IMPROVED ALGORITHMS FOR RARE EVENT SIMULATION WITH HEAVY TAILS

SØREN ASMUSSEN,* *University of Aarhus*

DIRK P. KROESE,** *The University of Queensland*

Abstract

The estimation of $P(S_n > u)$ by simulation, where S_n is the sum of independent, identically distributed random varibles $Y_1, \ldots, Y_n$, is of importance in many applications. We propose two simulation estimators based upon the identity $P(S_n > u) = nP(S_n > u, M_n = Y_n)$, where $M_n = \max(Y_1, \ldots, Y_n)$. One estimator uses importance sampling (for Y_n only), and the other uses conditional Monte Carlo conditioning upon $Y_1, \ldots, Y_{n-1}$. Properties of the relative error of the estimators are derived and a numerical study given in terms of the M/G/1 queue in which n is replaced by an independent geometric random variable N. The conclusion is that the new estimators compare extremely favorably with previous ones. In particular, the conditional Monte Carlo estimator is the first heavy-tailed example of an estimator with bounded relative error. Further improvements are obtained in the random-N case, by incorporating control variates and stratification techniques into the new estimation procedures.

Keywords: Bounded relative error; complexity; conditional Monte Carlo conditioning; control variate; logarithmic efficiency; M/G/1 queue; Pollaczek–Khinchin formula; rare event; regular variation; stratification; subexponential distribution; Weibull distribution

2000 Mathematics Subject Classification: Primary 65C60
Secondary 60K25

1. Introduction

This paper is concerned with the evaluation of $z(u) = P(S_n > u)$ by simulation, where $S_n = Y_1 + \cdots + Y_n$ and the Y_i are positive, independent, identically distributed, heavy-tailed random variables, in situations where u is large and $z(u)$ is thus small. By definition $S_0 = 0$. We will also consider the case in which n is replaced by a nonnegative, integer-valued random variable N independent of $\{Y_i\}$. An example where this is of relevance is the steady-state waiting time of the M/G/1 queue, which, according to the Pollaczek–Khinchin formula, has the same distribution as S_N for a certain choice of parameters (see [4, p. 237]). In insurance risk, $P(S_N > u)$ is also a representation of the ruin probability with initial reserve u (see [4, p. 399]). In financial mathematics, such compound Poisson sums with independent terms occur widely in models for operational risk (see, e.g. [11] and [14]). In the last two examples, the interesting values of $z(u)$ may be of order 10^{-2}. In the first (arising from problems in telecommunications

and data transmission, where $z(u)$ could be a bit loss rate) the magnitude could drop to, say, 10^{-10}. As is well known (see, e.g. [6] and [16]), the efficiency of crude Monte Carlo simulation greatly deteriorates as $z(u)$ decreases, making the simulation a nontrivial problem requiring variance reduction ideas.

Let F denote the common distribution of the Y_i. Their being heavy tailed means, in general terms, that their exponential moments fail to exist. However, most often a more narrow framework is considered, e.g. the case in which F is subexponential [12, pp. 36–57, 564–587], [20], or even just the special case (of primary importance) in which F is regularly varying with a relatively small index, α; this means that the tail $\bar{F}(x)$ equals $L(x)/x^\alpha$ for some slowly varying function $L(\cdot)$. We will not go into the general subexponential framework here but be satisfied with treating the regularly varying case as well as what is perhaps the second-most important example: the heavy-tailed Weibull case in which $\bar{F}(x) = e^{-u^\beta}$ with $0 < \beta < 1$. The relevance to applications of such modeling assumptions has been convincingly argued (see, e.g. [12, pp. 1–19] and [2]).

We call a random variable $Z \equiv Z(u)$ an *estimator* for $z(u)$ if Z can be generated by simulation and $\mathrm{E}^*[Z] = z(u)$, where E^* is the expectation corresponding to $\mathrm{P}^* \equiv \mathrm{P}^{*,u}$, the probability measure used in the simulation (the given distribution of $Y_1, Y_2, \ldots$ or an importance sampling distribution). As is standard, we say that $Z(u)$ has *bounded relative error* to indicate that $\mathrm{var}(Z(u))/z(u)^2$ is bounded in u, and say that $Z(u)$ is *polynomial time* to indicate that $\mathrm{var}(Z(u))/z(u)^{2-\varepsilon}$ is bounded in u for any $\varepsilon > 0$ (often the term *logarithmic efficient*, or just *efficient*, is also used to describe this [6], [16], [7], [17]). Similar terminology is used in the case of a random number, N, of random variables. With light tails, the most established approach for simulation of $z(u)$ is the *exponential change of measure*, as determined by the saddle-point method; see [4, pp. 373–376]. As discussed there, this scheme can be seen as an implementation of the general principle in importance sampling of taking the changed measure used for the simulation to be as close as possible to the conditional distribution given the rare event. In the present case this means sampling $Y_1, \ldots, Y_n$ using an asymptotic description of their conditional distribution $\mathrm{P}^{n,u}$ given $S_n > u$. The traditional description in the subexponential setting states that one of the Y_i is larger than u and the rest are in some sense 'typical', that is, unaffected by the conditioning (for the precise statements, see Proposition 1.2 of [3, p. 252] and Lemma 6.6 of [4, p. 405]). However, as noted in [7], the most straightforward ideas of how to use this asymptotics as a basis for importance sampling fail. The first polynomial time algorithm reported in the literature (see [5]) in fact uses a different idea, namely conditional Monte Carlo conditioning, invoking the order statistics $Y_{(1)}, \ldots, Y_{(n)}$, $Y_{(1)} < \cdots < Y_{(n)}$, for the conditioning. The estimator is

$$\mathrm{P}(S_n > u \mid Y_{(1)}, \ldots, Y_{(n-1)}) = \frac{\bar{F}(Y_{(n-1)} \vee (u - S_{(n-1)}))}{\bar{F}(Y_{(n-1)})}, \qquad (1.1)$$

where $S_{(n-1)} = Y_{(1)} + \cdots + Y_{(n-1)}$ and $x \vee y$ is shorthand notation for $\max(x, y)$. Later, polynomial time importance sampling ideas were given in [7] and [17]. Here we only consider the (more efficient) ideas of [17], which are based upon the hazard rate, $\Lambda(x) = -\log \bar{F}(x)$. A key ingredient (not the only one) in the algorithms of [17] is *hazard rate twisting*, which changes $\Lambda(x)$ to $\theta \Lambda(x)$ for some small θ. That is, the tail $\bar{F}$ is changed to $\bar{F}_\theta = \bar{F}^\theta$. The more refined *weighted delayed hazard rate twisting* simulates the Y_i from a distribution which is a mixture of F conditioned to $(0, x^*]$ and F_θ conditioned to (x^*, ∞), with the weights, θ and x^*, chosen to depend appropriately on u.

Despite the good computational and theoretical properties of the algorithms reported in [17], it appears intuitively unnatural that the importance sampling change of measure is independent

and identically distributed. Indeed, the above description of $P^{n,u}$ is highly asymmetric, showing the particular role taken by one of the Y_i (the large one). Our contribution in this paper is to present some modified algorithms that take this into account, and which will turn out to yield very substantial efficiency improvements over existing algorithms. Our algorithms depart from the identity

$$P(S_n > u) = nP(S_n > u, \, M_n = Y_n),$$

where, since $M_k = \max(Y_1, \ldots, Y_k)$, we have $Y_{(n)} = M_n$. If f is the density of F and f^* the density of an importance sampling distribution (we return to the choice of f^* later), we twist only the distribution of Y_n, and arrive at the estimator

$$n \frac{f(Y_n)}{f^*(Y_n)} \, \mathbf{1}(S_n > u, \, M_n = Y_n). \tag{1.2}$$

Using conditional Monte Carlo conditioning instead yields the estimator

$$nP(S_n > u, \, M_n = Y_n \mid Y_1, \ldots, Y_{n-1}) = n \bar{F}(M_{n-1} \vee (u - S_{n-1})). \tag{1.3}$$

The paper is organized as follows. In Section 2, we give a theoretical proof of the efficiency of our estimators for the Pareto case and present some numerical studies. In Section 3, we address the same issues for the Weibull case, where it turns out that there is a certain critical value of β for which the estimator in (1.3) is polynomial time. The empirical performance of the estimators in (1.2) and, in particular, (1.3) is excellent. For example, for an M/G/1 queue with Pareto service times of index $\alpha = 1.5$, Table 1 below shows that (1.3) reduces the variance by a factor of between five and 14 compared to [5], and a factor of between 33 and 529 compared to [17]. The performance improvement over [17] for Weibull tails is similar, while the algorithm of [5] is not even polynomial in this case. In addition to this, we would like to point out the simplicity of (1.3), which makes the code much shorter and more transparent than it is for some of the estimators we compare it with.

In Section 4, we discuss combinations with other variance reduction ideas (control variates and stratification) in the case in which $n = N$ is random as in the examples above.

2. The regularly varying case

In this section, we assume that $f(x) = L(x)/(1+x)^{1+\alpha}$, $x > 0$, with L slowly varying and $\alpha > 0$ (we do not need conditions like $\alpha \geq 1$, to exclude infinite mean, etc.). Then $\bar{F}(x) \sim L(x)/(1+x)^{\alpha}$ by Karamata's theorem (see [13, pp. 275–284]). Here the notation '$\sim$' has the following meaning: $a(x) \sim b(x) \Leftrightarrow \lim_{x \to \infty} a(x)/b(x) = 1$.

We first consider the estimator in (1.2). It remains to specify the importance sampling density $f^*(y)$. As in [17] and [9], we take f^* to be regularly varying with index α^* of the form $b/\log u$, and, for simplicity, to be just Pareto: $f^*(x) = \alpha^*/(1+x)^{1+\alpha^*}$.

Theorem 2.1. (a) *With f^* as stated, the estimator (1.2) is polynomial time for any fixed n.*

(b) *More precisely, if there exist u_0 and z_0 such that $L(uz)/L(u)$ is either monotonically increasing or monotonically increasing in u, $u \geq u_0$, for all $z \geq z_0$, then an asymptotic upper bound for the squared relative error is $c_1 n^2 \log u$, where c_1 is a constant independent of n.*

(c) *In the random-N case, the estimator is polynomial time provided that $E[N^{3\alpha+3}] < \infty$.*

(Some discussion of the condition in part (b) is given in Remark 2.1, below.)

In the proof, we need the density, g^*, proportional to $L(x)^2/(1+x)^{1+2\alpha-\alpha^*}$ (note the dependence on u via $\alpha^* = b/\log u$). That is, $g^* = f^2/(f^*c^*)$, where

$$c^* = \int_0^\infty \frac{f^2(x)}{f^*(x)}\, \mathrm{d}x \sim c_2 \log u, \qquad c_2 = b^{-1} \int_0^\infty \frac{L(x)^2}{(1+x)^{2\alpha+1}}\, \mathrm{d}x.$$

Lemma 2.1. *Let Y^* have density g^*. Then,* (a) *for any $\varepsilon > 0$, there is a constant, c_ε, such that $\mathrm{P}(Y^* > u) \le c_\varepsilon(1+u)^{-(2\alpha-\varepsilon)}$ for all large u, and,* (b) *under the conditions of Theorem 2.1(b), $\mathrm{P}(Y^* > u) \sim c_3 \bar{F}(u)^2 \log u$.*

Proof. We have

$$\mathrm{P}(Y^* > u) \sim c_4 \int_u^\infty \frac{L(z)^2}{(1+z)^{2\alpha-\alpha^*+1}}\, \mathrm{d}z \sim c_4 \int_u^\infty \frac{L(z)^2}{z^{2\alpha-\alpha^*+1}}\, \mathrm{d}z.$$

From this (a) follows by noting that we can bound α^* by ε for sufficiently large u, and that $\int_u^\infty L(z)^2/z^{1+\beta}\, \mathrm{d}z \sim L(u)^2/u^\beta$ by Karamata's theorem. For (b), substitute $z = uy$ to obtain

$$\mathrm{P}(Y^* > u) \sim c_4 \frac{L(u)^2}{u^{2\alpha-\alpha^*}} \int_1^\infty \frac{L(uy)^2/L(u)^2}{(1+y)^{2\alpha-\alpha^*+1}}\, \mathrm{d}y.$$

Here $u^{\alpha^*} \to b$ whereas $L(uy)/L(u) \to 1$ for all y, as $u \to \infty$. The convergence is in fact uniform on $[1, z_0]$, so using monotone convergence on (z_0, ∞) ensures the existence of a limit of the integral. This completes the proof.

Proof of Theorem 2.1. Let $A = \{x_1 + \cdots + x_n > u,\ x_n = \max_{k \le n} x_k\}$. The second moment of (1.2) is then

$$n^2 \int \cdots \int_A \frac{f^2(x_n)}{f^*(x_n)}\, \mathrm{d}x_n \prod_{i=1}^{n-1} f(x_i)\, \mathrm{d}x_i = n^2 c^* \int \cdots \int_A g^*(x_n)\, \mathrm{d}x_n \prod_{i=1}^{n-1} f(x_i)\, \mathrm{d}x_i$$

$$= n^2 c^* \mathrm{P}^{**}(M_n = Y_n,\ S_n > u),$$

where P^{**} is the probability measure under which the Y_i are independent, with Y_n having density g^* and the rest having density f.

Let $\beta \in (\tfrac{2}{3}, 1)$ and let

$$u' = \frac{u^\beta}{n-1} \quad \text{and} \quad u'' = u - (n-1)u' = u(1 - u^{\beta-1}).$$

Then $Y_n > u''$ on the event $\{M_{n-1} \le u'\}$. It follows that

$$\mathrm{P}^{**}(S_n > u,\ M_n = Y_n) \le \mathrm{P}^{**}(Y_n > u'') + \mathrm{P}^{**}(M_{n-1} > u',\ Y_n > u')$$

$$\le \mathrm{P}^{**}(Y_n > u'') + n\mathrm{P}^{**}(Y_n > u')\mathrm{P}^{**}(Y_1 > u').$$

Choose $\delta, \varepsilon > 0$ such that $2\alpha + \delta < 2\alpha + \delta + \varepsilon\beta < 3\alpha\beta$. Then, using Lemma 2.1(a), the second term, i.e. $n\mathrm{P}^{**}(Y_n > u')\mathrm{P}^{**}(Y_1 > u')$, can be bounded by

$$c_5 n \left(\frac{n-1}{u^\beta}\right)^{2\alpha-\varepsilon} \left(\frac{n-1}{u^\beta}\right)^{\alpha-\varepsilon} = c_5 n \frac{(n-1)^{3\alpha-2\varepsilon}}{u^{3\alpha\beta-2\varepsilon\beta}} \le c_5 n \frac{(n-1)^{3\alpha-2\varepsilon}}{u^{2\alpha+\delta}}.$$

Again by Lemma 2.1(a), $(u(1 - u^{\beta-1}))^{-(2\alpha-\varepsilon)}$ is an upper bound for the first term. Since the asymptotics of this expression is $u^{-(2\alpha-\varepsilon)}$, part (a) of the theorem follows.

Part (b) of the theorem now follows immediately by invoking part (b) of Lemma 2.1 rather than part (a). To prove part (c) of the theorem, simply condition on $\{N = n\}$ and use dominated convergence to control the second term.

Remark 2.1. The improvement over hazard rate twisting obtained when using (1.2) can be understood as follows. Assume, for simplicity, that F is Pareto with density $\alpha/(1 + x)^{\alpha+1}$, $x > 0$, and that n is fixed. The second moment in hazard rate twisting is

$$\int_{x_1+\cdots+x_n>u} \cdots \int \prod_{i=1}^{n} \frac{f^2(x_i)}{f^*(x_i)} \, dx_i = \left(\frac{\alpha^2}{\alpha^*\alpha_x} \right)^n \int_{x_1+\cdots+x_n>u} \cdots \int \prod_{i=1}^{n} f_u(x_i) \, dx_i$$

$$= \left(\frac{\alpha^2}{\alpha^*\alpha_u} \right)^n \bar{F}_u^{*n}(x),$$

where F_u is the Pareto distribution with parameter $\alpha_u = 2\alpha - \alpha^*$.

Simple modifications of arguments of [13, pp. 278–279] show that $\bar{F}_u^{*n}(u) \sim n e^b u^{-2\alpha}$, as expected from the fact that $u^{\alpha^*} \to e^b$. Thus, the asymptotics of the second moment is

$$e^b \left(\frac{\alpha}{2b} \right)^n n (\log u)^n \bar{F}(u)^2.$$

Comparison with Theorem (2.1) then shows that (1.2) improves the squared relative error by a factor of $(\log u)^{n-1}$.

An immediate corollary of Theorem 2.1 is that the estimator (1.3) is also polynomial time. To see this, note that (1.3) is the conditional expectation of (1.2) given $Y_1, \ldots, Y_{n-1}$ (the choice of f^* is immaterial here) and recall the well-known fact that conditioning reduces variance. In fact, the following, stronger result holds.

Theorem 2.2. *The estimator (1.3) has bounded relative error, assuming, in the case of a random N, that*

$$\limsup_{u\to\infty} \frac{E[L(u/2N)^2 N^{2\alpha+4}]}{L_1(u)^2} < \infty \qquad (2.1)$$

for some function $L_1(u)$ satisfying $L_1(u) \sim L(u)$.

Remark 2.2. In the examples we considered above, condition (2.1) is equivalent to or only marginally stronger than $E[N^{2\alpha+4}] < \infty$. To see this, note that $L(u/2N)/L_1(u) \to 1$ since $L(u)$ is slowly varying. Thus, a sufficient condition for the left-hand side of (2.1) to equal $E[N^{2\alpha+4}]$ is that $L(ut)/L_1(u)$ be bounded or monotone in u for a fixed t (which is essentially the condition of Theorem 2.1). If, e.g. $L(x) = (\log(a+x))^\beta$, then we can let $L_1(x) = (\log x)^\beta$, to have

$$\frac{L(ut)}{L_1(u)} = \left(1 + \frac{\log(t + a/u)}{\log u} \right)^\beta,$$

which is decreasing for $\beta > 0$ and increasing for $\beta < 0$ (note that $\log(t + a/u)/\log u$ is decreasing, it being the ratio between a decreasing function and an increasing function).

Proof of Theorem 2.2. We split the second moment in (1.3) (with $n + 1$ instead of n for notational convenience, and omitting the factor $(n + 1)^2$) into the three parts

$$v_1 = \mathrm{E}[\bar{F}(M_n \vee (u - S_n))^2; M_n > u/2],$$
$$v_2 = \mathrm{E}[\bar{F}(M_n \vee (u - S_n))^2; M_n \leq u/2, Y_{(n-1)} \leq \varepsilon u],$$
$$v_3 = \mathrm{E}[\bar{F}(M_n \vee (u - S_n))^2; M_n \leq u/2, Y_{(n-1)} > \varepsilon u],$$

where $\varepsilon = 1/2n$. Here $v_1 \leq \bar{F}(u/2)^2 \sim 2^{2\alpha} \bar{F}(u)^2$. If $M_n \leq u/2$ and $Y_{(n-1)} \leq \varepsilon u$, then $S_n \leq u(1 - 1/2n)$ and, hence,

$$v_2 \leq \bar{F}\left(\frac{u}{2n}\right)^2 \leq \frac{L(u)^2}{u^{2\alpha}} \frac{L(u/2n)^2 (2n)^{2\alpha}}{L(u)^2} \sim \bar{F}(u)^2 (2n)^{2\alpha}.$$

Finally,

$$v_3 \leq \mathrm{P}(Y_{(n-1)} > \varepsilon u) \leq n^2 \bar{F}(\varepsilon u)^2 \leq n^2 \frac{L(u)^2}{u^{2\alpha}} \frac{L(u/2n)^2 (2n)^{2\alpha}}{L(u)^2} \sim 4^{\alpha} n^{2+2\alpha} \bar{F}(u)^2.$$

Combining these estimates completes the proof.

For the case of a random N, just condition upon $\{N = n+1\}$, to find that the squared relative error of the estimator, Z, given by the right-hand side of (1.3) (where N replaces n) is

$$\frac{\mathrm{E}^*[Z^2]}{\bar{F}(u)^2} \sim \sum_{n=0}^{\infty} \mathrm{P}(N = n + 1)(n + 1)^2 \left(1 + (n^{2\alpha} + n^{2\alpha+2})4^{\alpha} \frac{L(u/2n)^2}{L(u)^2}\right).$$

Therefore, the relative error is bounded provided that (2.1) holds.

Theorem 2.2 appears to give the first example of an estimator with bounded relative error in a heavy-tailed setting. However, asymptotic efficiency does not guarantee efficiency for any given set of parameters (a good example of this is the importance sampling algorithm of [7]). Nevertheless, complexity studies can guide the decision of whether or not to proceed with an estimator. Encouraged by Theorems 2.1 and 2.2, we performed a comparison of different estimators for the Pareto case $\bar{F}(x) = (1+x)^{-\alpha}$ and the M/G/1 queue in which N is geometric with

$$\mathrm{P}(N = n) = (1 - \rho)\rho^n, \qquad n = 0, 1, 2, \ldots,$$

and

$$f(y) = \frac{\mathrm{P}(U > y)}{\mathrm{E}[U]},$$

where U is a generic service time. Here, $\rho = \lambda \mathrm{E}[U]$ is the traffic intensity, with λ the arrival rate. Thus, F plays the role of the integrated tail distribution, so the service time distribution itself is Pareto, but with index $\alpha_U = \alpha + 1$ instead of α. The most important range for α_U is often argued to be the interval $(1, 2)$ (implying finite mean but infinite variance), so we chose $\alpha_U = \frac{3}{2}$, corresponding to $\alpha = \frac{1}{2}$ and supplemented this with the larger value $\alpha = \frac{3}{2}$, corresponding to $\alpha_U = \frac{5}{2}$. Three different traffic intensities, $\rho = 0.25, 0.5, 0.75$, were considered, whereas for u we considered the four values for which the standard approximation

$$\mathrm{P}(W > u) \sim \frac{\rho}{1 - \rho} \bar{F}(u) \tag{2.2}$$

is 10^{-k}, $k = 2, 5, 8, 11$.

TABLE 1: Results for the Pareto case with $\alpha = 0.5$.

ρ	k	(1.1)	(1.2)$_{\text{CE}}$	(1.3)	JS	(1.2)$_{\text{JS}}$	Estimator (1.3)
0.25	2	0.071	0.152	0.032	0.185	0.224	9.9928×10^{-3}
0.25	5	0.105	0.260	0.031	0.421	0.506	1.0000×10^{-5}
0.25	8	0.122	0.335	0.031	0.582	0.703	9.9980×10^{-9}
0.25	11	0.115	0.397	0.031	0.713	0.859	9.9985×10^{-12}
0.50	2	0.111	0.192	0.045	0.253	0.402	9.9945×10^{-3}
0.50	5	0.144	0.301	0.044	0.515	0.812	1.0004×10^{-5}
0.50	8	0.146	0.380	0.044	0.702	1.098	9.9989×10^{-9}
0.50	11	0.153	0.445	0.044	0.855	1.337	9.9996×10^{-12}
0.75	2	0.141	0.232	0.054	0.314	0.744	9.9958×10^{-3}
0.75	5	0.205	0.341	0.054	0.591	1.381	1.0003×10^{-5}
0.75	8	0.188	0.423	0.054	0.795	1.888	1.0005×10^{-8}
0.75	11	0.180	0.494	0.054	0.960	2.272	1.0003×10^{-11}

TABLE 2: Results for the Pareto case with $\alpha = 1.5$.

ρ	k	(1.1)	(1.2)$_{\text{CE}}$	(1.3)	JS	(1.2)$_{\text{JS}}$	Estimator (1.3)
0.25	2	0.100	0.169	0.051	0.200	0.255	1.1216×10^{-2}
0.25	5	0.150	0.260	0.031	0.420	0.518	1.0021×10^{-5}
0.25	8	0.124	0.335	0.031	0.583	0.702	1.0001×10^{-8}
0.25	11	0.102	0.396	0.031	0.715	0.861	9.9998×10^{-12}
0.50	2	0.161	0.234	0.077	0.282	0.492	1.2606×10^{-2}
0.50	5	0.201	0.302	0.044	0.514	0.835	1.0027×10^{-5}
0.50	8	0.152	0.381	0.044	0.703	1.113	1.0002×10^{-8}
0.50	11	0.149	0.446	0.044	0.854	1.340	9.9966×10^{-12}
0.75	2	0.212	0.333	0.114	0.361	0.933	1.5297×10^{-2}
0.75	5	0.201	0.342	0.054	0.589	1.469	1.0044×10^{-5}
0.75	8	0.189	0.422	0.054	0.795	1.848	9.9948×10^{-9}
0.75	11	0.231	0.492	0.054	0.961	2.298	1.0005×10^{-11}

In the implementation we used

$$P(W > u) = P(Y_1 + \cdots + Y_N > u) = \rho P(Y_1 + \cdots + Y_{N^*} > u),$$

where

$$P(N^* = n) = (1 - \rho)\rho^{n-1}, \qquad n = 1, 2, \ldots,$$

and simulated using N^*. The algorithms were replicated $R = 10^7$ times. Tables 1 and 2 give the corresponding relative errors defined as the half-widths of the 95% confidence interval divided by the point estimate.

In the list of algorithms in the tables, (1.1) and (1.3) are self-explanatory. JS is the weighted delayed hazard rate twisting of [17], implemented using the parameters suggested there. The asymmetric importance sampling (1.2) was implemented for two different densities: (1.2)$_{\text{CE}}$ means a Pareto f^* with $\alpha^* = 1/\log u$, as suggested by a cross-entropy argument similar to that of [9] and [19] (we omit the details of the derivation), and (1.2)$_{\text{JS}}$ corresponds to the same f^* as was used in algorithm JS. The final column gives the point estimates, calculated using the most accurate procedure, namely (1.3).

The point estimates show excellent agreement with approximation (2.2) in the present Pareto case, but we note that regularly varying cases in which (2.2) is quite inaccurate (and, hence, simulation is a realistic alternative) have been reported in, e.g. [1] and [18].

The findings in Tables 1 and 2 are that the conditional Monte Carlo algorithms (1.1) and (1.3) perform better than any of the importance sampling algorithms, with (1.3) representing a substantial improvement over (1.1). The weighted delayed hazard rate twisting in algorithm JS is substantially more efficient than the naive twisting of α in the CE algorithm, but when combined with asymmetric importance sampling, the CE algorithm is improved very substantially and the JS algorithm is not. A possible intuitive argument as to why this is the case is that the role of the delay in the JS algorithm (represented by the change point x^*) is to ensure that a sufficient number of the Y_i are not large, even if $Y_1 + \cdots + Y_N$ is.

3. The Weibull case

In this section, we assume that F is *Weibull-like* (in the terminology of [18]), meaning that the density $f(x)$ is asymptotically of the form $cx^\gamma e^{-x^\beta}$ where $0 < \beta < 1$. The tail then satisfies

$$\bar{F}(x) \sim \frac{cx^{1+\gamma-\beta}e^{-x^\beta}}{\beta}.$$

We will only give a theoretical study of the efficiency properties of the estimator (1.3), and not of (1.2). This is motivated by our findings (both theoretical and empirical) in the Pareto case and the numerical studies below.

Theorem 3.1. *Assume that* $\beta < \bar{\beta} = \log \frac{3}{2} / \log 2 = 0.585$, *i.e.* $2^{1+\beta} < 3$. *Then the estimator (1.3) is polynomial time for any fixed n.*

Remark 3.1. Inspection of the proof of Theorem 3.1 given in Appendix A shows that the condition on β is indeed necessary. This may not be surprising, since the algorithms are specifically designed for heavy tails and, as $\beta \uparrow 1$, we approach the border, $\beta = 1$, to the light tails region. The occurrence of a critical value of β is not uncommon for the Weibull distribution; see, e.g. [8], where some of the results take a different form for $\frac{1}{2} < \beta < 1$, $\frac{1}{3} < \beta < \frac{1}{2}$, and so on. The criticality of having $\beta < \frac{1}{2}$ for a certain result to hold has been found in many later cases and is often referred to as *square root insensitivity* (see, e.g. [10] and references therein). The authors have not seen the present critical value, 0.585, appear before.

Theorem 3.1 is obviously weaker than the corresponding Theorem 2.1 for the Pareto case, which additionally features bounded relative error and is valid in the random-N case. The numerical examples presented below strongly suggest that both of these extensions also hold for the Weibull case, but we have no proof of this.

Tables 3–5 contain the results of a geometric sum numerical study similar to that for the Pareto case, considering a (standard) Weibull F with tail e^{-x^β} and three different values, 0.25, 0.50, and 0.75, for β (we omit algorithm (1.1) since it is not polynomial time in the Weibull case). The choices of F and $\beta = 0.5$ are as in [17] and, as a check, we reconstructed Table 2 of [17] and obtained very similar point estimates and confidence bands (in fact, algorithm JS gave slightly better results here than in [17], probably due to our using N^* instead of N). We do not include the table here. Note, however, that e^{-x^β} is not the tail of an integrated tail distribution (as the derivative is not monotone) and, therefore, that there is no direct interpretation in terms of an M/G/1 queue.

TABLE 3: Results for the Weibull case with $\beta = 0.25$.

ρ	k	$(1.2)_{\text{CE}}$	(1.3)	JS	$(1.2)_{\text{JS}}$	Estimator (1.3)
0.25	2	0.154	0.035	0.186	0.228	1.0152×10^{-2}
0.25	5	0.261	0.032	0.419	0.508	1.0040×10^{-5}
0.25	8	0.335	0.031	0.583	0.702	1.0008×10^{-8}
0.25	11	0.396	0.031	0.716	0.860	1.0004×10^{-11}
0.50	2	0.200	0.052	0.258	0.420	1.0545×10^{-2}
0.50	5	0.303	0.045	0.513	0.821	1.0097×10^{-5}
0.50	8	0.380	0.044	0.702	1.099	1.0018×10^{-8}
0.50	11	0.446	0.044	0.854	1.336	1.0005×10^{-11}
0.75	2	0.256	0.071	0.331	0.823	1.1468×10^{-2}
0.75	5	0.347	0.056	0.588	1.406	1.0215×10^{-5}
0.75	8	0.424	0.054	0.795	1.867	1.0049×10^{-8}
0.75	11	0.494	0.054	0.963	2.238	1.0023×10^{-11}

TABLE 4: Results for the Weibull case with $\beta = 0.5$.

ρ	k	$(1.2)_{\text{CE}}$	(1.3)	JS	$(1.2)_{\text{JS}}$	Estimator (1.3)
0.25	2	0.171	0.054	0.201	0.255	1.4027×10^{-2}
0.25	5	0.334	0.072	0.432	0.677	2.0509×10^{-5}
0.25	8	0.378	0.056	0.590	0.859	1.9035×10^{-8}
0.25	11	0.414	0.070	0.719	0.886	1.6725×10^{-11}
0.50	2	0.261	0.098	0.295	0.515	2.8526×10^{-2}
0.50	5	0.614	0.185	0.650	1.415	9.9016×10^{-5}
0.50	8	0.465	0.097	0.733	1.939	1.1539×10^{-7}
0.50	11	0.502	0.101	0.876	1.486	7.3482×10^{-11}
0.75	2	0.388	0.153	0.315	0.863	9.6750×10^{-2}
0.75	5	7.128	0.665	3.820	8.015	2.7456×10^{-3}
0.75	8	1.319	0.597	1.101	3.077	4.0061×10^{-5}
0.75	11	0.683	0.118	1.068	10.251	2.5753×10^{-7}

TABLE 5: Results for the Weibull case with $\beta = 0.75$.

ρ	k	$(1.2)_{\text{CE}}$	(1.3)	JS	$(1.2)_{\text{JS}}$	Estimator (1.3)
0.25	2	0.198	0.082	0.222	0.290	1.1380×10^{-2}
0.25	5	1.580	0.476	1.208	2.534	1.1059×10^{-5}
0.25	8	6.673	1.296	2.652	7.588	1.0519×10^{-8}
0.25	11	2.951	4.532	1.927	6.643	1.0337×10^{-11}
0.50	2	0.304	0.135	0.268	0.487	1.5223×10^{-2}
0.50	5	2.815	1.110	3.689	7.396	1.3290×10^{-5}
0.50	8	10.451	8.043	19.517	25.075	1.1556×10^{-8}
0.50	11	7.396	13.190	13.483	61.513	1.1001×10^{-11}
0.75	2	0.329	0.141	0.189	0.516	3.1421×10^{-2}
0.75	5	2.088	0.713	1.326	4.067	2.5138×10^{-5}
0.75	8	13.584	5.239	12.860	37.925	1.5581×10^{-8}
0.75	11	86.575	43.958	70.867	193.427	1.3302×10^{-11}

The algorithms show a considerable performance degradation not just for $\beta = 0.75$ (which is above the critical value, $\bar{\beta}$) but even for $\beta = 0.5$ (which is close to $\bar{\beta}$), when $\rho = 0.75$. For other parameter values, the conclusions of Tables 3 and 4 are very similar to those in the Pareto case.

4. Combination with stratification and control variates

The subexponential asymptotics

$$P(S_n > u) \sim n\bar{F}(u), \tag{4.1}$$

which is valid for a fixed n, indicates that a substantial part of the variability of the estimators in the random-N tables may be due to the variability in N. Two ways to eliminate this are to use N as control variate (as suggested by the linearity of (4.1) in n) or to stratify N. Note that both methods guarantee variance reduction. We refer the reader to, e.g. [15, pp. 186–279] for the basic facts on these variance reduction techniques (for control variates, we used the version with an estimated control coefficient [15, p. 187, Equation (4.5)]). Tables 6–10 contain the results of numerical studies pertaining to this issue, using the same parameter values as in Tables 1–5. For the stratification (using proportional allocation; cf. [15, pp. 209–235]), we took eight strata, $N^* = 1, \ldots, 7$, and $N^* > 7$ for $\rho = 0.25$, and 17 strata, $N^* = 1, \ldots, 16$, and $N^* > 16$ for both $\rho = 0.50$ and $\rho = 0.75$. The tables again give the half-widths of the confidence intervals, and the subscripts 'CV' and 'str.' respectively denote combination of the relevant algorithm with control variates and stratification. Our results are precise to three decimal places, so '0.000' simply indicates a result less than 5×10^{-4}.

The conclusion to be drawn is that the control variate method and stratification perform rather similarly. The variance reduction is by far the largest for algorithm (1.3), which we take as indication that the algorithm estimates $P(S_n > u)$ very accurately for a fixed n and that the variability of the M/G/1 estimators is largely due to the variability in N. In contrast, for the other estimators the variability in the estimates of $P(Y_1 + \cdots + Y_n > u)$ is nonnegligible compared to the variation in N.

TABLE 6: Variance reduction results for the Pareto case with $\alpha = 0.5$.

ρ	k	$(1.2)_{CV}$	$(1.2)_{str.}$	$(1.3)_{CV}$	$(1.3)_{str.}$	JS_{CV}	$JS_{str.}$
0.25	2	0.150	0.121	0.008	0.008	0.132	0.181
0.25	5	0.258	0.210	0.000	0.000	0.303	0.419
0.25	8	0.335	0.272	0.000	0.000	0.420	0.583
0.25	11	0.396	0.323	0.000	0.000	0.513	0.713
0.50	2	0.192	0.179	0.009	0.009	0.222	0.248
0.50	5	0.301	0.285	0.000	0.000	0.458	0.512
0.50	8	0.381	0.362	0.000	0.000	0.625	0.700
0.50	11	0.445	0.424	0.000	0.000	0.761	0.855
0.75	2	0.232	0.225	0.009	0.011	0.301	0.308
0.75	5	0.341	0.334	0.000	0.005	0.572	0.588
0.75	8	0.423	0.419	0.000	0.005	0.773	0.793
0.75	11	0.491	0.487	0.000	0.005	0.938	0.961

TABLE 7: Variance reduction results for the Pareto case with $\alpha = 1.5$.

ρ	k	$(1.2)_{\text{CV}}$	$(1.2)_{\text{str.}}$	$(1.3)_{\text{CV}}$	$(1.3)_{\text{str.}}$	JS_{CV}	$\text{JS}_{\text{str.}}$
0.25	2	0.169	0.143	0.025	0.024	0.159	0.193
0.25	5	0.259	0.211	0.001	0.001	0.302	0.418
0.25	8	0.334	0.272	0.000	0.000	0.419	0.582
0.25	11	0.396	0.323	0.000	0.000	0.514	0.712
0.50	2	0.234	0.220	0.043	0.038	0.259	0.273
0.50	5	0.302	0.286	0.001	0.001	0.457	0.512
0.50	8	0.380	0.362	0.000	0.000	0.626	0.702
0.50	11	0.445	0.425	0.000	0.000	0.760	0.851
0.75	2	0.329	0.315	0.074	0.069	0.344	0.348
0.75	5	0.342	0.336	0.002	0.006	0.573	0.587
0.75	8	0.423	0.418	0.000	0.005	0.772	0.797
0.75	11	0.492	0.486	0.000	0.005	0.937	0.963

TABLE 8: Variance reduction results for the Weibull case with $\beta = 0.25$.

ρ	k	$(1.2)_{\text{CV}}$	$(1.2)_{\text{str.}}$	$(1.3)_{\text{CV}}$	$(1.3)_{\text{str.}}$	JS_{CV}	$\text{JS}_{\text{str.}}$
0.25	2	0.153	0.124	0.011	0.011	0.135	0.182
0.25	5	0.260	0.212	0.005	0.005	0.302	0.418
0.25	8	0.335	0.273	0.002	0.001	0.419	0.582
0.25	11	0.396	0.322	0.000	0.000	0.511	0.714
0.50	2	0.200	0.188	0.018	0.018	0.230	0.253
0.50	5	0.303	0.289	0.007	0.007	0.457	0.511
0.50	8	0.381	0.361	0.002	0.002	0.625	0.698
0.50	11	0.446	0.424	0.001	0.001	0.762	0.853
0.75	2	0.257	0.248	0.031	0.032	0.318	0.323
0.75	5	0.347	0.341	0.011	0.012	0.571	0.584
0.75	8	0.424	0.419	0.003	0.006	0.771	0.790
0.75	11	0.494	0.488	0.001	0.006	0.939	0.964

TABLE 9: Variance reduction results for the Weibull case with $\beta = 0.5$.

ρ	k	$(1.2)_{\text{CV}}$	$(1.2)_{\text{str.}}$	$(1.3)_{\text{CV}}$	$(1.3)_{\text{str.}}$	JS_{CV}	$\text{JS}_{\text{str.}}$
0.25	2	0.170	0.146	0.029	0.028	0.162	0.195
0.25	5	0.344	0.318	0.054	0.057	0.339	0.431
0.25	8	0.412	0.326	0.063	0.046	0.446	0.587
0.25	11	0.411	0.350	0.020	0.025	0.534	0.718
0.50	2	0.261	0.247	0.062	0.058	0.272	0.283
0.50	5	0.789	0.565	0.140	0.140	0.620	0.607
0.50	8	0.635	0.585	0.090	0.126	0.678	0.733
0.50	11	0.503	0.486	0.044	0.032	0.802	0.874
0.75	2	0.373	0.367	0.109	0.101	0.290	0.295
0.75	5	2.266	2.269	0.638	0.788	2.846	2.241
0.75	8	1.063	0.948	0.241	0.259	1.063	1.101
0.75	11	0.736	0.662	0.074	0.077	1.059	1.072

TABLE 10: Variance reduction results for the Weibull case with $\beta = 0.75$.

ρ	k	$(1.2)_{CV}$	$(1.2)_{str.}$	$(1.3)_{CV}$	$(1.3)_{str.}$	JS_{CV}	$JS_{str.}$
0.25	2	0.197	0.176	0.049	0.047	0.190	0.209
0.25	5	1.492	1.469	0.471	0.418	1.864	2.263
0.25	8	4.205	4.374	1.726	1.581	1.893	3.403
0.25	11	2.303	1.778	2.099	1.864	3.474	1.725
0.50	2	0.290	0.283	0.091	0.085	0.240	0.247
0.50	5	3.050	3.249	1.105	0.956	3.619	3.578
0.50	8	14.864	27.744	11.819	15.355	17.372	34.213
0.50	11	17.251	9.784	11.914	20.099	28.343	10.382
0.75	2	0.266	0.311	0.101	0.096	0.156	0.159
0.75	5	2.020	2.009	0.659	0.594	1.268	1.306
0.75	8	15.617	16.337	4.926	4.950	13.290	12.996
0.75	11	46.885	81.367	32.731	41.419	174.065	74.014

Appendix A. Proof of Theorem 3.1

In the proof of Theorem 3.1 we make use of Figure 1, where the two areas shaded differently together form the support of the distribution of (M_n, S_n). Note that $u - y > x$ in the dark-shaded area and that $x > u - y$ in the light-shaded area. For each u, we divide the support into the $2n - 1$ regions labeled $0, 1', \ldots, (n-1)', 1'', \ldots, (n-1)''$ shown in Figure 1 (e.g. region k' is the triangle with border lines $u = x + y$, $y = kx$, and $y = (k+1)x$). We will denote by $\gamma_1, \gamma_2, \ldots$ certain powers of x or u whose particular values are unimportant (note that it suffices to bound the relative error by $u^{\gamma_k} e^{-2u^\beta}$) and by $c_1, c_2, \ldots$ denote constants.

Lemma A.1. *Let* $(x, y) \in k' \cup k''$ *for some* $k = 1, \ldots, n - 1$. *Then the conditional density,* $g(\cdot \mid x)$, *of* S_n *given* $M_n = Y_n = x$ *satisfies*

$$g(y \mid x) \leq c_1 x^{\gamma_1} \exp\{-(k-1)x^\beta - (y - kx)^\beta\}, \qquad y \geq x.$$

Proof. We have $f(x) \leq c_2(1 + x)^\gamma e^{-x^\beta} \psi(x)$, where

$$c_3 = \sup_{y \geq a} \psi(y) < \infty, \quad \psi(y) \leq c_4 f(y), \qquad y \leq a, \tag{A.1}$$

for some $a > 0$. Let S be the compact simplex in $\mathbb{R}^{n-1}$ specified by the constraints

$$0 \leq x_1 \leq x, \quad \ldots, \quad 0 \leq x_{n-1} \leq x, \qquad x_1 + \cdots + x_{n-1} = y - x,$$

and let μ be the Lebesgue measure on S. We then have

$$\begin{aligned}
g(y \mid x) &= \int_S f(x_1) \cdots f(x_{n-1}) \, d\mu(x_1, \ldots, x_{n-1}) \\
&\leq [c_2(1 + x)^{\gamma_2}]^{n-1} \sup_S \exp\{-x_1^\beta - \cdots - x_{n-1}^\beta\} \\
&\quad \times \int_S \psi(x_1) \cdots \psi(x_{n-1}) \, d\mu(x_1, \ldots, x_{n-1}) \\
&\leq c_5(1 + x)^{\gamma_1} \sup_S \exp\{-x_1^\beta - \cdots - x_{n-1}^\beta\},
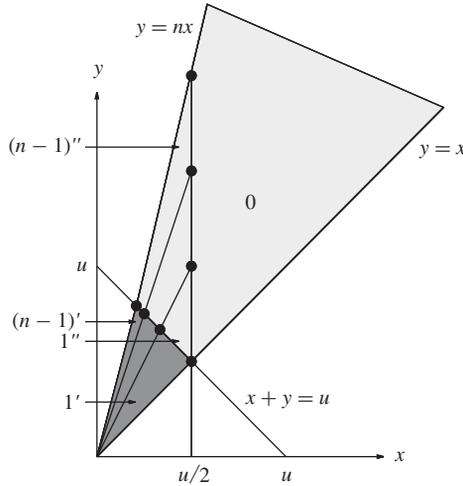\end{aligned}$$

FIGURE 1: Division of the support of the distribution of (M_n, S_m) into $2n - 1$ regions.

where in the last step we have used (A.1) to bound the μ-integral by $c_4^k[c_3(x - a)]^{n-k-1}$ over a region of the form

$$S \cap \{x_1 \leq a, \ldots, x_k \leq a, x_{k+1} > a, \ldots, X_{n-1} > a\}.$$

Being concave, $x_1^\beta + \cdots + x_{n-1}^\beta$ attains its minimum on S at an extremal point, which is easily seen to imply that at a minimum point $k-1$ of the x_i must equal x, one of the x_i must equal $y - kx$, and the rest must equal 0. Thus, the supremum can be bounded by $\exp\{-(k-1)x^\beta - (y-kx)^\beta\}$.

Proof of Theorem 3.1. Let $h(x, y)$ denote the joint density of M_n and S_n. The second moment of the estimator is then

$$\iint k(x, y)\, dx\, dy, \qquad k(x, y) = \exp\{-2(x \vee (u - y))^\beta\}h(x, y).$$

Consider the partition, shown in Figure 1, of the support of h (the shaded area) into the regions $0, 1', \ldots, (n - 1)', 1'', \ldots, (n - 1)''$. We will prove the theorem by showing that the contributions, $I_0, I_{1'}, \ldots$, to the above integral from the separate regions grow no faster than $c_m u^{\gamma_m} e^{-2u^\beta}$. This is simple for region 0, and for the other regions we will use ideas based upon concavity and extremal points (marked by bullet points in Figure 1) similar to those in the proof of Lemma A.1; note that the areas grow at rate at most u^2.

In region 0, $x > u - y$ and $x > u/2$. Thus, recalling the well-known fact that the density of M_n is $nf(x)F(x)^{n-1} \leq nf(x)$, we obtain

$$I_0 \leq e^{-2(u/2)^\beta} P(M_n \geq u/2) \leq c_6 u^{\gamma_3} e^{-3(u/2)^\beta} \leq c_6 u^{\gamma_3} e^{-2u^\beta}.$$

In region k', $u - y \geq x$ and, by Lemma A.1,

$$h(x, y) \leq c_7 x^{\gamma_4} \exp\{-kx^\beta - (y - kx)^\beta\}.$$

Thus (note that $x \leq u$ when $x \in k'$),

$$I_{k'} \leq c_7 u^{\gamma_4} \iint_{k'} \exp\{-kx^\beta - (y - kx)^\beta - 2(u - y)^\beta\}\, dx\, dy.$$

Since the area of region k' grows like u^2, it suffices to check that minus the value of the exponent is at least $2u^\beta$ at each of the three boundary points. This is clear at $(0, 0)$. The two other boundary points (on the line $x+y=u$) are $(u/(k+1), ku/(k+1))$ and $(u/(k+2), (k+1)u/(k+2))$, where minus the values of the exponents are respectively $(k+2)u^\beta/(k+1)^\beta$ and $(k+3)u^\beta/(k+2)^\beta$, which are both greater than $2u^\beta$ (the function $(x+2)/(x+1)^\beta$ is $3/2^\beta > 2$ at $x = 1$ and increasing, as is easily seen by checking that the log derivative is positive for $\beta < \bar\beta$ and $x \geq 1$).

In region k'', $u - y \leq x$ and, so, as in the argument for region k',

$$I_{k''} \leq c_8 u^{\gamma_5} \iint_{k''} nx^n \exp\{-(k+2)x^\beta - (y - kx)^\beta\} \, \mathrm{d}x \, \mathrm{d}y.$$

We must again check that minus the value of the exponent is at least $2u^\beta$ at each of the four boundary points. The two boundary points on the line $u = x + y$ and, thus, the values to be checked, are the same as for region k', so these behave as they should. The two boundary points on the line $x = u/2$ are $(u/2, (k+1)u/2)$ and $(u/2, ku/2)$, and at each of these points minus the value of the exponent is at least $(k+2)x^\beta = (k+2)u^\beta/2^\beta \geq 2u^\beta$.

References

[1] ABATE, J., CHOUDHURY, G. L. AND WHITT, W. (1994). Waiting-time tail probabilities in queues with long-tailed service-time distributions. *Queueing Systems* **16,** 311–338.

[2] ADLER, R. J., FELDMAN, R. AND TAQQU, M. S. (eds) (1998). *A Practical Guide to Heavy Tails*. Birkhäuser, Boston, MA.

[3] ASMUSSEN, S. (2000). *Ruin Probabilities*. World Scientific, River Edge, NJ.

[4] ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd edn. Springer, New York.

[5] ASMUSSEN, S. AND BINSWANGER, K. (1997). Simulation of ruin probabilities for subexponential claims. *ASTIN Bull.* **27,** 297–318.

[6] ASMUSSEN, S. AND RUBINSTEIN, R. Y. (1995). Steady state rare events simulation in queueing models and its complexity properties. In *Advances in Queueing. Theory, Methods, and Open Problems*, ed. J. H. Dshalalow, CRC Press, Boca Raton, FL, pp. 429–466.

[7] ASMUSSEN, S., BINSWANGER, K. AND HØJGAARD, B. (2000). Rare events simulation for heavy-tailed distributions. *Bernoulli* **6,** 303–322.

[8] ASMUSSEN, S., KLÜPPELBERG, C. AND SIGMAN, K. (1999). Sampling at subexponential times, with queueing applications. *Stoch. Process. Appl.* **79,** 265–286.

[9] ASMUSSEN, S., KROESE, D. P. AND RUBINSTEIN, R. Y. (2005). Heavy tails, importance sampling and cross-entropy. *Stoch. Models* **21,** 57–76.

[10] BALTRUNAS, A., DALEY, D. J. AND KLÜPPELBERG, C. (2004). Tail behaviour of the busy period of a GI/G/1 queue with subexponential service times. *Stoch. Process. Appl.* **111,** 237–258.

[11] CRUZ, M. G. (2002). *Modeling, Measuring and Hedging Operational Risk*. John Wiley, Chichester.

[12] EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

[13] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edn. John Wiley, New York.

[14] FRACHOT, A., MOUDOULAUD, O. AND RONCALLI, T. (2004). Loss distribution approach in practice. In *The Basel Handbook. A Guide for Financial Practitioners*, ed. M. K. Ong, Risk Books, London.

[15] GLASSERMAN, P. (2004). *Monte Carlo Methods in Financial Engineering*. Springer, New York.

[16] HEIDELBERGER, P. (1995). Fast simulation of rare events in queueing and reliability models. *ACM Trans. Model. Comput. Simul.* **5,** 43–85.

[17] JUNEJA, S. AND SHAHABUDDIN, P. (2002). Simulating heavy tailed processes using delayed hazard rate twisting. *ACM Trans. Model. Comput. Simul.* **12,** 94–118.

[18] MIKOSCH, T. AND NAGAEV, A. V. (1998). Large deviations of heavy-tailed sums with applications in insurance. *Extremes* **1,** 81–110.

[19] RUBINSTEIN, R. Y. AND KROESE, D. P. (2004). *The Cross-Entropy Method. A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. Springer, New York.

[20] SIGMAN, K. (1999). Appendix: a primer on heavy-tailed distributions. *Queueing Systems* **33,** 261–275.