Rare events simulation for heavy-tailed distributions

SØREN ASMUSSEN,[1] KLEMENS BINSWANGER[2] and BJARNE HØJGAARD[3]

[1] *Department of Mathematical Statistics, University of Lund, Box 118, S-221 00 Lund, Sweden. E-mail: asmus@maths.lth.se*
[2] *Swiss Reinsurance Company, Mythenquai 50/60, CH-8022 Zürich, Switzerland. E-mail: Klemens-Binswanger@swissre.com*
[3] *Department of Mathematics, Aalborg University, Fr. Bajersv. 7E, DK-9220 Aalborg Ø, Denmark. E-mail: malik@math.auc.dk*

This paper studies rare events simulation for the heavy-tailed case, where some of the underlying distributions fail to have the exponential moments required for the standard algorithms for the light-tailed case. Several counterexamples are given to indicate that in the heavy-tailed case, there are severe problems with the approach of developing limit results for the conditional distribution given the rare event; this is used as a basis for importance sampling. On the positive side, two algorithms having a relative error which is almost bounded are presented, one based upon order statistics and the other upon a different importance sampling idea.

Keywords: conditional Monte Carlo; importance sampling; large deviations; logarithmic efficiency; $M/G/1$ queue; order statistics; random walk; regular variation; subexponential distribution

1. Introduction

Estimation of small probabilities by simulation is one of the key issues of today's simulation literature. The reason is two-fold: the topic is of major practical importance in areas such as reliability, telecommunications systems and insurance risk; it also presents considerable challenges from a methodological point of view.

The most established approach for rare events simulation is importance sampling via exponential change of measure. The present paper deals with the case where this technique is intrinsically impossible because the required exponential moments do not exist: some underlying distribution G is *heavy-tailed*, for example

- lognormal, the distribution of $e^{\mu+\sigma U}$ where U is standard normal;
- Weibull with decreasing failure rate, $\bar{G}(x) = e^{-x^\beta}$ with $0 < \beta < 1$, where $\bar{G}(x) = 1 - G(x)$ is the tail;
- regularly varying, $\bar{G}(x) = L(x)/x^\alpha$, where L is slowly varying, $L(tx)/L(x) \to 1$ as $x \to \infty$ for all $t > 0$ (this covers Pareto, stable and log-gamma distributions).

We return to a more general mathematical definition in Section 2, but the three examples just given are the key ones.

The relevance of heavy-tailed modelling has recently been argued strongly in areas such as telecommunications and insurance risk. In insurance risk, heavy-tailed behaviour occurs mainly as result of claims with a heavy-tailed distribution, as thoroughly discussed in the recent book by Embrechts *et al.* (1997). In telecommunications, it may arise in a traditional queueing model because heavy-tailed service-time distribution; see Abate *et al.* (1994) for a recent study and references. In a fluid model, it may result from a heavy-tailed distribution of the period of the individual sources – when sufficiently many sources are at the same time, the buffer content will build up, and this build-up period then has a heavy-tailed distribution; see Boxma (1996; 1997), Choudhury and Whitt (1997), Heath *et al.* (1997; 1998), Jelenković and Lazar (1995; 1996) and Rolski *et al.* (1999). This situation can alternatively be viewed as an instance of long-range dependence, for which empirical evidence is presented in Willinger *et al.* (1997; 1998), Crovella and Bestravros (1996) and Cunha *et al.* (1995).

In application areas such as these, explicit or easily computable solutions are only available in light-tailed setting, imposing conditions such as phase-type distributions of claims, service times or periods; see, for example, Asmussen and Rolski (1991) and Asmussen (1992; 1995). Approximations are available in quite a few cases in the heavy-tailed case but are notoriously mostly imprecise, as documented numerically in Abate *et al.* (1994), Asmussen and Binswanger (1997) and Kalashnikov (1999). Thus, simulation may be required even for quite simple problems.

The formal set-up of the paper is that of a family $\{A(x)\}$ of events defined on a probability space $(\Omega, \mathscr{F}, P)$ and rare in the sense that $z(x) = P(A(x)) \to 0$, $x \to \infty$. An *estimator* for $z(x)$ is a random variable $Z(x)$ such that $z(x) = E Z(x)$. The simulation is performed by the Monte Carlo method: produce N independent and identically distributed (i.i.d.) replications $Z_1, \ldots, Z_N$ of $Z = Z(x)$, estimate $z = z(x)$ by the empirical average $(Z_1 + \cdots + Z_N)/N$ and form a confidence interval based upon the empirical variance of the Z_i.

The difficulty in rare events simulation is to produce estimators which have not only a small variance $\mathrm{var}(Z(x))$ but also a small relative error $\sqrt{\mathrm{var}(Z(x))}/z(x)$. Asymptotically, the best performance which has been observed in realistic situations is a bounded relative error in the limit $x \to \infty$. Following the general custom of the literature (see, for example, Sadowsky 1993), we use a slightly weaker criterion,

$$\mathrm{var}(Z(x)) \overset{\log}{\leqslant} z(x)^2, \tag{1.1}$$

and call $\{Z(x)\}$ *asymptotically efficient* if (1.1) holds. Here $a(x) \overset{\log}{\leqslant} b(x)$ means that a, b are functions converging to 0 in such a way that $\underline{\lim}|\log a(x)|/|\log b(x)| \geqslant 1$. We refer to Heidelberger (1995) or Asmussen and Rubinstein (1995) for surveys of the area of rare events simulation with particular emphasis on such complexity issues.

In the light-tailed case, many examples of estimators statisfying (1.1) are known. A general method is importance sampling, simulating not $Z(x) = I(A(x))$ from P as in the crude Monte Carlo method but rather $Z(x) = I(A(x))L$ from $\tilde{P}$, where $\tilde{P}$ is a different probability measure on $(\Omega, \mathscr{F})$ and L is the likelihood ratio in the sense that

$$P(B) = \tilde{E}[L; B], \qquad \text{for all } B \subseteq A(x). \tag{1.2}$$

How do we choose $\tilde{P}$? It follows by general results on importance sampling that the optimal $\tilde{P}$ is the conditional distribution $P^{(x)} = P(\cdot|A(x))$ given the rare event (in fact, this choice even gives a zero variance). However, this choice is infeasible because $L = I(A)z(x)$ involves $z(x)$, which is unknown. Nevertheless, one obtains some form of guidance: if $\tilde{P}$ is not too far from $P^{(x)}$, we can hope that the variance is small. Thus, one looks for a probability measure $\tilde{P}^{(x)}$ which, in an appropriate asymptotic sense, is close to $P^{(x)}$. This can often be obtained by a large-deviations argument, and the result is typically that $\tilde{P}^{(x)}$ is chosen via an exponential change of measure; see Section 3 for a simple example.

In the heavy-tailed case, exponential change of measure is intrinsically impossible, and also the relevant results on the asymptotic form of $P^{(x)}$ are recent (see the survey in Section 2). In view of this, it is not surprising that a more sophisticated approach to rare events simulation for heavy-tailed distributions is only in its infancy at present. The present paper surveys the current state of the area by presenting some asymptotically efficient algorithms and discussing some of the difficulties. We do not claim to have any general approach with the promise of eventually applying to complex practical situations. Rather, it is our hope that the paper will stimulate further research in the area.

In Section 2, we give the necessary background on heavy-tailed distributions, introducing the class $\mathscr{S}$ of *subexponential distributions*. We also briefly summarize some of the known results on rare events behaviour which are relevant for discussing the approach via the conditional distribution given the rare event. In Section 3, we consider the rare event $\{S_n > x\}$, where S_n is a sum of i.i.d. positive random variables. Two asymptotically efficient algorithms are presented, one from Asmussen and Binswanger (1997) which uses a conditional Monte Carlo idea and order statistics, and a new one using importance sampling. Section 4 then contains a discussion of the method of identifying $P^{(x)}$ and simulating $\tilde{P}^{(x)}$ which is the obvious route inspired by the light-tailed case. Our analysis contains, however, several counterexamples and indicates severe difficulties with this approach. The more technical proofs, basically variance calculations, are deferred to the Appendix.

2. Heavy-tailed distributions and their rare events behaviour

2.1. Subexponential distributions

Let G be a distribution on $(0, \infty)$ and let $Y_1, Y_2 \ldots$ be i.i.d. with common distribution G. An established common framework for the heavy-tailed setting is that $G \in \mathscr{S}$, the subexponential class introduced by Chistyakov (1964) and defined by the requirement

$$\lim_{x \to \infty} \frac{P(Y_1 + \cdots + Y_n > x)}{P(Y_1 > x)} = n \tag{2.1}$$

for all n (actually, it suffices that (2.1) holds for $n = 2$). Since, for any distribution G, $P(\max(Y_1, \ldots, Y_n) > x) \sim nP(Y_1 > x)$ (here and in the following, $\sim$ means that the ratio is one in the limit $x \to \infty$), this formal definition contains what is the crucial intuition behind

heavy tails: *the only way the sum can get large is by one of the summands getting large* (in contrast, in the light-tailed case all summands are large if the sum is so). See Feller (1971) and Bingham *et al.* (1987) for the regularly varying case, and Embrechts *et al.* (1997) and Goldie and Klüppelberg (1998) for the general subexponential class.

For a distribution F on the whole line, we write $F \in \mathscr{S}$ if $\bar{F}(x) \sim \bar{G}(x)$, $x \to +\infty$, for some $G \in \mathscr{S}$ which is concentrated on $(0, \infty)$.

A property that is often used is

$$\lim_{x \to \infty} \frac{\bar{G}(x + y)}{\bar{G}(x)} = 1 \tag{2.2}$$

uniformly for y in compact sets.

>The class $\mathscr{S}$ splits naturally into two, the class of distributions with a regularly varying tail and the rest. This classification is closely related to extreme value theory. Recall that $G \in MDA(H)$ (maximum domain of attraction) if there exist constant a_n, b_n such that $a_n(\max(Y_1, \ldots, Y_n) - b_n) \xrightarrow{\mathscr{D}} H$. Since subexponential distributions have unbounded support, H must be either the Fréchet law Φ_α or the Gumbel law Λ. Further, $G \in MDA(\Phi_\alpha)$ is equivalent to regular variation with index $-\alpha$, and if G is subexponential but not regulary varying, then $G \in MDA(\Lambda)$ subject to some weak smoothness conditions; see Goldie and Resnick (1988). A further fundamental result (a sharpening of (2.2)) is that then there exist constants $\gamma(x) \to \infty$ (which can be taken as $E[Y - x | Y > x]$, the mean residual lifetime) such that

$$P\left(\frac{Y - x}{\gamma(x)} > y \mid Y > x\right) = \frac{\bar{G}(x + \gamma(x)y)}{\bar{G}(x)} \to P(V_\alpha > y), \tag{2.3}$$

where V_α is Pareto with mean one when $G \in MDA(\Phi_\alpha)$ with $\alpha < \infty$ and standard exponential when $G \in MDA(\Lambda)$ ($\alpha = \infty$). This fact is often the crucial one for rare events behaviour rather than the equivalent characterization $G \in MDA(H)$.

2.2. Some basic probabilistic results

The framework is that of a random walk $\{S_n\}_{n=0,1,2,\ldots}$ in discrete time, $S_n = X_1 + \cdots + X_n$ where $X_1, X_2, \ldots$ are i.i.d. with common distribution F and mean $\mu < 0$. Also the reflected version $\{W_n\}_{n=0,1,2,\ldots}$ given by $W_0 = 0$,

$$W_n = (W_{n-1} + X_n)^+ = S_n - \min_{k=0,\ldots,n} S_k$$

is of interest because if X_n is the difference between a service time and an independent inter-arrival time, then $\{W_n\}$ is the $GI/G/1$ actual waiting process. For this reason, rare events for $\{S_n\}$ are benchmark examples in the simulation literature. A random variable with the limiting stationary distribution is denoted by W_∞ and has the same distribution as $M = \max_{n=0,1,2,\ldots} S_n$; see Asmussen (1987, Ch. III.7). We write $\tau(x) = \inf\{n > 0 : S_n > x\}$.

Consider, first, the rare event $A(x) = \{M > x\}$, where x is large, or, equivalently, the event of a large steady-state waiting time. In insurance risk, $A(x)$ can also be interpreted as

the event of ruin with initial reserve x. It is a classical fact (Embrechts and Veraverbeke 1982, and references therein) that

$$P(M > x) = P(W_\infty > x) = P(\tau(x) < \infty) \sim \frac{1}{|\mu|} \int_x^\infty \bar{F}(y)\, dy, \qquad (2.4)$$

where $\bar{F}(y) = 1 - F(y)$, whenever $\int_x^\infty \bar{F}(y)\, dy$ is a subexponential tail; in practice, this is the case at the same time as $F \in \mathscr{S}$ but none of the conditions imply the other.

How does $\{S_n\}$ attain a large value x? That is, what does the conditional distribution $P^{(x)}$ given $M > x$ look like? The heavy-tail intuition predicts exceedances of level x to occur as the consequence of one big jump. Thus, what remains is to find the asymptotic distribution of the time of the big jump and to make precise the fact that the random walk behaves in its typical way except for this one big jump. The answer to these questions was recently obtained by Asmussen and Klüppelberg (1996). In particular:

Proposition 2.1. *Assume that* $F \in MDA(H) \cap \mathscr{S}$. *Then the* $P^{(x)}$-*distribution of* $\tau(x)/\gamma(x)$ *converges weakly to that of* $V_\alpha/|\mu|$.

Remark 2.1. Asmussen and Klüppelberg (1996) also give a number of results which show that the large-deviations path $\{S_n\}_{n=0,1,\dots,\tau(x)}$ in an appropriate sense is 'typical' of the random walk. For example, the empirical distribution function $\tau(x)^{-1} \sum_1^{\tau(x)-1} I(X_n \leqslant \cdot)$ converges in $P^{(x)}$-probability to $F(\cdot)$.

We will discuss a simulation algorithm inspired by these results in Section 4. For light tails, the large-deviations path is described in Asmussen (1982) – see also Anantharam (1988) – and explains the efficiency of the exponential change of measure in Siegmund (1976).

Another rare event related to $\{S_n\}$ which is often studied is buffer overflow within a cycle. The precise definition is $A(x) = \{M_C > x\}$, where

$$C = \min\{k = 1, 2, \dots : W_k = 0\} = \min\{k = 1, 2, \dots : S_k \leqslant 0\}$$

is the regenerative cycle and $M_C = \max_{k=0,\dots,C-1} W_k$ the cycle maximum.

The asymptotic form of the rare event probability was given in Iglehart (1972) for light tails, but was only recently found in the heavy-tailed case (Asmussen 1998; see also Heath *et al.* 1997):

Proposition 2.2. *If* $F \in \mathscr{S}$, *then* $P(M_C > x) \sim EC \cdot \bar{F}(x)$, $x \to \infty$.

This is intuitive since the probability of making a big jump above x is $\bar{F}$ and the expected time the process has to do it in a cycle is EC. The proof in Asmussen (1998) is, however, surprisingly tricky. It shows also that indeed the level $S_{\tau(x)-1}$ before the big jump is $O(1)$.

Remark 2.2. Some further relevant references in the area of approximations and conditioned limit theorems for rare events behaviour in the presence of heavy tails are Asmussen and Højgaard (1996), Asmussen *et al.* (1999), Durret (1980) and Klüppelberg and Stadtmüller (1998), but there are still many open problems. We mention in particular that steady-state tail asymptotics is not at present available for the following random variables:

1. The buffer content in a fluid model with i.i.d. on–off sources with a heavy-tailed on period. All of the references Boxma (1996; 1997), Choudhury and Whitt (1997), Heath *et al.* (1997; 1998), Jelenković and Lazar (1995; 1996) and Rolski *et al.* (1998) impose some special condition which allows one reduction to reduce to random walks but which are not the most natural ones from the point of view of fluid models. The problem is that in general for a build-up of the fluid queue, it is required that more than one source is on, so that one expects that for large buffer contents to arise, several sources must have simultaneous long on periods. This heuristic easily gives a lower bound (see Choudhury and Whitt 1997), which one expects to be asymptotically correct, but no proof of this is available: the problem is to show that simultaneous long on periods are asymptotically the only way that a build-up can occur.

2. The waiting time in a $GI/G/c$ queue ($c > 1$) with heavy-tailed service distribution. The problem is similar to that for the fluid queue, that service times at more than one server must be long at the same time for a build-up (again, one can easily derive a lower bound).

3. Sojourn times in a queueing network with heavy-tailed service times distributions at one or several nodes. A conjecture is given in Asmussen (1996).

3. Heavy-tailed sums

In this section, we consider a problem which is simpler than the random walk problems considered in Section 2: that of producing a simulation estimator which is efficient for $z(x) = P(S_n > x)$ in the limit $x \to \infty$ (n is fixed!), where $X_1, \ldots, X_n$ are non-negative and heavy-tailed with common distribution F. The crude Monte Carlo estimator is $Z_1(x) = I(S_n > x)$ and has variance $F(x)\bar{F}(x) \sim \bar{F}(x)$.

The light-tailed case. Earlier, Bucklew *et al.* (1990) considered the problem of estimating $P(S_n > x)$ in a rare events setting. However, in their set-up $x = n(\mu + \varepsilon)$, where $\mu = \mathbb{E}X$ and the limit is $n \to \infty$, so that the rare event occurs as a consequence of large deviations in the law of large numbers. The algorithm is based upon the saddlepoint technique (see Jensen 1995, for a survey): let $d\tilde{F}/dF = e^{\theta x}/\hat{F}[\theta]$, where $\hat{F}$ is the moment generating function of F and $\theta = \theta(\epsilon)$ is chosen such that $\tilde{\mathbb{E}}X = \mu + \epsilon$ (i.e. $\tilde{\mathbb{E}}S_n = x$), and use the estimator

$$Z_2(x) = e^{-\theta S_n}\hat{F}[\theta]^n I(S_n > x),$$

where $X_1, \ldots, X_n$ are simulated as i.i.d. with distribution $\tilde{F}$. This algorithm is asymptotically efficient in the sense of (1.1), and, in fact, this particular choice of measure is the unique one with $X_1, \ldots, X_n$ i.i.d. for which (1.1) holds.

If instead, as in the situation we consider here, n is fixed and $x \to \infty$, the exponential change of measure with $\tilde{\mathbb{E}}S_n = x$ quite often also works well. For example, one has asymptotic efficiency if the tail of F is either asymptotically exponential, $\bar{F}(x) \sim ce^{-\delta x}$, or normal-like, $\bar{F}(x) \sim ce^{-\delta z^2}$ (we know of no precise reference for these results, but they can be verified by straightforward variance calculations).

The conditional Monte Carlo idea. The first and obvious idea when using conditional Monte Carlo is to condition upon $X_1, \ldots, X_{n-1}$, which leads to

$$Z_3(x) = \mathrm{P}(S_n > x | X_1, \ldots, X_{n-1}) = \bar{F}(x - S_{n-1}).$$

Thus, we generate only $X_1, \ldots, X_{n-1}$. As a conditional Monte Carlo estimator, $Z_3(x)$ has a smaller variance than $Z_1(x)$. However, asymptotically it represents no improvement: the variance is of the same order of magnitude $\bar{F}(x)$. To see this, just note that

$$\mathrm{E} Z_3(x)^2 \geqslant \mathrm{E}[\bar{F}(x - S_{n-1})^2; X_1 > x] = \mathrm{P}(X_1 > x) = \bar{F}(x)$$

(here we used the fact that, by positivity of the X_i, $S_{n-1} > x$ when $X_1 > x$, and that $\bar{F}(y) = 1$, $y < 0$).

The reason why this algorithm does not work well is that the probability of a single X_i becoming large is too big. This problem has arisen in many of the algorithms we have tried to implement: it quite often happens that the probability that S_n 'by itself' exceeds x is so large that its contribution to the variance is $O(\bar{F}(x))$ rather than $O(\bar{F}(x))^2$ as one would have hoped.

An algorithm involving order statistics. Asmussen and Binswanger (1997) suggested avoiding the problem we met for the naïve implementation of the conditional Monte Carlo idea by discarding the largest X_i and considering only the remaining ones. For the simulation, we thus generate $X_1, \ldots, X_n$, form the order statistics

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}$$

(assuming for simplicity that F has a density f), throw away the largest one $X_{(n)}$, and let

$$Z_4(x) = \mathrm{P}(S_n > x | X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}) = \frac{\bar{F}((x - S_{n-1}^*) \vee X)_{(n-1)})}{\bar{F}(X_{(n-1)})},$$

where $S_{n-1}^* = X_{(1)} + X_{(2)} + \cdots + X_{(n-1)} = S_n - X_{(n)}$. To check the formula for the conditional probability, note first that

$$\mathrm{P}(X_{(n)} > x | X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}) = \frac{\bar{F}(X_{(n-1)} \vee x)}{\bar{F}(X_{(n-1)})}.$$

We then obtain

$$\mathrm{P}(S_n > x | X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}) = \mathrm{P}(X_{(n)} + S_{n-1}^* > x | X_{(1)}, X_{(2)}, \ldots, X_{(n-1)})$$

$$= \mathrm{P}(X_{(n)} > x - S_{n-1}^* | X_{(1)}, X_{(2)}, \ldots, X_{(n-1)})$$

$$= \frac{\bar{F}((x - S_{n-1}^*) \vee X_{(n-1)})}{\bar{F}(X_{(n-1)})}.$$

Theorem 3.1. *Assume that* $\bar{F}(x) = L(x)/x^\alpha$ $(\alpha > 1)$ *with* $L(x)$ *slowly varying. Then the algorithm given by* $\{Z_4(x)\}$ *is logarithmically efficient.*

The proof is given in the Appendix.

Algorithms using importance sampling. Theorem 3.1 leaves open what to do with distributions like the Weibull or lognormal. We shall here present a new algorithm involving importance sampling. Let H be the importance sampling distribution (not to be confused with the extreme value distribution of Section 2!), h its density and f the density of F. It turns out that what works is to take H as much more heavy-tailed than F in the following sense:

Condition 3.1. *The distribution K with density $k(x) = f^2(x)/(ch(x))$, where $c = \int_0^\infty f^2/h$, is subexponential and satisfies $\overline{K}(x) \lesssim \overline{F}(x)^2$.*

It is easily verified that taking $\bar{F}(x) \sim c/\log x$, say,

$$h(x) = \begin{cases} \dfrac{\eta}{x(\log x)^2}, & x > a, \\ \gamma g(x), & 0 \leqslant x \leqslant a, \end{cases} \tag{3.1}$$

where $a \geqslant e$, $\eta > 0$, g is an arbitrary strictly positive density on $[0, a]$ and

$$\int_0^\infty h(x)\, dx = \eta \int_a^\infty x^{-1}(\log x)^{-2}\, dx + \gamma = \frac{\eta}{\log a} + \gamma = 1$$

works in all of our three subexponential examples (lognormal, Weibull, regular variation). Note, in particular, that this distribution has infinite mean. That is, $h(x)$ decreases only slightly faster than $1/x$ (the choice $h(x) \sim x^{-1}$ itself does not work because we need h to be integrable). For the Weibull case, one can take h somewhat more moderately heavy-tailed, say Pareto where $h(x) = (\alpha - 1)(1 + x)^{-\alpha}$ with $\alpha \in (1, \infty)$.

It is not a priori obvious just how to implement the importance sampling, and here are two ways. One could just simulate X_1 from H and $X_2, \ldots, X_n$ from F, in which case the estimator is

$$Z_5(x) = \frac{f(X_1)}{h(X_1)} I(X_1 + \cdots + X_n > x);$$

or one could simulate all of $X_1, \ldots, X_n$ from H, leading to the estimator

$$Z_6(x) = \frac{f(X_1)}{h(X_1)} \cdots \frac{f(X_n)}{h(X_n)} I(X_1 + \cdots + X_n > x).$$

However, the estimator $Z_5(x)$ can easily be discarded because the event that one of $X_2, \ldots, X_n$ 'by itself' is large contributes too much to the variance:

Proposition 3.1. *For any F,*

$$\varliminf_{x \to \infty} \frac{\mathrm{E}\, Z_5(x)^2}{z(x)} > 0.$$

Proof.

$$\mathrm{E}Z_5(x)^2 \geqslant \mathrm{E}_H\left[\frac{f(X_1)^2}{h(X_1)^2}; X_2 > x\right] = c\mathrm{P}(X_2 > x) = cz(x).$$

$\square$

However:

Proposition 3.2. *If $F \in \mathscr{S}$ and Condition 3.1 holds, then the estimator $Z_6(x)$ is logarithmically efficient.*

Proof.

$$\mathrm{E}Z_6(x)^2 = \int \cdots \int_{\{x_1+\cdots+x_n>x\}} \frac{f(x_1)^2}{h(x_1)^2} \cdots \frac{f(x_n)^2}{h(x_n)^2} h(x_1) \ldots h(x_n)\,\mathrm{d}x_1 \ldots \mathrm{d}x_n$$

$$= c^n \int \cdots \int_{\{x_1+\cdots+x_n>x\}} k(x_1) \ldots k(x_n)\,\mathrm{d}x_1 \ldots \mathrm{d}x_n$$

$$= c^n\mathrm{P}_K(X_1 + \cdots + X_n > x) \sim nc^n\overline{K}(x) \overset{\log}{\leqslant} nc^n\overline{F}(x)^2.$$

$\square$

Random sums and the $M/G/1$ *queue.* Now consider the case where $n = N$ is a random variable independent of $X_1, X_2 \ldots$. Here are two main examples. The first is the Pollaczek–Khinchine formula for the distribution of the $M/G/1$ waiting time W, or the compound Poisson ruin probability $\psi(x)$ as in Asmussen and Binswanger (1997),

$$z(x) = \mathrm{P}(W > x) = \psi(x) = \mathrm{P}(X_1 + \cdots + X_N > x), \qquad (3.2)$$

where N, X_1, X_2, ... are independent, $\mathrm{P}(N = n) = (1 - \rho)\rho^n$ and $\mathrm{P}(X_k \leqslant x) = F(x)$, in which F is the stationary excess distribution corresponding to the service-time distribution B, $F(x) = \mu_B^{-1}\int_0^x \overline{B}(y)\,\mathrm{d}y$ ($\rho = \beta\mu_B$, where β is the arrival intensity). The second is the accumulated claims for an insurance company in a year, where the most standard assumption is that N is Poisson, say with rate λ.

It is straightforward to see from the variance calculations in the proofs of Theorem 3.1 and Proposition 3.2 that both algorithms work in this setting too, provided $\mathrm{E}N^2 < \infty$ for the order statistics algorithm and $\mathrm{E}Nc^N < \infty$ for the importance sampling algorithm. In the setting of the $M/G/1$ queue, $\mathrm{E}N^2 < \infty$ is automatic but the order statistics algorithm only works for the regularly varying case (and, in fact, for the lognormal case; see Binswanger 1997). For subexponential distributions such as the Weibull, the importance sampling algorithm is applicable provided we can choose the parameters in (3.1) in such a way that $\rho c < 1$. In fact, define

$$\gamma(a) = 1 - \frac{1}{\log a}, \qquad \varepsilon(a) = \int_a^\infty f^2(x)x\log^2(x)\,\mathrm{d}x.$$

Then:

Theorem 3.2. *Let a be chosen so large that*

$$\frac{F^2(a)}{\gamma(a)} + \varepsilon(a) < \frac{1}{\rho},$$

and define

$$h(x) = \begin{cases} \gamma(a)\dfrac{f(x)}{F(a)}, & x < a, \\[2mm] \dfrac{1}{x\log^2(x)}, & x > a. \end{cases}$$

Then the estimator

$$Z_6^*(x) = \frac{f(X_1)}{h(X_1)} \ \cdots \ \frac{f(X_N)}{h(X_N)} I(X_1 + \cdots + X_N > x)$$

where $N, X_1, X_2, \ldots$ are as in (3.2), is logarithmically efficient for estimating $z(x) = \mathrm{P}(W > x)$ in the $M/G/1$ queue.

Proof. It is easily seen that $\int h(x)\,\mathrm{d}x = 1$ and that $\gamma(a) \to 1$ and $\varepsilon(a) \to 0$. Hence the choice of a is feasible, and we get

$$\begin{aligned} c &= \int_0^\infty \frac{f^2(x)}{h(x)}\,\mathrm{d}x \\[2mm] &= \frac{1}{\gamma(a)} \int_0^a f(x)F(a)\,\mathrm{d}x + \int_a^\infty f^2(x)x\log^2(x)\,\mathrm{d}x \\[2mm] &= \frac{1}{\gamma(a)} F^2(a) + \varepsilon(a) < \frac{1}{\rho}. \end{aligned}$$

$\square$

4. The large-deviations approach

We now consider the idea of finding a probability $\tilde{\mathrm{P}}^{(x)}$ which in an appropriate asymptotic sense is close to the conditional distribution $\mathrm{P}^{(x)}$ given the rare event, and use $\tilde{\mathrm{P}}^{(x)}$ as importance sampling distribution.

We will start with two extremely simple examples, indicating the difficulties with this idea in the heavy-tailed case.

Example 4.1. Assume that $A(x) = I(X_1 + \cdots + X_n > x)$. The subexponential asymptotics in its crudest form says that $A(x)$ occurs by one of $X_1, \ldots, X_n$ becoming large (and even exceeding x) while the remaining ones are typical (unaffected). Thus, one would simulate one

X_i, say X_1, from $F^x(y) = (F(y) - F(x))/\bar{F}(x)$, $y \geqslant x$, and $X_2 \ldots, X_n$ from F. However, this importance sampling distribution $\tilde{P}$ fails to satisfy the crucial absolute continuity criterion (1.2) because $\tilde{P}(B(x)) = 0$ but $P(B(x)) > 0$, where

$$B(x) = \left\{ X_1 + \cdots + X_n > x, \ \max_{i=1,\ldots,n} X_i \leqslant x \right\}.$$

Example 4.2. Assume that $A(x) = I(X > x)$, where X is Weibull (β) with $\beta < 1$, that is, with density $f(y) = \beta y^{\beta-1} e^{-y^\beta}$, $y > 0$. Then by (2.3), the conditional distribution of $(X - x)/\gamma(x)$ given $X > x$ is limiting standard exponential where $\gamma(x) = x^{1-\beta}/\beta$. Thus we may try the estimator $Z_7(x) = f(x + Y)/h(Y)$, where Y is generated from the exponential density $h(y) = \gamma(x)^{-1} e^{-y/\gamma(x)}$. But

$$\mathrm{E} Z_7(x)^2 = \int_0^\infty \frac{f(x+y)^2}{h(y)^2} h(y) \, \mathrm{d}y = \int_0^\infty \beta^2 \gamma(x)(x+y)^{2\beta-2} e^{y/\gamma(x)-2(x+y)^\beta} \, \mathrm{d}y = \infty.$$

In contrast, if $\beta = 1$ we are in the light-tailed case with $z(x) = P(A(x)) = e^{-x}$ and obtain $\mathrm{E} Z_7(x)^2 = e^{-2x}$ so that (1.1) is satisfied because $z(x) = e^{-x}$ in this case.

We will encounter a similar phenomenon in the random walk setting $z(x) = P(\tau(x) < \infty)$. Here $\tilde{P}^{(x)}$ is described in Proposition 2.1 and Remark 2.1 (we assume without loss of generality that $\mu = -1$ in the following), which leads to the following algorithm: if $\alpha = \infty$, that is $F \in MDA(\Lambda)$ (say the lognormal or Weibull case), we generate T from a geometric distribution with mean $\gamma(x)$ (note that $T/\mathrm{E}T$ is then limiting standard exponential) and in the regularly varying case $\bar{F}(x) = L(x)/x^\alpha$, $1 < \alpha < \infty$, we generate T from a 'discrete' Pareto distribution with mean $\gamma(x)$ (see below). We force the random walk to cross level x at time T, that is, simulate X_T from $F^{x-S_{T-1}}$ where $F^x(y) = (F(y) - F(x))/\bar{F}(x)$, $y \geqslant x$. This approach has (at least) two variants: one could just simulate $X_1, X_2, \ldots, X_{T-1}$ from F, in which case the estimator is

$$Z_8(x) = \frac{\prod_{i=1}^{\tau} f(X_i)}{\sum_{k=1}^{\tau} p_k f^{x-S_{k-1}}(X_k) \prod_{i=1}^{k-1} f(X_i) + \bar{p}_\tau \prod_{i=1}^{\tau} f(X_i)},$$

where $p_n = P(T = n)$, $\bar{p}_n = P(T > n)$ and $\tau = \tau(x)$ (note that $\tau \leqslant T$). Since $f^{x-S_{i-1}}(X_i) = 0$ for $i < \tau$, we obtain

$$Z_8(x) = \frac{\bar{F}(x - S_{\tau-1})}{p_\tau + \bar{p}_\tau \bar{F}(x - S_{\tau-1})}.$$

In this case a level crossing before time T is possible. Another variant is to exclude such a level crossing by simulating $X_k (k = 1, 2, \ldots, T-1)$ from $F_{x-S_{k-1}}$ where $F_x(y) = F(y)/F(x)$, $y \leqslant x$, in which case the estimator is

$$Z_9(x) = \frac{\prod_{i=1}^{\tau} f(X_i)}{\sum_{k=1}^{\tau} p_k f^{x-S_{k-1}}(X_k) \prod_{i=1}^{k-1} f_{x-S_{i-1}}(X_i) + \bar{p}_\tau \prod_{i=1}^{\tau} f_{x-S_{i-1}}(X_i)}$$

$$= \frac{1}{p_\tau} \bar{F}(x - S_{\tau-1}) \prod_{i=1}^{\tau-1} F(x - S_{i-1}).$$

If $\alpha = \infty$ then $p_n = \rho(1 - \rho)^n$, where $\rho = \rho(x) = 1 - 1/\gamma(x)$, and then $\bar{p}_n = (1 - \rho)^n$. In this case we then have

$$Z_8(x) = \frac{1}{(1 - \rho)^\tau} \frac{\bar{F}(x - S_{\tau-1})}{\rho + \bar{F}(x - S_{\tau-1})}$$

$$Z_9(x) = \frac{1}{\rho(1 - \rho)^\tau} \bar{F}(x - S_{\tau-1}) \prod_{i=1}^{\tau-1} F(x - S_{i-1}).$$

Theorem 4.1. *Assume that F is in $MDA(\Lambda) \cap \mathscr{S}$ and that $\int_x^\infty \bar{F}(y)\,dy$ is a subexponential tail. Then for all $x > 0$ we have* (a) $E[Z_8(x)^2] = \infty$ *and* (b) $E[Z_9(x)^2] = \infty$.

The proof is given in the Appendix.

In the regularly varying case $\bar{F}(x) = L(x)/(1 + x)^\alpha$, $\gamma(x) = (1 + x)/(\alpha - 1)$, and we choose T with point probabilities $p_k = c(x)(1 + x)^{\alpha-1}/(1 + x + k)^\alpha$, where c is a normalizing constant. It is easy to see that then $c(x) \to \alpha - 1$ and that $P(T > (1 + x)y) \to (1 + y)^{1-\alpha}$ as required.

Theorem 4.2. *Assume that $\bar{F}(y) = (1 + y)^{-\alpha}$ for some $\alpha > 1$. Then for all $x > 0$ we have* $E[Z_9(x)^2] = \infty$.

The proof is given in the Appendix.

Theorem 4.3. *Assume that $\bar{F}(y) = (1 + y)^{-\alpha}$ for some $\alpha > 2$. Then the estimator $Z_8(x)$ fails to satisfy the efficiency criterion* (1.1).

The proof is given in the Appendix.

5. Concluding remarks

A main conclusion to be drawn from the present paper is that the rare events simulation method which is prevalent in the light-tailed case, importance sampling via large-deviations arguments, does not seem to hold promise of being equally useful in the heavy-tailed case (one may note that there are recent indications that it also has limitations in the light-tailed case; see Glasserman and Wang 1997).

We have presented two algorithms which have as good asymptotic efficiency properties as

the standard estimators in the light-tailed case. However, the situations in which these estimators can be used are quite special. We consider it premature to predict whether a general methodology using features of these algorithms can be developed. At this stage, one must note that there is no efficient algorithm for simulating the tail waiting time probability $P(W_\infty > x)$ in a general $GI/G/1$ queue with heavy-tailed service-time distribution. For the buffer overflow probability $P(M_C > x)$ (see Proposition 2.2), we did not manage to come up with an efficient algorithm even for M/G/1.

A technical question which is left open is the behaviour of the estimator $Z_8(x)$ when $1 < \alpha \leq 2$ (finite mean but infinite variance). We conjecture that also in this case $Z_8(x)$ will fail to satisfy (1.1). One reason is that $Z_8(x)$ captures the large-deviations reasoning less well than $Z_9(x)$ (which has infinite variance in all cases). For example, the distribution of τ subject to the importance sampling measure is not asymptotically correct.

A possible line for further research is to relax the rather strict requirements we have imposed upon our estimators. Here are two possible relaxations of the assumptions, for which is may be easier to find algorithms:

(a) One could allow some bias so that $E Z(x)$ is close to but not necessarily equal to $z(x)$.
(b) One could look for estimators which improve the crude Monte Carlo estimators without attaining the efficiency criterion (1.1).

Appendix: Proofs

The key step in the proof of Theorem 3.1 is the following estimate:

Lemma A.1.

$$E[Z_4(x)^2] \leq n(n-1)\left[\frac{1}{2}\bar{F}^2\left(\frac{x}{2}\right) - \bar{F}^2\left(\frac{x}{n}\right)\log\bar{F}\left(\frac{x}{2}\right)\right]. \tag{A.1}$$

Proof. We first recall (see, for example, Gut 1995, p. 106) that the density $f_{X_{(n-1)}}(x)$ of the random variable $X_{(n-1)}$ is

$$f_{X_{(n-1)}}(x) = n(n-1)F^{n-2}(x)\bar{F}(x)f(x).$$

We then obtain

$$E[Z_4(x)^2] = E\left[\frac{\bar{F}(x - S_{n-1}^*)}{\bar{F}(X_{(n-1)})}; X_{(n-1)} \leq \frac{x}{n}\right]^2 + E\left[\frac{\bar{F}((x - S_{n-1}^*) \vee X_{(n-1)})}{\bar{F}(X_{(n-1)})}; \frac{x}{n} < X_{(n-1)} \leq \frac{x}{2}\right]^2$$

$$+ E\left[1; X_{(n-1)} > \frac{x}{2}\right]^2. \tag{A.2}$$

The first term on the right-hand side of (A.2) can be bounded as follows. If $X_{(n-1)} \leq x/n$ then $\bar{F}(x - S_{n-1}^*) \leq \bar{F}(x/n)$, so that

$$\mathrm{E}\left[\frac{\bar{F}(x - S_{n-1}^*)}{\bar{F}(X_{(n-1)})}; X_{(n-1)} \leqslant \frac{x}{n}\right]^2 \leqslant \bar{F}^2\left(\frac{x}{n}\right)\int_0^{x/n} \frac{f_{X_{(n-1)}}(y)}{\bar{F}^2(y)}\,\mathrm{d}y$$

$$\leqslant n(n-1)\bar{F}^2\left(\frac{x}{n}\right)\int_0^{x/n} \frac{f(y)}{\bar{F}(y)}\,\mathrm{d}y$$

$$= -n(n-1)\bar{F}^2\left(\frac{x}{n}\right)\log\bar{F}\left(\frac{x}{n}\right).$$

The second term in (A.2) can be bounded in the same way. For $x/n < X_{(n-1)} \leqslant x/2$, $\bar{F}((x - S_{n-1}^*) \vee X_{(n-1)} \leqslant \bar{F}(x/n)$, yielding

$$\mathrm{E}\left[\frac{\bar{F}((x - S_{n-1}^*) \vee X_{(n-1)})}{\bar{F}(X_{(n-1)})}; \frac{x}{n} < X_{(n-1)} \leqslant \frac{x}{2}\right]^2 \leqslant \bar{F}^2\left(\frac{x}{n}\right)\int_{x/n}^{x/2} \frac{f_{X_{(n-1)}}(y)}{\bar{F}^2(y)}\,\mathrm{d}y$$

$$\leqslant n(n-1)\bar{F}^2\left(\frac{x}{n}\right)\int_{x/n}^{x/2} \frac{f(y)}{\bar{F}(y)}\,\mathrm{d}y$$

$$= -n(n-1)\bar{F}^2\left(\frac{x}{n}\right)x\left[\log\bar{F}\left(\frac{x}{2}\right) - \log\bar{F}\left(\frac{x}{n}\right)\right].$$

To find an upper bound for the third term in (A.2) we write

$$\mathrm{E}\left[1; X_{(n-1)} > \frac{x}{2}\right]^2 = \int_{x/2}^{\infty} f_{X_{(n-1)}}(y)\,\mathrm{d}y$$

$$= n(n-1)\int_{x/2}^{\infty} F^{n-2}(y)\bar{F}(y)f(y)\,\mathrm{d}y$$

$$\leqslant n(n-1)\int_{x/2}^{\infty} \bar{F}(x)f(x)\,\mathrm{d}x$$

$$= n(n-1)\frac{1}{2}\bar{F}^2\left(\frac{x}{2}\right).$$

Adding the above inequalities leads to the desired result. □

Proof of Theorem 3.1. Since $F \in \mathscr{S}$, we have $z(x) = \mathrm{P}(S_n > x) \sim n\bar{F}(x) = nL(x)/x^{\alpha}$. It follows from $X^{\epsilon}L(x) \to \infty$, $x^{-\epsilon}L(x) \to 0$, $x \to \infty$, for any $\epsilon > 0$, that $|\log z(x)| \sim \alpha \log x$. Since $L(x/d) \sim L(x)$, we have $\bar{F}(x/d) \sim d^{\alpha}\bar{F}$, and hence Lemma A.1 yields

$$|\log \mathrm{var}\, Z_4(x)| = -\log \mathrm{var}\, Z_4(x) \geqslant -\log \mathrm{E}[Z_4(x)^2]$$

$$\sim -\log \bar{F}(x)^2 \sim 2\alpha \log x. \qquad \square$$

Proof of Theorem 4.1. Throughout the proof we let C denote any positive constant and denote the importance sampling measure by $\tilde{\mathrm{P}}^{(x)}$.

(a) Let $\tilde{S}_n$ be a random walk under $\tilde{\mathrm{P}}^{(x)}$ with increment distribution F and $\tilde{\tau} = \inf\{n : \tilde{S}_n \geq x\}$. Then, by the strong law of large numbers, we can find constants $\varepsilon, \eta, \delta > 0$, with $\eta < 1 < \varepsilon$, such that

$$\tilde{\mathrm{P}}^{(x)}(-\varepsilon n - \delta \leq \tilde{S}_n \leq -\eta n + \delta \text{ for all } n) \geq C. \tag{A.3}$$

Since under $\tilde{\mathrm{P}}^{(x)}$, $\{S_n\}_{n<\tau} \overset{\mathscr{D}}{=} \{\tilde{S}_n\}_{n<\tilde{\tau}}$ (see, for example, Asmussen and Klüppelberg 1996) we have

$$\mathrm{E}[Z_8(x)^2] \geq C\mathrm{E}\left[\frac{1}{(1-\rho)^{2\tau}} \left(\frac{\bar{F}(x+\varepsilon\tau+\delta)}{\bar{F}(x+\varepsilon\tau+\delta)+\rho}\right)^2\right].$$

Clearly,

$$\tilde{\mathrm{P}}^{(x)}(T = n, \tilde{\tau} > T) \geq (1 - z(x))\tilde{\mathrm{P}}^{(x)}(T = n).$$

Thus

$$\tilde{\mathrm{E}}^{(x)}[Z_8(x)^2] \geq C\tilde{\mathrm{E}}^{(x)}\left[\frac{1}{(1-\rho)^{2T}} \left(\frac{\bar{F}(x+\varepsilon T+\delta)}{\bar{F}(x+\varepsilon T+\delta)+\rho}\right)^2; \tilde{\tau} > T\right]$$

$$+ C\tilde{\mathrm{E}}^{(x)}\left[\frac{1}{(1-\rho)^{2\tilde{\tau}}} \left(\frac{\bar{F}(x+\varepsilon\tau+\delta)}{\bar{F}(x+\varepsilon\tilde{\tau}+\delta)+\rho}\right)^2; T > \tilde{\tau}\right]$$

$$\geq C\tilde{\mathrm{E}}^{(x)}\left[\frac{1}{(1-\rho)^{2T}} \left(\frac{\bar{F}(x+\varepsilon T+\delta)}{\bar{F}(x+\varepsilon T+\delta)+\rho}\right)^2\right].$$

Now the right-hand side equals

$$\frac{C\rho}{1-\rho}\sum_{n=1}^{\infty}\frac{1}{(1-\rho)^{n-1}}\left(\frac{\bar{F}(x+\varepsilon n+\delta)}{\bar{F}(x+\varepsilon n+\delta)+\rho}\right)^2 = \infty$$

since $F \in \mathscr{S}$ implies

$$\lim_{n\to\infty} \theta^n \bar{F}(x+\varepsilon n) > 0$$

for all x and $\theta > 1$.

(b) In this case let $\tilde{S}_n$ have increment distribution $F_{x-\tilde{S}_{n-1}}$. If $S_n \leq x$ for all x, then $\tilde{S}_n = S_n$ for all n. Since $\mathrm{P}(S_1 \leq x, S_2 \leq x, \ldots) \to 1$ as $x \to \infty$, we conclude that (A.3) holds also in this case. This leads to

$$\tilde{\mathrm{E}}^{(x)}[Z_9(x)^2] \geqslant C\tilde{\mathrm{E}}^{(x)}\left[\frac{1}{\rho^2(1-\rho)^{2(\tau-1)}}\left(\prod_{n=0}^{\tau-1}F(x+\eta n-\delta)\right)^2\bar{F}(x+\varepsilon\tau+\delta)^2\right]$$

$$\geqslant C\tilde{\mathrm{E}}^{(x)}\left[\frac{1}{\rho^2(1-\rho)^{2(\tau-1)}}\left(\prod_{n=0}^{\infty}F(x+\eta n-\delta)\right)^2\bar{F}(x+\varepsilon\tau+\delta)^2\right].$$

Now $\Pi_{n=0}^{\infty}F(x+\eta n-\delta)=c>0$ if and only if $\sum_{n=0}^{\infty}\bar{F}(x+\eta n-\delta)<\infty$, which follows from $\mu_F<\infty$. Since in this case $\tau=T$ almost surely we obtain, as above,

$$\tilde{\mathrm{E}}^{(x)}[Z_9(x)^2] \geqslant C\sum_{n=1}^{\infty}\frac{1}{\rho(1-\rho)^{n-1}}\bar{F}(x+\varepsilon n+\delta)^2=\infty,$$

Since $F\in\mathscr{S}$.   □

Proof of Theorem 4.2. Again we let C denote any positive constant and denote the importance sampling measure by $\tilde{\mathrm{P}}^{(x)}$.

Obviously

$$\tilde{\mathrm{E}}^{(x)}[Z_9(x)^2] \geqslant \tilde{\mathrm{E}}^{(x)}\left[\frac{1}{p_\tau^2}\left\{\prod_{k=1}^{\tau-1}F(x-S_{k-1})\bar{F}(1)\right\}^2 I(X_{\tau-1}\in(x-1-S_{\tau-2};x-S_{\tau-2}))\right].$$

As in the previous proof, we can apply (A.3) to obtain, for some $\varepsilon,\delta>0$,

$$\tilde{\mathrm{E}}^{(x)}[Z_9(x)^2] \geqslant C\tilde{\mathrm{E}}^{(x)}\left[\frac{1}{p_\tau^2}\left\{\prod_{k=1}^{\tau-1}F(x+\eta k-\delta)\right\}^2 I(X_{\tau-1}\in(x-1+\varepsilon\tau+\delta;x+\eta\tau+\delta))\right].$$

$$\geqslant C\tilde{\mathrm{E}}^{(x)}\left[\frac{1}{p_\tau^2}I(X_{\tau-1}\in(x-1+\varepsilon\tau+\delta;x+\eta\tau+\delta))\right]$$

$$\geqslant C\sum_{t=0}^{\infty}\frac{1}{p_t}f(x+\eta t+\delta)[(\eta-\varepsilon)t+1]$$

$$\geqslant C\sum_{t=0}^{\infty}\frac{1}{p_t}f(x+\eta t+\delta)=\frac{C}{(1+x)^{\alpha-1}}\sum_{t=0}^{\infty}(1+x+t)^{\alpha}\frac{1}{(1+x+\eta t+\delta)^{\alpha+1}}.$$

It is easily seen that

$$\frac{1+x+t}{1+x+\eta t+\delta} \geqslant \min[1/\eta,(1+x)/(1+x+\delta)]$$

for $t\geqslant 0$, and we obtain

$$\tilde{\mathrm{E}}^{(x)}[Z_9(x)^2] \geqslant \frac{C}{(1+x)^{\alpha-1}}\sum_{t=0}^{\infty}\frac{1}{(1+x+\eta t+\delta)}=\infty.$$

□

In the proof of Theorem 4.3, let $\tilde{S}_t$ be a random walk with increments $\tilde{X}_1, \tilde{X}_2, \ldots$ distributed according to F and $\tilde{\tau} = \inf\{t > 0 : \tilde{S}_t > x\}$. Let $\beta = 1/\alpha$ and let $T_{x,t}$ be the interval

$$(x - (x + \eta t)^\beta, \, x - 2(x + \eta t)^\beta);$$

cf. (A.3).

Lemma A.2. *Let $g_{x,t}(z)$ be the density of the absolutely continuous part of the random variable $\tilde{S}_t I(\tilde{\tau} > t)$. Then, for $t \geqslant x$,*

$$g_{x,t}(z) C \frac{(x + t)^\beta}{(x + t)^{\alpha+1}}, \qquad z \in I_{x,t}.$$

Proof. Let $t' = 1, \ldots, t$ be fixed, let $\tilde{S}_t^{(t')} = \tilde{S}_t - \tilde{X}_{t'}$ and let $A(t')$ be the event that

$$\tilde{S}_{t''} \leqslant 0 \quad \text{for all } t'' < t', \qquad \tilde{S}_{t''} - \tilde{S}_{t'} \leqslant 0 \quad \text{for all } t' \leqslant t'' < t \tag{A.4}$$

$$\tilde{S}_t^{(t')} \in (-\epsilon t - \delta, -\eta t + \delta), \quad \tilde{S}_t^{(t')} \tilde{S}_{t'}^{(t')} \in (-\epsilon(t - t') - \delta, -\eta(t - t') + \delta) \quad \text{for all } t' \leqslant t'' < t. \tag{A.5}$$

Then, by a variant of (A.3), $P(A(t')) \geqslant C > 0$ with C independent of t, t'. Write $z = z_1 + z_2 + z_3$, where $z_1 = \tilde{S}_{t'}^{(t')}$, $z_2 = \tilde{X}_{t'}$, $z_3 = \tilde{S}_t^{(t')} - \tilde{S}_{t'}^{(t')}$, and assume $t - t' \leqslant (x + \eta t)^\beta/\epsilon - \delta$. On $A(t')$, we then have

$$\tilde{S}_{t'} = z - z_3 < x - (x + \eta t)^\beta + \epsilon(t - t') + \delta \leqslant x.$$

Then (A.4) yields $\tilde{S}_{t''} < x$ for all $t'' \leqslant t$ and hence $\tilde{\tau} > t$. It follows that the contribution to $g_{x,t}(z)$ from $A(t')$ is at least

$$f(z_2) \geqslant f(C(x + \eta t)) \geqslant C(x + \eta t)^{-\alpha-1} \tag{A.6}$$

because, by (A.5),

$$z_2 = z - z_1 - z_3 \geqslant z + \eta t - \delta \geqslant x - 2(x + \eta t)^\beta + \eta t - \delta \geqslant C(x + \eta t).$$

Now, finally, note that the $A(t')$ are mutually exclusive (say, the first upcrossing of level zero occurs at time t' by (A.4)) and sum (A.6) over the $(x + \eta t)^\beta/\epsilon - \delta \geqslant C(x + t)^\beta$ allowed values of t' to obtain the result. $\qquad\square$

Proof of Theorem 4.3. If $S_{t-1} \in I_{x,t}$, we have

$$\bar{F}(x - S_{t-1}) \geqslant \bar{F}(2(x + \eta t)^\beta) \geqslant C(x + \eta t)^{-\alpha\beta} \geqslant C(x + t)^{-1},$$

$$p_t + \bar{p}_t \bar{F}(x - S_{t-1}) \leqslant p_t\{1 + C(x + t)\bar{F}(x - S_{t-1})\}$$

$$\leqslant p_t\{1 + C(x + t)\bar{F}((x + \eta t)^\beta)\}$$

$$\leqslant p_t\{1 + C(x + t)\bar{F}((x + \eta t)^\beta)\} \leqslant C p_t.$$

Hence

$$\mathrm{E}Z_8(x)^2 \geqslant \sum_{t=1}^{\infty} \mathrm{E}\left[\left(\frac{\bar{F}(x - S_{t-1})}{p_t + \bar{p}_t \bar{F}(x - S_{t-1})}\right)^2 ; T = t \leqslant \tilde{\tau}, S_{t-1} \in I_{x,t}\right]$$

$$\geqslant C \sum_{t=1}^{\infty} p_t \frac{1}{(p_t(1 + x + t))^2} \int_{I_{x,t}} g_{x,t}(z)\,\mathrm{d}z$$

$$\geqslant C \sum_{t=1}^{\infty} \frac{1}{p_t} \frac{1}{(1 + x + t)^2} \frac{(1 + x + t)^{2\beta}}{(1 + x + t)^{\alpha+1}}$$

$$= C \sum_{t=1}^{\infty} \frac{(1 + x + t)^\alpha}{x^{\alpha-1}} \frac{1}{(1 + x + t)^2} \frac{(1 + x + t)^{2\beta}}{(1 + x + t)^{\alpha+1}}$$

$$= C \frac{1}{x^{\alpha-1}} \sum_{t=1}^{\infty} \frac{1}{(1 + x + t)^{3-2\beta}} \geqslant C \frac{1}{x^{\alpha-1}} \frac{1}{x^{2-2\beta}}.$$

Now $z(x)$ is of order $x^{-(\alpha-1)}$, cf. (2.4), so that it suffices to note that $2 - 2\beta = 2 - 2/\alpha < \alpha - 1$ when $\alpha > 2$. □

References

Abate, J., Choudhury, G.L. and Whitt, W. (1994) Waiting-time tail probabilities in queues with long-tail service-time distributions. *Queuing Syst.*, **16**, 311–338.

Anantharam, V. (1988) How large delays build up in a $GI/GI/1$ queue. *Queueing Syst.*, **5**, 345–368.

Asmussen, S. (1982) Conditioned limit theorems relating a random walk to its associate, with applications to risk reserve processes and the $GI/G/1$ queue. *Adv. Appl. Probab.*, **14**, 143–170.

Asmussen, S. (1987) *Applied Probability and Queues*. Chichester: Wiley.

Asmussen, S. (1992) Phase-type representations in random walk and queueing problems. *Ann. Probab.*, **20**, 772–789.

Asmussen, S. (1995) Stationary distributions for fluid flow models with or without Brownian noise. *Comm. Statist. Stochastic Models*, **11**, 21–49.

Asmussen, S. (1996) Rare events in the presence of heavy tails. In P. Glasserman, K. Sigman and D. Yao (eds), *Stochastic Networks: Rare Events and Stability*, Lecture Notes in Statist. 117, pp. 197–214. New York: Springer-Verlag.

Asmussen, S. (1998) Subexponential asymptotics for stochastic processes: extremal behaviour, stationary distributions and first passage times. *Ann. Appl. Probab.*, **8**, 354–374.

Asmussen, S. and Binswanger, K. (1997) Ruin probability simulation for subexponential claims. *ASTIN Bull.*, **27**, 297–318.

Asmussen, S. and Højgaard, B. (1996) Finite horizon ruin probabilities for Markov-modulated risk procesess with heavy tails. *Teor. Sluch. Protsess.*, **2**, 96–107.

Asmussen, S. and Klüppelberg, C. (1996) Large deviations results for subexponential tails, with applications to insurance risk. *Stochastic Process. Appl.*, **64**, 103–125.

Asmussen, S. and Rolski, T. (1991) Computational methods for ruin probabilities: a matrix-algorithmic approach. *Insurance Math. Econom.*, **10**, 259–274.

Asmusssen, S. and Rubinstein, R.Y. (1995) Steady-state rare events simulation and its complexity

properties. In J. Dshalalow (ed.), *Advances in Queueing: Models, Methods & Problems*, pp. 429–466. Boca Raton, FL: CRC Press.

Asmussen, S., Schmidli, H. and Schmidt, V. (1999) Tail asymptotics for non-standard risk and queueing processes with subexponential jumps. *Adv. Appl. Probab.* To appear.

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation*, Encyclopedia of Mathematics and its Applications, **27**. Cambridge: Cambridge University Press.

Binswanger, K. (1997) *Rare Events and Insurance*. Doctoral dissertation 12233, ETH, Zurich.

Boxma, O.J. (1996) Fluid queues and regular variation. *Perform. Eval.*, **27/28**, 699–712.

Boxma, O.J. (1997) Regular variation in a multi-source fluid queue. In V. Ramaswami and P.E. Wirth (eds), *Teletraffic Contributions for the Information Age – Proceedings ITC 15*, pp. 391–402. Amsterdam: North-Holland.

Bucklew, J.A., Ney, P. and Sadowsky, J.S. (1990) Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *J. Appl. Probab.*, **27**, 44–59.

Chistyakov, V.P. (1964) A theorem on sums of independent random variables and its application to branching random processes. *Theory Probab. Appl.*, **9**, 640–648.

Choudhury, G.L. and Whitt, W. (1997) Long-tail buffer-content distributions in broadband networks. *Perform. Eval.*, **30**, 177–190.

Crovella, M. and Bestravros, A. (1996) Self-similarity in World Wide Web traffic – evidence and possible causes. *Proc. Sigmetrics*, **96**, 160–169.

Cunha, Bestravros, A. and Crovella, M. (1995) Characteristics of WWW client-based traces. Manuscript.

Durrett, R. (1980) Conditioned limit theorems for random walks with negative drift. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **52**, 277–287.

Embrechts, P. and Veraverbeke, N. (1982) Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance Math. Econom.*, **1**, 55–72.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Finance and Insurance.* Heidelberg: Springer-Verlag.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications, Vol II* (2nd edn). New York: Wiley.

Goldie, C.M. and Klüppelberg, C. (1998) Subexponential distributions. In R. Adler, R. Feldman and M.S. Taqqu, (eds), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications.* Boston: Birkhäuser Verlag.

Goldie, C.M. and Resnick, S.I. (1988) Distributions that are both subexponential and in the domain of attraction of an extreme-value distribution. *J. Appl. Probab.*, **20**, 706–718.

Glasserman, P. and Wang, Y. (1997) Counterexamples in importance sampling for large deviations probabilities. *Ann. Appl. Probab.*, **7**, 731–746.

Gut, A. (1995) *An Intermediate Course in Probability.* New York: Springer-Verlag.

Heath, D., Resnick, S. and Samorodnitsky, G. (1997) Patterns of buffer overflow in a class of queues with long memory in the input stream. *Ann. Appl. Probab.*, **7**, 1021–1057.

Heath, D., Resnick, S. and Samorodnitsky, G. (1998) Heavy tails and long range dependence in on/off processes and associated fluid models. *Math. Oper. Res.*, **23**, 145–165.

Heidelberger, P. (1995) Fast simulation of rare events in queueing and reliability models, *ACM Trans. Model. Comput. Simulation*, **6**, 43–85.

Iglehart, D.L. (1972) Extreme values in the GI/G/1 queue. *Ann. Math. Statist.*, **43**, 627–635.

Jelenković, P.R. and Lazar, A.A. (1995) Subexponential asymptotics of a network multiplexer. Research Report, Centre for Telecommunications Research, Columbia University, New York.

Jelenković, P.R. and Lazar, A.A. (1996) Multiplexing on-off sources with subexponential on periods.

Research Report, Centre for Telecommunications Research, Columbia University, New York.

Jensen, J.L. (1995) *Saddlepoint Approximations*. Oxford: Oxford University Press.

Kalashnikov, V. (1999) Bounds for ruin probabilities in the presence of large claims and their comparison. *North Amer. Actuar. J.*, **3**(2).

Klüppelberg, C. and Stadtmüller, U. (1998) Ruin probabilities in the presence of heavy-tails and interest rates. *Scand. Actuar. J.*, 49–58.

Rolski, T., Schlegel, S. and Schmidt, V. (1999) Asymptotics of Palm-stationary buffer content distributions in fluid flow queues. *Adv. Appl. Probab.*, **31**(1), 234–253.

Sadowsky, J.S. (1993) On the optimality and stability of exponential twisting in Monte Carlo simulation. *IEEE Trans. Inform. Theory*, **IT-39**, 119–128.

Siegmund, D. (1976) Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.*, **4**, 673–684.

Willinger, W., Paxson, V. and Taqqu, M. (1998) Self-similarity and heavy tails: Structural modeling of network traffic. In R. Adler, R. Feldman and M.S. Taqqu (eds). *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Boston: Birkhäuser Verlag.

Willinger, W., Taqqu, M., Sherman, R. and Wilson, D. (1997) Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Trans. Networking*, **5**, 71–86.