t-y_0-s}^{t-s} P(X > u)\, du P(X \in ds) \sim P(X > t)\, y_0 P(X > 0).$$

Similarly, we obtain that

$$P(X > 0)\int_{t-y_0}^{t} P(X > u)\, du \sim P(X > t)\, y_0 P(X > 0)$$

as $t \nearrow \infty$ which yields

$$I_2 \sim -P(X > t)\int_0^{y_0} P(X > s)\, ds$$

Combining these estimates, we obtain

$$\begin{aligned} & P(X + Z > t, Z \leq t) \\ \sim\ & (I_1 + I_2)/(-EX) + P(X > t - y_0)\, P(Z = y_0) \\ \sim\ & 2P(X > t)\, EX^+/(-EX) - P(X > t)\int_0^{y_0} P(X > s)\, ds/(-EX) \\ & + P(X > t - y_0)\, P(Z = y_0). \end{aligned}$$

Since

$$P(Z = y_0) = 1 - \frac{1}{(-EX)}\int_0^{y_0} P(X > s)\, ds,$$

we have

$$\begin{aligned} & P(X + Z > t; Z \leq t) \\ \sim\ & P(X > t)\, [2EX^+ + (-EX) - EX^+]/(-EX) \\ =\ & P(X > t)\, EX^-/(-EX). \end{aligned}$$

On the other hand,

$$\begin{aligned}
P\left(X+Z\leq t,Z>t\right) &= -\frac{1}{EX}\int_{t}^{\infty}P\left(X\leq t-s\right)P\left(X>s\right)ds \\
&= -\frac{1}{EX}\int_{-\infty}^{0}P\left(X\leq s\right)P\left(X>t-s\right)ds \\
&\sim P\left(X>t\right)EX^{-}/\left(-EX\right)
\end{aligned}$$

as $t\nearrow\infty$, this yields the proof of the result. ■

The constant a_* that characterizes the region where inequality (11) holds will play an important role in the construction of our algorithm. The bound (11) indicates that on the region $(-\infty,a^*]$ the approximation to the zero-variance change-of-measure based on $v(\cdot)$ is good enough to control the variance of the likelihood ratio in our simulations. Finding a_* can be done numerically or analytically depending on the problem at hand. For implementation, the simulator can choose any value of γ (for instance, $\gamma=1/2$) or optimize the asymptotic upper bound that we shall obtain in Theorem 3, which we now are ready to state and prove.

Consider the importance sampling change-of-measure generated by

$$\begin{aligned}
Q_{a_*}\left(y,dz\right) &= \frac{P\left(y+X\in z+dz\right)v\left(z+a_*\right)}{w\left(y+a_*\right)} \\
&= P\left(y+X\in z+dz|\,Z+X\geq -y-a_*\right).
\end{aligned} \tag{12}$$

Then, we will show that the corresponding estimator defined as

$$R=I\left(\tau\left(0\right)<\infty\right)\prod_{j=1}^{\tau(0)}\frac{w\left(S_{k-1}+a_*\right)}{v\left(S_k+a_*\right)}$$

has bounded relative variance as $S_0=y\searrow-\infty$.

Theorem 3 *Suppose that assumption A is in force. Fix $\gamma\in(0,1)$ and select $a_*=a_*(\gamma)\in(-\infty,0]$ as in (11). Then,*

$$E_y^{Q_{a_*}}R^2\leq\left(1-\gamma\right)^{-1}\kappa\left(a_*\right)^{-2}v\left(y+a_*\right)^2,$$

where $\kappa(a_*)=\inf_{z\geq0}[v(z+a_*)]=P(Z>-a_*)$. Consequently,

$$\overline{\lim}_{b\nearrow\infty}E_{-b}^{Q_{a_*}}[R(b)^2]/P\left(M>b\right)^2\leq\left(1-\gamma\right)^{-1}\kappa\left(a_*\right)^{-2}<\infty.$$

Proof of Theorem 3. Define

$$h\left(y\right)=I\left(y+a_*\leq0\right)+\left(1-\gamma\right)I\left(y+a_*>0\right),$$

We wish to apply Proposition 2 so we must satisfy bound (7), which in our case can be written as

$$w(y+a_*)^{-1} Ev(X+y+a_*) h(X+y) \leq \left(\frac{v(y+a_*)}{w(y+a_*)}\right)^2, \tag{13}$$

for all $y \leq 0$. Here we have used the fact that $h(y) = 1$ for $y \leq 0$. Using the interpretation of $v(\cdot)$ as a tail probability we note that bound (13) can be expressed, for all $y \leq 0$, as

$$E(h(X+y) - 1 | X + Z > -y - a_*) \leq \frac{v(y+a_*)^2 - w(y+a_*)^2}{w(y+a_*)^2}.$$

Observe that

$$h(X+y) - 1 = -\gamma I(X \geq -y - a_*).$$

Therefore, it suffices to verify that for all $y \leq 0$,

$$\begin{aligned} & -\gamma P(X > -y - a_* | X + Z \geq -y - a_*) \\ \leq \; & \frac{v(y+a_*)^2 - w(y+a_*)^2}{w(y+a_*)^2}. \end{aligned}$$

However, it follows since $Z \geq 0$ and using the fact that $w(y) = P(X + Z \geq -y)$, that the previous inequality holds if and only if for all $y \leq 0$,

$$-\gamma \leq \frac{v(y+a_*)^2 - w(y+a_*)^2}{P(X > -y - a_*) w(y+a_*)}$$

which is true by definition of a_*. The conclusion of the result follows directly from Propositions 2 and 3 and the fact that $P(M > b) \sim v(-b + a_*)$ as $b \nearrow \infty$. ■

As we indicated before, in terms of implementation, the simulator can just fix any value of γ or find the value of γ that minimizes $(1-\gamma)^{-1} \kappa(a_*(\gamma))^{-2}$ (choosing a value of γ moderately high $\gamma \in [1/2, 2/3]$ would typically be fine).

4 The Algorithm and Complexity Analysis

In order to implement the ideas suggested by the previous analysis we shall first assume that $v(\cdot)$ and $w(\cdot)$ are either available in closed form or can be easily computed numerically. For fixed γ (say $\gamma = 1/2$) set $a_* = a_*(\gamma) \leq 0$ satisfying (11). Note that Theorem 3 suggests choosing a_* as small as possible subject to the constraint (11).

We wish to estimate

$$u^*(-b) = P(\tau(0) < \infty | S_0 = -b),$$

for $b > 0$. Our proposed algorithm proceeds as follows.

Algorithm 1
STEP 1 Initialize $s = -b$, $R = 1$.
STEP 2 Set $s = y$, generate a random variable Y with law

$$P(Y \in t + dt) = P(X \in t + dt | X + Z > -y - a_*),$$

and update $s = y + Y$,

$$\begin{aligned} R &\longleftarrow w(y + a^*) v(s + a^*)^{-1} R \\ &= P(Z + X > -y - a^*) P(Z > -s - a^*)^{-1} R. \end{aligned}$$

STEP 3 If $s > 0$ then return R and STOP, otherwise, go to STEP 2.

The overall complexity of the algorithm involves the relative efficiency and the number of operations required to generate a single replication of R. This involves the time required to terminate the procedure and the complexity (as measured by the number of uniform rv's required) associated with the generation of each increment. In the previous section we proved that our estimator achieves asymptotic optimality in terms of relative efficiency. In this section, we will argue that the number of uniform rv's required to obtain a single realization of R using the previous algorithm is of order $O(b)$ as $b \nearrow \infty$.

Let us first recall some basic facts from extreme value theory (see, for instance, Embrechts et al (1997), Section 3.3). We say that X_1 belongs to the domain of attraction of H (denoted by $X_1 \in MDA(H)$) if there exists a sequence of constants $c_n \geq 0$ and $d_n \in \mathbb{R}$ (for $n \geq 1$) such that

$$c_n^{-1} (\max(X_1, ..., X_n) - d_n) \Longrightarrow H$$

as $n \nearrow \infty$. The random variable H follows a so-called extreme value distribution which, due to the Fisher-Tippett theorem (see Embrechets et al (1997), p. 121), can be of only three types. Only the cases when H has Frechet distribution, given by

$$\Phi_\alpha(x) = \exp\left(-x^{-\alpha}\right) I(x > 0), \quad \alpha > 0$$

or H follows a Gumbel distribution described via

$$\Lambda(x) = \exp(-\exp(-x))$$

are of interest to us. The class $MDA(\Phi_\alpha)$ consists of regularly varying distributions with index $\alpha > 0$ (i.e. $P(X > x) = x^{-\alpha} L(x)$ where $L(\cdot)$ is slowly varying at infinity), whereas $MDA(\Lambda)$ contains other commonly used heavy-tailed distributions, such as log-normal and Weibull.

Using extreme value theory, Asmussen and Kluppelberg (1996) show that under the zero-variance change-of-measure the random walk takes at most $O(b)$ steps to reach the origin. The following result (whose proof is given at the end of the section) provides sufficient conditions to ensure that **Algorithm 1** completes in at most $O(b)$ steps, given $S_0 = -b$.

Proposition 4 *If assumption A is in force and either $X_1 \in MDA(\Phi_\alpha)$for $\alpha > 1$ or $X \in MDA(\Lambda)$, then*

$$E_{-b}^{Q_{a*}} \tau(0) \leq O(b)$$

as $b \nearrow \infty$.

Remark We point out that the previous Proposition remains valid if assumption A holds and, for some $\varepsilon > 0$, the mapping $t \longrightarrow t^{\varepsilon} P(Z > t)$ is eventually decreasing. These conditions are easy to verify and slightly weaker than the MDA assumptions imposed. We framed the above result in terms of extreme value theory in order to establish a connection with well known results in the literature such as Asmussen and Kluppelberg (1996).

In order to complete our complexity analysis, it is important to observe that the execution of STEP 2 of the algorithm involves a one dimensional rare-event type simulation problem. We have assumed that $v(\cdot)$ and $w(\cdot)$ can be easily evaluated. Nevertheless, it could be the case that finding explicitly the distribution of Y in STEP 2 could be difficult or numerically expensive. We shall argue that the variates in STEP 2 can be simulated through a suitable acceptance / rejection scheme. Note, however, that one has to design the scheme in such a way that the acceptance probability remains uniformly bounded (in y) away from zero. By doing this, the generation of the random walk increments in STEP 2 under the importance sampling distribution has uniformly bounded complexity as $b \nearrow \infty$. Consequently, given Proposition 4, the expected number of variates required to run **Algorithm 1** will be of order $O(b)$ as $b \nearrow \infty$.

Typically, acceptance / rejection schemes such as those indicated in the previous paragraph, although not complicated, must be designed based on specific characteristics of the problem at hand. Assume that X has a continuous density $f_X(\cdot)$. STEP 2 of **Algorithm 1** requires sampling a rv Y with density $f_Y(\cdot)$ defined, for $b \geq 0$, as

$$f_Y(z; b) = v(-b + z) f_X(z) / w(-b).$$

The objective is to find an easy to simulate rv $\widetilde{Z}$ with computable density $f_{\widetilde{Z}}(z; b)$ such that for all $z \in \mathbb{R}$

$$f_Y(z; b) \leq p_{acc}(b)^{-1} f_{\widetilde{Z}}(z; b) \tag{14}$$

where the acceptance probability, $p_{acc}(b)$, satisfies $\inf_{b \geq 0} p_{acc}(b) > 0$.

In order to construct $f_{\widetilde{Z}}(\cdot)$ let us assume that $f_X(\cdot)$ is regularly varying. We pick $\theta \in (0,1)$ and define

$$\begin{aligned}
c(b) &= P(X \leq b - \theta b) \frac{P(Z > \theta b)}{P(Z > b)} + \frac{P(X > b - \theta b)}{P(Z > b)}, \\
\lambda_0(b) &= c(b)^{-1} P(X \leq b - \theta b) P(Z > \theta b) / P(Z > b), \\
\lambda_1(b) &= c(b)^{-1} P(X > b - \theta b) / P(Z > b).
\end{aligned}$$

Then, the mixture density

$$\begin{aligned}
f_{\widetilde{Z}}(z; b) &= \lambda_0(b) \frac{f_X(z) I(z \leq b - \theta b)}{P(X \leq b - \theta b)} \\
&\quad + \lambda_1(b) \frac{f_X(z)}{P(X > b - \theta b)} I(z > b - \theta b)
\end{aligned}$$

satisfies

$$f_Y(z; b) \leq mc(b) f_{\widetilde{Z}}(z; b),$$

where

$$m \geq \sup_{b \geq 0} [P(Z > b) / P(Z + X > b)].$$

The acceptance probability using $f_{\widetilde{Z}}(z; b)$ as proposal is $[mc(b)]^{-1}$. Using elementary properties of regularly varying functions it follows that $\inf_{b \geq 0} [c(b) m]^{-1} > 0$.

We conclude the section with a proof of Proposition 4.

Lemma 1 *If assumption A is in force and either* $X_1 \in MDA(\Phi_\alpha)$ *for* $\alpha > 1$ *or* $X \in MDA(\Lambda)$, *then there exists* $t_0 > 0$ *and* $\varepsilon > 0$ *such that for all* $t \geq t_0$

$$E[X | X + Z > t] \geq \varepsilon$$

Proof. The assumptions imply that X_1 and Z must be subexponential. In particular, it follows that $P(X + Z > t) \sim P(Z > t)$ as $t \nearrow \infty$. Thus, it suffices to show that

$$\underline{\lim}_{t \longrightarrow \infty} \left(\frac{-E[X^-; X + Z > t]}{P(Z > t)} + \frac{E[X^+; X + Z > t]}{P(Z > t)} \right) > 0. \tag{15}$$

The Bounded Convergence Theorem implies that

$$\frac{-E[X^-; X + Z > t]}{P(Z > t)} = \int_{-\infty}^{0} s \frac{P(X > t - s)}{P(Z > t)} P(X \in ds) \longrightarrow -EX^-$$

as $t \nearrow \infty$. On the other hand, we have that

$$\begin{aligned} E[X^+; X+Z &> t] = E[X; X+Z > t; X \geq 0] \\ &= E\int_0^\infty I\left(X+Z > t; X \geq 0; X \geq s\right) ds \\ &= \int_0^\infty P\left(X+Z>t; X\geq s\right) ds \\ &= \int_0^{t-y_0} P\left(X+Z>t; X\geq s\right) ds + P\left(Z > t-y_0\right)\left(-EX\right), \end{aligned}$$

where $y_0 = \inf\{t \in \mathbb{R} : P\left(Z > t\right) < 1\}$. Now,

$$\begin{aligned} \int_0^{t-y_0} P\left(X+Z>t; X\geq s\right) ds &= \int_0^{t-y_0} P\left(X+Z>t; s \leq X \leq t-y_0\right) ds \\ &\quad + \left(t-y_0\right) P\left(X > t-y_0\right). \end{aligned}$$

The first integral in the right hand side of the previous display is greater or equal to

$$\int_0^{t-y_0} P\left(Z > t-s\right) P\left(s \leq X \leq t-y_0\right) ds \sim P\left(Z>t\right) EX^+.$$

On the other hand, it follows that

$$tP\left(X>t\right) \geq \int_t^{2t} P\left(X>s\right) ds = \left[P\left(Z>t\right) - P\left(Z>2t\right)\right]\left(-EX\right).$$

Note that if $X_1 \in MDA\left(\Phi_\alpha\right)$ then

$$P\left(Z > 2t\right)/P\left(Z>t\right) \longrightarrow \left(1/2\right)^{\alpha-1} < 1$$

as $t \nearrow \infty$, whereas if $X_1 \in MDA\left(\Lambda\right)$ then X_1 exhibits rapid variation and the previous limit is zero (see Corollary 3.3.32, p. 148 and Definition A3.11, p. 570 in Embrechts et al (1997)). Putting all the previous estimates together (and using the fact that Z has a long tail) we obtain that the limit in (15) is greater or equal to

$$\begin{aligned} & -EX^- - EX + EX^+ - \left(1-\left(1/2\right)^{\alpha-1}\right) EX \\ = & -\left(1-\left(1/2\right)^{\alpha-1}\right) EX > 0 \end{aligned}$$

which is more than we need in order conclude the proof of the lemma. ■

Finally, we provide the proof of Proposition 4.

Proof of Proposition 4. It follows from Lemma 1 and Chebyshev's inequality that there exists $a < 0$ and $\varepsilon > 0$ such that

$$\sup_{y \leq a} E\left(X \middle| X+Z > -y-a\right) > \varepsilon. \tag{16}$$

Now, set $\tau(a) = \inf\{n \geq 1 : S_n > a\}$. It follows from (16) then that on $\{\tau(a) > n\}$, there exists $\varepsilon > 0$ such that

$$E^{Q_{a*}}\left(S_{n+1} | S_n\right) - S_n > \varepsilon$$

and therefore (letting $\min(n, \tau(0)) \triangleq n \wedge \tau(0)$) it is not hard to see that $M_n = S_{n \wedge \tau(0)} - (n \wedge \tau(a))\,\varepsilon$ is a submartingale (under Q_{a_*}). In particular, we obtain that

$$\varepsilon E_y^{Q_{a*}}[n \wedge \tau(a)] \leq E_y^{Q_{a*}} S_{n \wedge \tau(a)} - y \leq -y$$

Finally, the monotone convergence theorem yields

$$E_y^{Q_{a*}} \tau(a) \leq |y| / \varepsilon.$$

On the other hand, we have that

$$\sup_{y \in [a,0]} E_y^{Q_{a*}} \tau(a) \leq 1 - \frac{1}{\varepsilon} \sup_{a \leq y \leq 0} E[X + y; X + y \leq a | X + Z > -y - a_*] < \infty.$$

Therefore, it follows from a geometric trials argument that

$$\begin{aligned} & E_{-b}^{Q_{a*}} \tau(0) \\ \leq\ & E_{-b}^{Q_{a*}} \tau(a) + [\sup_{a \leq y \leq 0} P(X > -a | X + Z > -y - a^*)]^{-1} \sup_{y \in [a,0]} E_y^{Q_{a*}} \tau(a) \\ \leq\ & b \cdot m \end{aligned}$$

for some $m \in (0, \infty)$, which yields the proof of the result. ■

5 Empirical Validation

We shall illustrate the performance of the algorithm empirically via simulation experiments. In particular, we analyze tail probabilities for the steady-state waiting time in the $D/G/1$ queue with unit interarrival time. A given service time V has tail distribution given by

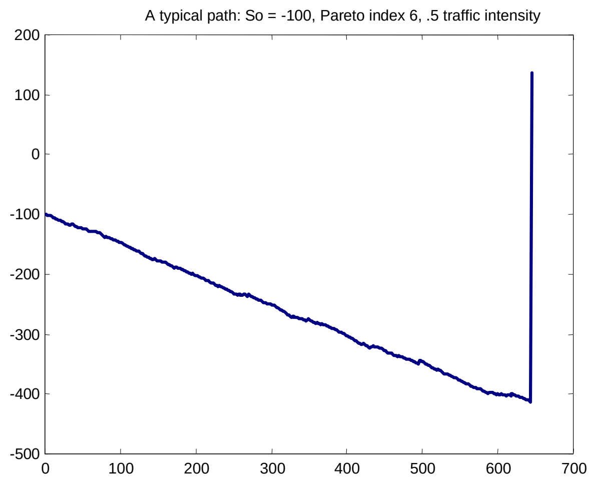
$$P(V > t) = (1 + t/\lambda)^{-p}.$$

The increments X_n's have distribution given by $X = V - 1$ so that

$$EX = \lambda \int_0^1 (1 + t)^{-p} \, dt - 1 = \frac{\lambda}{p - 1} - 1.$$

The traffic intensity is $\rho = \lambda / (p - 1)$ and we assume $\rho < 1$.

The next plot shows a typical path of the algorithm when $p = 6$, $\rho = .5$ and $S_0 = -100$. In this case, $u^*(-100) \approx 8.2 \times 10^{-9}$.



Note the this figure is in agreement with the fact that, in the presence of heavy-tailed claims, a typical sample path leading to ruin (which in this case corresponds to hitting the positive half line) will look like a "regular path" just prior to the ruin which is caused by a single large claim inducing a large deficit at the time of ruin (see Embrechts et al (1997), p. 453).

The following table illustrates the performance of the previous algorithm for a Pareto service time distribution and different traffic intensity values. The column HTA provides $P(Z > -S_0)$ (i.e. the heavy-tailed approximation), IS is the estimator based on our importance sampling scheme using 10,000 replications. We performed several tests increasing the sample size up to 25,000 and the estimators remained very close to those reported below.

$p = 3$	S_0	HAP	IS	CV
$\rho = .8$	-100	9.72e-04	9.01e-04	24.4
	-500	4.05e-05	3.94e-05	6.57
$\rho = .5$	-10	6.94e-03	9.95e-05	.09
	-350	8.07e-06	8.14e-06	.03

Table 1: Numerical Results

Acknowledgement 1 *Thanks to Bert Zwart for pointing out the need for condition S^* in our proof of Proposition 3.*

References

[1] Adler, J., Feldman, R., and Taqqu, M. (Eds.) (1998) *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhauser. Boston.

[2] Asmussen, S. (2003) *Applied Probability and Queues*. Springer-Verlag. New York.

[3] Asmussen, S. and Glynn, P. (2007) *Stochastic Simulation: Algorithms and Analysis.* To appear.

[4] Asmussen, S., and Binswanger, K. (1997) Simulation of ruin probabilities for subexponential claims. Astin Bulletin 27, 297-318.

[5] Asmussen, S., Binswanger, K., and Hojgaard, B. (2000) Rare event simulation for heavy-tailed distributions. Bernoulli 6, Vol. 2, 303-322.

[6] Asmussen, S., and Kluppelberg, C. (1996) Large deviation results for subexponential tails, with applications to insurance risk. Stoch. Proc. Appl. 64, 103-125.

[7] Asmussen, S. and Kroese, D. (2005) Improved algorithms for rare-event simulation with heavy tails. MaPhySto, Research Report No. 3.

[8] Bassamboo, A., Juneja, S., and Zeevi, A. (2006). On the efficiency loss of state-independent importance sampling in the presence of heavy-tails. Operations Research Letters. 34, 521-531.

[9] Borovkov, A. A. (2000) Estimates for the distribution of sums and maxima of sums of random variables without the Cramer condition. Siberian Math. J. 41, 811-848.

[10] Borovkov, A. A. and Borovkov, K. A. (2001) On probabilities of large deviations for random walks. I: Regularly varying distribution tails. Theory Prob. Appl. 49, No. 2, 189-205.

[11] Bucklew, J. (2004) *Introduction to Rare-event Simulation.* Springer, New York.

[12] Dupuis, P. and Wang, H. (2004) Importance sampling, large deviations, and differential games. Stochastics and Stochastics Reports 76, 481-508.

[13] Dupuis, P., Leder, K., and Wang, H. (2006) Notes on importance sampling for random variables with regularly varing tails. Preprint.

[14] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag. New York.

[15] Hammersley, J. and Morton, K. (1954) Poor man's Monte Carlo. Journal of the Royal Stat. Soc, Series B. 16 - 1, p. 23 - 38.

[16] Juneja, S. and Shahabuddin, P. (2002) Simulating heavy-tailed processes using delayed hazard rate twisting. ACM TOMACS, 12 - 2, p. 94 - 118.

[17] Juneja, S. and Shahabuddin, P. (2006) Rare event simulation techniques: An introduction and recent advances. *Handbook on Simulation.* Elsevier. Editors: Shane Henderson and Barry Nelson p. 291-350.

[18] Liu, J. (2001) *Monte Carlo Strategies in Scientific Computing.* Springer-Verlag, New York.

[19] Meyn, S. and Tweedie, R. (1993) *Markov Chains and Stochastic Stability.* http://decision.csl.uiuc.edu/~meyn/pages/book.html.

[20] Rosenbluth, M. and Rosenbluth, A. (1955) Monte Carlo calculation of the average extension of molecular chains. Journal of Chemical Physics. 23 - 2, p. 356 - 359.

[21] Siegmund, D. (1976) Importance sampling in the Monte Carlo study of sequential tests. Ann. Statist. 4, 673-684.