THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

Error rates and improved algorithms for rare event simulation with heavy Weibull tails

Søren Asmussen and Dominik Kortschak

13

Research Report

No. 02 I January 2013

Error rates and improved algorithms for rare event simulation with heavy Weibull tails*

Søren Asmussen[1] and Dominik Kortschak[2]

[1]Department of Mathematic, Aarhus University, asmus@imf.au.dk
[2]Institut de Science Financière et d'Assurances, Université Lyon 1, kortschakdominik@gmail.com

Abstract

Let $Y_1, \ldots, Y_n$ be i.i.d. subexponential and $S_n = Y_1 + \cdots + Y_n$. Asmussen and Kroese (2006) suggested a simulation estimator for evaluating $\mathbb{P}(S_n > x)$, combining an exchangeability argument with conditional Monte Carlo. The estimator was later shown by Hartinger & Kortschak (2009) to have vanishing relative error. For the Weibull and related cases, we calculate the exact error rate and suggest improved estimators. These improvements can be seen as control variate estimators, but are rather motivated by second order subexponential theory which is also at the core of the technical proofs.

Keywords: Complexity, conditional Monte Carlo, control variates, lognormal distribution, M/G/1 queue, Pollaczeck-Khinchine formula, rare event, regular variation, ruin theory, second order subexponentiality, subexponential distribution, vanishing relative error, Weibull distribution

1 Introduction

This paper is concerned with the efficient simulation of

$$z = z(x) = \mathbb{P}(S_n > x)\,,$$

where $Y_1, \ldots, Y_n$ are i.i.d. with a common subexponential distribution, $S_n = Y_1 + \cdots + Y_n$ and x is large so that z is small. By definition of subexponentiality (e.g., [18], [3, X.1], or [19]), we have $z \sim n\overline{F}(x)$ as $n \to \infty$ where $\overline{F}(x) = 1 - F(x)$ is the tail. Our main set-up is that F is heavy-tailed Weibull with tail $\overline{F}(x) = \mathrm{e}^{-x^\beta}$ with

$$0 < \beta < \beta_0 = \log(3/2)/\log(2) \approx 0.585 \tag{1.1}$$

*The second author was supported by the the MIRACCLE-GICC project and the Chaire d'excellence "Generali – Actuariat responsable: gestion des risques naturels et changements climatiques."

(the Weibull distribution with $\beta_0 \le \beta < 1$ is also heavy-tailed, but (1.1) is in fact essential for our results as well as for our key references [8], [21]). We chose the Weibull distribution since it is the prototype of a distribution with sublinear hazard rate and also keeps the expressions in the proofs simple. Nevertheless, we will give in Section 6 the main ideas that are needed to extend the results to a more general class of distributions that also includes the lognormal distribution as well as distributions close to but more general than the Weibull.

The general subexponential problem has a long history. As is traditional in the literature ([6]), we denote by a simulation estimator a r.v. $Z = Z(x)$ that can be generated by simulation and is unbiased, $\mathbb{E}Z = z$. The usual performance measure is the relative error $e(x) = \mathbb{V}\mathrm{ar}^{1/2}(Z)/z$. The relative error is bounded if $\limsup_{x\to\infty} e(x) < \infty$, and the estimator Z is logarithmically efficient if

$$\limsup_{x\to\infty} e(x)/z(x)^{\epsilon} < \infty \qquad \text{for all } \epsilon > 0.$$

Efficient estimators have long been known with light tails (see e.g. [6, VI.2], [15], [22] and [23] for surveys), and are typically based on ideas from large deviations theory implemented via exponential change of measure. The heavy-tailed case is more recent. In [5], some of the difficulties in a literal translation of the light-tailed ideas are explained. However, in the regularly varying case [4] gave the first logarithmically efficient estimator for $\mathbb{P}(S_n > x)$ using a conditional Monte Carlo idea. The idea was further improved in [8], which as of today stands as a model of an efficient and at the same time easily implementable algorithm. It is also at the core of this paper. The idea is to combine an exchangeability argument with the conditional Monte Carlo idea. More precisely (for convenience assuming existence of densities to exclude multiple maxima) one has

$$z = n\,\mathbb{P}(S_n > x,\, M_n = Y_n)$$

where $M_k = \max_{i\le k} Y_i$. An unbiased simulation estimator of z based on simulated values $Y_1, \ldots, Y_n$ is therefore the conditional expectation

$$Z_{\mathrm{AK}} = n\,\overline{F}\big(M_{n-1} \vee (x - S_{n-1})\big)$$

of this expression given $Y_1, \ldots, Y_{n-1}$, where $S_{n-1} = Y_1 + \cdots + Y_{n-1}$. In the Weibull case, this estimator (baptized the Asmussen-Kroese estimator by the simulation community) is shown in [8] to be logarithmically efficient when $\beta < \beta_0$ and in [21] to have vanishing relative error ($e(x) \to 0$), though the argument for this is rather implicit and no quantitative rates are given. A survey of the area (that also includes some importance sampling algorithms) is in [6, VI.3].

The contribution of this paper is two-fold: to compute the exact error rate of Z_{AK}; and to produce different estimators with better rates. Both aspects combine with ideas of higher order subexponential methodology (cf. Remark 4). A companion paper [7] gives similar results for the regularly varying case, though it should be remarked that the analysis is rather different and in fact easier than in the Weibull case.

Our main results are:

Theorem 1. *If $0 < \beta < \beta_0$, then the Asmussen-Kroese estimator's variance is asymptotically given by*

$$\mathbb{V}\mathrm{ar}(Z_{\mathrm{AK}}) \sim n^2\, \mathbb{V}\mathrm{ar}(S_{n-1}) f(x)^2$$

This Theorem is just a special case of the following more general result.

Theorem 2. *Denote with $f^{(k)}$ the k-th derivative of the density f. Define the estimator*

$$Z_m = Z_{\mathrm{AK}} + n \sum_{k=1}^{m} \frac{(-1)^{k-1}}{k!} \left(\mathbb{E}S_{n-1}^k - S_{n-1}^k\right) f^{(k-1)}(x). \tag{1.2}$$

If $0 < \beta < \beta_0$, then the estimator Z_m in (1.2) *has vanishing relative error. More precisely,*

$$\mathbb{V}\mathrm{ar}(Z_m(x)) \sim \frac{n^2}{(m+1)!^2}\, \mathbb{V}\mathrm{ar}((S_{n-1})^{m+1}) f^{(m)}(x)^2.$$

Remark 3. The rates for the variances in Theorems 1 and 2 have to be compared with the rate $\mathrm{e}^{-2\beta}$ for the bounded relative error case. Note that $f(x) = \beta x^{\beta-1}\mathrm{e}^{-x^\beta}$ and $f^{(k)}(x) = (-1)^k p_k(x)\overline{F}(x)$ where p_k is regularly varying with index $(k{+}1)(\beta{-}1)$. Thus Z_{AK} improves the bounded relative error rate by a factor of $x^{1-\beta}$ and (1.2) by a factor of $x^{(k+1)(1-\beta)}$

The feature of vanishing relative error is quite unusual. The few further examples we know of are [14] and [17] in the setting of dynamic importance sampling, though it should be remarked that the algorithms there are much more complicated than those of this paper and [7], and that the rate results in [14], [17] are not very explicit.

Remark 4. A main idea of higher order subexponential methodology is the Taylor expansion

$$\overline{F}(x - S_{n-1}) = \overline{F}(u) + f(x)S_{n-1} - \tfrac{1}{2}f'(x)S_{n-1}^2 + \cdots \tag{1.3}$$

which leads to the refinement

$$z(x) = \mathbb{P}(S_n > x) = n\overline{F}(x) + nf(x)\mathbb{E}S_{n-1} - \tfrac{1}{2}f'(x)\mathbb{E}S_{n-1}^2 + \cdots ,$$

cf. [24], [11], [10] and [9]. Technically, the Taylor expansion is only useful for moderate S_{n-1}, and large values have to be shown to be negligible by a separate argument; this also is the case in the present paper. One may note that (1.3) is only useful for heavy-tailed distributions where typically $\overline{F}(x) \gg f(x) \gg f'(x) \gg \cdots$ – for light-tailed distributions like the exponential typically $\overline{F}(x), f(x), f'(x), \ldots$ have the same magnitude.

Remark 5. Main applications of the problem under study occur in ruin theory and the M/G/1 queue. These cases are connected by $\psi(x) = \mathbb{P}(W > x)$ where $\psi(x)$ is the ruin probability in a Cramér-Lundberg risk process and W is the steady-state waiting time of the queue. These quantities are in turn given by the Pollaczeck-Khinchine formula, where the number n of terms in S_n is an independent geometric r.v. N and the Y_i have the integrated tail distribution of the claim size/service time distribution, which is again subexponential. By means of dominated convergence our theory can be refined to this case (see Section 4).

A further application is credit risk, where N is the number of defaults and $Y_1, Y_2, \dots$ their sizes. Here the treatise Basel II calls for $\mathbb{P}(S_N > x)$ to be of magnitude $e-2$ to $3e-4$, which is also the relevant order for ruin theory. In queueing, $\mathbb{P}(W > x)$ could go all the way down to $e-12$, for example when studying bit loss rates in data transmission.

Remark 6. The main properties of the Weibull distribution $\overline{F}(x) = \mathrm{e}^{-x^\beta}$ that are used in the proofs are that the Weibull distribution is subexponential, has moments of all orders, that the density is infinitely often differentiable and that the hazard rate behaves like a power tail. Hence the results can be broadened, say to $\overline{F}(x) = c_1 x^\gamma \mathrm{e}^{-c_2 x^\beta}$ or the lognormal distribution, see Section 6 for more details.

2 First proofs

In this section we will prove Theorem 2. Since we want to extend the results to a random N, we will provide the constants as functions of n which is not needed if we are only interested in a fixed n.
Define $\widehat{V} = I\left(S_{n-1} \le x/2\right)$,

$$\begin{aligned} V_1 &= \widehat{V}\Big(\overline{F}(x - S_{n-1}) - \overline{F}(x) - \sum_{k=1}^{m} \frac{(-1)^{k+1}}{k!} S_{n-1}^k f^{(k-1)}(x)\Big)\,, \\ V_2 &= (1-\widehat{V})\overline{F}\big(M_{n-1} \vee (x - S_{n-1})\big)\,, \quad V_3 = -(1-\widehat{V})\overline{F}(x)\,, \\ V_{3+k} &= \frac{(-1)^k}{k!}(1-\hat{V})(S_{n-1})^k f^{(k-1)}(x), \quad k \ge 1\,. \end{aligned}$$

Then the estimator in (1.2) satisfies

$$Z_m = n\Big(\sum_{k=1}^{m+3} V_k\Big) + n\Big(\overline{F}(x) + \sum_{k=1}^{m} \frac{(-1)^{k-1}}{k!} \mathbb{E} S_{n-1}^k f^{(k-1)}(x)\Big). \tag{2.1}$$

Note that the second summand in (2.1) is constant. In the proofs, we will need two lemmas that are proved in Section 3:

Lemma 7.

$$\frac{\overline{F}\big(M_{n-1} \vee (x - S_{n-1})\big)}{\overline{F}(x)} \le \frac{\overline{F}(M_{n-1})}{\overline{F}(M_{n-1} + S_{n-1})}\,.$$

Lemma 8. *If $\beta < \beta_0$ then for all $k > 0$, $\ell \in \{1, 2\}$, $\gamma > 0$ and $\epsilon > 0$ there exist a C such that for all $n \ge 0$.*

$$\mathbb{E}\Big[M_n^k \Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n + S_n)}\Big)^\ell\Big] < C(1+\epsilon)^n\,, \tag{2.2}$$

$$\mathbb{E}\Big[M_n^k \Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n + S_n)}\Big)^\ell;\, S_n > x/2\Big] \le C(1+\epsilon)^n x^{-\gamma}. \tag{2.3}$$

Remark 9. In the following proofs we will sometimes consider bounds similar to

$$\overline{F}(x)^2\mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^{\ell};\, S_n > x/2\Big] \leq C(1+\epsilon)^n x^{-\gamma}\overline{F}(x)^2 = \mathrm{o}(f^{(m)}(x)^2).$$

Since $\overline{F}(x)^2 f^{(k)}(x)^2 \sim x^{2(m+1)(\beta-1)}$ we have to choose $\gamma > 2(m+1)(\beta-1)$ for the above inequality to be true.

Proof of Theorem 2. Since

$$\mathbb{V}\mathrm{ar}(Z_m) = n^2\Big(\sum_{i=1}^{m+3}\mathbb{V}\mathrm{ar}(V_i) + \sum_{i,j=1,\, i\neq j}^{m+3}\mathbb{C}\mathrm{ov}(V_i,V_j)\Big)$$

and $\big|\mathbb{C}\mathrm{ov}(V_i,V_j)\big| \leq \sqrt{\mathbb{V}\mathrm{ar}(V_i)\,\mathbb{V}\mathrm{ar}(V_j)}$ it is enough to show that $\mathbb{V}\mathrm{ar}(V_i) = \mathrm{o}(f^{(m)}(x)^2)$ for $i > 1$ and

$$\mathbb{V}\mathrm{ar}(V_1) \sim \frac{1}{(m+1)!^2}\,\mathbb{V}\mathrm{ar}(S_{n-1}^{m+1})f(m)(x)^2.$$

V_1*:* A Taylor expansion leads to

$$V_1(x) = (-1)^m\hat{V}(x)\frac{(S_{n-1})^{m+1}}{(m+1)!}f^{(m)}(x-\xi_{S_{n-1}})$$

with $0 \leq \xi_{S_{n-1}} \leq S_{n-1}$. Since $f^{(m)}(x)$ is long tailed (i.e., $f^{(m)}(x)/f^{(m)}(x+y) \to 1$ for all y), it follows that for fixed S_{n-1}

$$\lim_{x\to\infty}\frac{V_1(x)}{(-1)^m f^{(m)}(x)} = \frac{(S_{n-1})^{m+1}}{(m+1)!}.$$

Remember that $f^{(m)}(x) = (-1)^m p_m(x)\overline{F}(x)$ with $p_m(x)$ is regularly varying and $C_m = \sup_{x\geq 0}\sup_{x/2\leq y\leq x} p(y)/p(x) < \infty$. In the following, we will use that when $\widehat{V} \neq 0$ then $S_{n-1} \leq x/2$ and hence $M_{n-1} \leq x - S_{n-1}$, so that by Lemma 7

$$\begin{aligned}\frac{V_1}{(-1)^m f^{(m)}(x)} &= \frac{p_m(x-\xi_{S_{n-1}})}{p_m(x)}\widehat{V}\frac{(S_{n-1})^{m+1}}{(m+1)!}\frac{\overline{F}(x-\xi_{S_{n-1}})}{\overline{F}(x)}\\ &\leq C_m\widehat{V}\frac{(S_{n-1})^{m+1}}{(m+1)!}\frac{\overline{F}(x-S_{n-1})}{\overline{F}(x)}\\ &= C_m\widehat{V}\frac{(S_{n-1})^{m+1}}{(m+1)!}\frac{\overline{F}(M_{n-1}\vee(x-S_{n-1}))}{\overline{F}(x)}\\ &\leq C_m\frac{(S_{n-1})^{m+1}}{(m+1)!}\frac{\overline{F}(M_{n-1})}{\overline{F}(M_{n-1}+S_{n-1})}\end{aligned}$$

Since $(n-1)M_{n-1} > S_{n-1}$, we get for $k \in \{1,2\}$ that

$$\mathbb{E}\left|\frac{V_1(x)}{(-1)^m f^{(m)}(x)}\right|^k \leq \left(\frac{C_m(n-1)^{(m+1)k}}{(m+1)!}\right)^k\mathbb{E}\Big[M_{n-1}^{(m+1)k}\left(\frac{\overline{F}(M_{n-1})}{\overline{F}(M_{n-1}+S_{n-1})}\right)^k\Big]$$

which is finite by Lemma 8. Thus by dominated convergence,

$$\lim_{x\to\infty}\frac{\mathbb{V}\mathrm{ar}(V_1(x))}{f^{(m)}(x)^2}=\lim_{x\to\infty}\mathbb{E}\Big(\frac{V_1(x)^2}{f^{(m)}(x)^2}\Big)-\Big(\mathbb{E}\frac{V_1(x)}{f^{(m)}(x)}\Big)^2$$
$$=\frac{1}{(m+1)!^2}\left[\mathbb{E}((S_{n-1})^{2(m+1)})-\mathbb{E}((S_{n-1})^{m+1})^2\right].$$

V_2: from Lemmas 7, 8, (choose $\gamma>2(m+1)(\beta-1)$ in Lemma 8) we get

$$\mathbb{V}\mathrm{ar}(V_2)\le\mathbb{E}(V_2^2)=\mathbb{E}\big[\overline{F}\big(M_{n-1}\vee(x-S_{n-1})\big)^2;\,S_{n-1}>x/2\big]$$
$$\le\overline{F}(x)^2\mathbb{E}\Big[\Big(\frac{\overline{F}(M_{n-1})}{\overline{F}(M_{n-1}+S_{n-1})}\Big)^2;\,S_{n-1}>x/2\Big]=\mathrm{o}(f^{(m)}(x)^2).$$

V_3: Since $\hat V$ is a Bernoulli random variable and $\overline{F}(x)$ is constant we get

$$\mathbb{V}\mathrm{ar}(V_3)=\mathbb{P}(S_{n-1}>x/2)\mathbb{P}(S_{n-1}\le x/2)\overline{F}(x)^2=\mathrm{o}(f^{(m)}(x)^2).$$

We used $\mathbb{P}(S_{n-1}>x/2)\le K(1+\epsilon)^n\overline{F}(x/2)=o(x^{-\gamma})\ \forall\gamma>0$.
V_{3+k} $k\ge1$: We get

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}(V_{3+k})\\
&=\frac{f^{(k-1)}(x)^2}{k!^2}\,\mathbb{V}\mathrm{ar}\big((1-\widehat{V})(S_{n-1})^k\big)\le\frac{f^{(k-1)}(x)^2}{k!^2}\mathbb{E}\big((1-\widehat{V})S_{n-1}^k\big)^2\\
&=\frac{f^{(k-1)}(x)^2}{k!^2}\int_{x/2}^{\infty}y^{2k}\,\mathbb{P}(S_{n-1}\in\mathrm{d}y)\\
&=\frac{f^{(k-1)}(x)^2}{k!^2}\left(2k\int_{x/2}^{\infty}y^{2k-1}\mathbb{P}(S_{n-1}>y)\,\mathrm{d}y-(x/2)^{2k}\mathbb{P}(S_{n-1}>x/2)\right)\\
&\le K(1+\epsilon)^{n-1}\frac{f^{(k-1)}(x)^2}{k!^2}\left(2k\int_{x/2}^{\infty}y^{2k-1}\mathbb{P}(Y_1>y)\,\mathrm{d}y+(x/2)^{2k}\mathbb{P}(Y_1>x/2)\right)\\
&=\mathrm{o}(f^{(m)}(x)^2).
\end{aligned}$$
□

The estimator Z_1 in (1.2) has the form $Z_{\mathrm{AK}}+\alpha(S_{n-1}-\mathbb{E}S_{n-1})$, so it is a control variate estimator, using S_{n-1} as control for Z_{AK} (for $m\ge1$ Z_m can be interpreted as an estimator with multiple controls). It is natural to ask whether the $\alpha=-nf(x)$ at least asymptotically coincides with the optimal $\alpha^*=-\mathbb{C}\mathrm{ov}(Z_{\mathrm{AK}},S_{n-1})/\mathbb{V}\mathrm{ar}(S_{n-1})$ (cf. [6, V.2]). The following lemma shows that this is the case and further provides some more detailed expansions of $\mathbb{V}\mathrm{ar}(Z_{\mathrm{AK}})$, cf. Proposition 11 below. We get for the estimator $Z^*=Z_{\mathrm{AK}}+\alpha^*(S_{n-1}-\mathbb{E}S_{n-1})$ that

$$\mathbb{V}\mathrm{ar}(Z^*)\sim\mathbb{V}\mathrm{ar}(Z_{\mathrm{AK}})\left(1-\rho(S_{n-1}^2,S_{n-1}))^2\right),$$

where $\rho(S_{n-1}^2,S_{n-1})$ denotes the correlation between S_{n-1} and S_{n-1}^2. So Z_1 can be improved by the use of the optimal rate α^*.

Lemma 10.

$$\mathbb{C}\mathrm{ov}(Z_{\mathrm{AK}},S_{n-1})=n\,\mathbb{V}\mathrm{ar}(S_{n-1})f(x)-\tfrac{n}{2}\left(\mathbb{E}S_{n-1}^3-\mathbb{E}S_{n-1}\mathbb{E}S_{n-1}^2)\right)f'(x)$$
$$+\tfrac{n}{6}\left(\mathbb{E}S_{n-1}^4-\mathbb{E}S_{n-1}\mathbb{E}S_{n-1}^3)\right)f''(x)+\mathrm{o}\big(f''(x)\big).$$

Proof. Since

$$\mathbb{E}(Z_{\mathrm{AK}}S_{n-1}) = \mathbb{E}(Z_{\mathrm{AK}}S_{n-1}I\left(S_{n-1} > x/2\right)) + \mathbb{E}(Z_{Ak}S_{n-1}I\left(S_{n-1} \le x/2\right))$$

Now as in the proof for V_2 we get

$$\mathbb{E}(Z_{\mathrm{AK}}S_{n-1};\, S_{n-1} > x/2) = \mathrm{o}\big(x^{-k}\overline{F}(x)\big).$$

Further we get with a Taylor expansion that for some $0 \le \xi_y \le y$

$$\begin{aligned}
\tfrac{1}{n}\mathbb{E}(Z_{Ak}S_{n-1};\, S_{n-1} \le x/2) &= \int_0^{x/2} y\overline{F}(x-y)\mathrm{d}F_{S_{n-1}}(y) \\
= \int_0^{x/2} y\overline{F}(x)\mathrm{d}S_{n-1} &+ \int_0^{x/2} y^2 f(x)\mathrm{d}F_{S_{n-1}}(y) - \frac{1}{2}\int_0^{x/2} y^3 f'(x)\mathrm{d}F_{S_{n-1}}(y) \\
&+ \frac{1}{6}\int_0^{x/2} y^4 f''(x-\xi_y)\mathrm{d}F_{S_{n-1}}(y).
\end{aligned}$$

Since $f''(x)$ is monotonely decreasing we get that for every fixed y

$$1 \le \lim_{x\to\infty}\frac{f''(x-\xi_y)}{f''(x)} \le \lim_{x\to\infty}\frac{f''(x-y)}{f''(x)} = 1.$$

Denote with $c = \sup_{x>0}\frac{f''(x/2)}{\overline{F}(x/2)}\frac{\overline{F}(x)}{f''(x)} < \infty$. As in the proof for V_1, we get using Lemma 7 that

$$\begin{aligned}
S_{n-1}^4\frac{f''(x-\xi_{S_{n-1}})}{f''(x)}I\left(y < x/2\right) &\le S_{n-1}^4\frac{f''(x-S_{n-1})}{f''(y)}I\left(S_{n-1} < x/2\right) \\
\le S_{n-1}^4\frac{f''(x-S_{n-1})}{f''(S_{n-1})}I\left(S_{n-1} < x/2\right) &\le c_1 S_{n-1}^4\frac{\overline{F}(x-S_{n-1})}{\overline{F}(x)}I\left(S_{n-1} < x/2\right) \\
\le c_1 S_{n-1}^4\frac{\overline{F}(M_{n-1}\vee x - S_{n-1})}{\overline{F}(x)} &\le c_1 S_{n-1}^4\frac{\overline{F}(M_{n-1})}{\overline{F}(M_{n-1}+S_{n-1})}.
\end{aligned}$$

The last random variable is integrable by Lemma 8, hence we get by dominated convergence

$$\int_0^{x/2} y^4 f''(x-\xi_y)\mathrm{d}F_{S_{n-1}}(y) \sim \int_0^{x/2} y^4 f''(x)\mathrm{d}F_{S_{n-1}}(y).$$

Since for every $k > 0$

$$\int_{x/2}^{\infty} y^k\overline{F}^{(k-1)}(x)\mathrm{d}F_{S_{n-1}}(y) \sim (n-1)\overline{F}^{(k-1)}(x)\int_{x/2}^{\infty} y^k\mathrm{d}F(y) = o(f''(x))$$

it follows that

$$\begin{aligned}
&\tfrac{1}{n}\mathbb{E}(Z_{Ak}S_{n-1};\, S_{n-1} \le x/2) \\
&\quad = \mathbb{E}S_{n-1}\overline{F}(x) + \mathbb{E}S_{n-1}^2 f(x) - \tfrac{1}{2}\mathbb{E}S_{n-1}^3 f'(x) + \tfrac{1}{6}\mathbb{E}S_{n-1}^4 f''(x) + o(f''(x))\,.
\end{aligned}$$

Since (see [10])

$$\begin{aligned}
\mathbb{E}Z_{AK}\mathbb{E}S_{n-1} = \mathbb{E}S_{n-1}\mathbb{P}(S_n > u) &= n\mathbb{E}S_{n-1}\overline{F}(x) + n\left(\mathbb{E}S_{n-1}\right)^2 f(x) \\
&- \tfrac{n}{2}\mathbb{E}S_{n-1}\mathbb{E}S_{n-1}^2 f'(x) + \tfrac{n}{6}\mathbb{E}S_{n-1}\mathbb{E}S_{n-1}^3 f''(x) + \mathrm{o}(f''(x)).
\end{aligned}$$

and $\mathbb{C}\mathrm{ov}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$, the lemma follows. □

The following result gives more detailed expansions of the variance of the Asmussen-Kroese estimator than Theorem 1. We omit the proof.

Proposition 11. *The Asmussen-Kroese estimator has asymptotic variance*

$$\begin{aligned}\mathbb{V}\mathrm{ar}(Z_{\mathrm{AK}}) = n^2\, \mathbb{V}\mathrm{ar}(S_{n-1}) f(x)^2 - n^2 \left(\mathbb{E}S_{n-1}^3 - \mathbb{E}S_{n-1}\mathbb{E}S_{n-1}^2\right) f(x)f'(x) \\ + \tfrac{n^2}{3}\left(\mathbb{E}S_{n-1}^4 - \mathbb{E}S_{n-1}\mathbb{E}S_{n-1}^3\right) f(x)f''(x) \\ + \tfrac{n^2}{4}\left(\mathbb{E}S_{n-1}^4 - \left(\mathbb{E}S_{n-1}^2\right)^2\right) f'(x)^2 + \mathrm{o}(f'(x)^2).\end{aligned}$$

3 Further proofs

Proof of Lemma 7. This is essentially Lemma 4.2 of [21], but since the proof is short, we reproduce it here. The inequality is obvious if $x \le M_{n-1} + S_{n-1}$. Otherwise, let $z = x - M_{n-1} - S_{n-1}$. Since in the Weibull case, the failure rate $\lambda(x) = \beta/x^{1-\beta}$ is non-increasing for all $x > 0$, we have (recall that $\overline{F}(y) = \exp\{-\int_0^y \lambda(u)\,\mathrm{d}u\}$)

$$\begin{aligned}\log\frac{\overline{F}\big(M_{n-1}\vee(x-S_{n-1})\big)}{\overline{F}(x)} &= \log\frac{\overline{F}(x-S_{n-1})}{\overline{F}(x)} = \log\frac{\overline{F}(M_{n-1}+z)}{\overline{F}(z+M_{n-1}+S_{n-1})} \\ &= \int_{M_{n-1}+z}^{z+M_{n-1}+S_{n-1}} \lambda(u)\,\mathrm{d}u \le \int_{M_{n-1}}^{M_{n-1}+S_{n-1}} \lambda(u)\,\mathrm{d}u = \log\frac{\overline{F}(M_{n-1})}{\overline{F}(M_{n-1}+S_{n-1})}\,.\end{aligned}$$

□

Lemma 12. *Let $x > 0$, $c > 0$ and $\beta 2^\beta < 1$. Then*

$$\begin{aligned}&\int_0^x \beta y^{\beta-1}\exp\big\{2\,(2x+c+y)^\beta - y^\beta\big\}\,\mathrm{d}y \\ &\quad\le \exp\big\{2\,(2x+c)^\beta\big\}\Big[1-\exp\big\{-(1-\beta 2^\beta)x^\beta\big\} + \frac{2^\beta\Gamma(1/\beta)}{(1-\beta 2^\beta)^{1/\beta}}x^{\beta-1}\Big].\end{aligned}$$

Proof. By Taylor's theorem we get that for some $0 < \xi_y < y$

$$(2x+c+y)^\beta = (2x+c)^\beta + \beta y\,(2x+c+\xi_y)^{\beta-1} \le (2x+c)^\beta + \beta y\,(2x)^{\beta-1}\,.$$

Hence

$$\begin{aligned}&\int_0^x \beta y^{\beta-1}\exp\big\{2\,(2x+c+y)^\beta - y^\beta\big\}\,\mathrm{d}y \\ &\quad\le \exp\big\{2\,(2x+c)^\beta\big\}\int_0^x \beta y^{\beta-1}\exp\left\{\beta 2^\beta y x^{\beta-1} - y^\beta\right\}\,\mathrm{d}y.\end{aligned}$$

By partial integration and $x^{\beta-1} < y^{\beta-1}$,

$$\begin{aligned}\int_0^x \beta y^{\beta-1}\exp\left\{\beta 2^\beta y x^{\beta-1} - y^\beta\right\}\,\mathrm{d}y &= \int_0^x \beta y^{\beta-1}\mathrm{e}^{-y^\beta}\,\exp\big\{\beta 2^\beta y x^{\beta-1}\big\}\,\mathrm{d}y \\ &= -\exp\left\{\beta 2^\beta y x^{\beta-1} - y^\beta\right\}\Big|_0^x + \beta 2^\beta x^{\beta-1}\int_0^x \exp\left\{\beta 2^\beta y x^{\beta-1} - y^\beta\right\}\,\mathrm{d}y \\ &\le 1 - \exp\big\{-(1-\beta 2^\beta)x^\beta\big\} + \beta 2^\beta x^{\beta-1}\int_0^\infty \exp\big\{-(1-\beta 2^\beta)y^\beta\big\}\,\mathrm{d}y \\ &= \big(1-\exp\big\{-(1-\beta 2^\beta)x^\beta\big\}\big) + \frac{2^\beta\Gamma(1/\beta)}{(1-\beta 2^\beta)^{1/\beta}}x^{\beta-1}\,.\end{aligned}$$

$\square$

Proof of Lemma 8. At first note that it is enough to prove the Lemma with $(1+\epsilon)^n$ replaced by $n^\tau(1+\epsilon)^n$ where τ might dependent on ℓ, k. The reason for that is

$$\lim_{n\to\infty} \frac{n^\tau(1+\epsilon/2)^n}{(1+\epsilon)^n} = 0.$$

Since $\beta 2^\beta < \frac{3}{2}\log(3/2)/\log(2) < 1$ for $\beta < \log(3/2)/\log(2)$, we can choose x_0 such that for $x > x_0$

$$\big(1 - \exp\big\{\big(\beta 2^\beta - 1\big)\, x^\beta\big\}\big) + \frac{\beta 2^\beta \Gamma(1+1/\beta)}{(1-\beta 2^\beta)^{1/\beta}} x^{\beta-1} \le 1+\epsilon. \tag{3.1}$$

Since $\overline{F}(M_n)/\overline{F}(M_n+S_n) > 1$,

$$\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^\ell \le \Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^2.$$

Hence we only have to consider $\ell = 2$. First note that for every $\epsilon > 0$ there exists a C_1 with

$$\begin{aligned}\mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^2;\, M_n \le x_0\Big] &\le \frac{x_0^k}{\overline{F}\big((n+1)x_0\big)^2} \\ &= x_0^k\, \mathrm{e}^{2x_0^\beta(n+1)^\beta} \le C_1(1+\epsilon)^n\,.\end{aligned}$$

By the same exchangeability argument as for the Asmussen-Kroese estimator, we get that for every $x \ge 0$

$$\begin{aligned}&\mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^2; M_n > x\Big] \\ &\quad= n\,\mathbb{E}\Big[Y_n^k\Big(\frac{\overline{F}(Y_n)}{\overline{F}(Y_n+S_n)}\Big)^2;\, Y_n > x,\, M_n = Y_n\Big] \\ &\quad= n\,\mathbb{E}\Big[Y_n^k\Big(\frac{\overline{F}(Y_n)}{\overline{F}(2Y_n+S_{n-1})}\Big)^2;\, Y_n > x,\, M_n = Y_n\Big].\end{aligned}$$

If $x > x_0$ we get with an iterative application of Lemma 12 and (3.1) that

$$\begin{aligned}&\mathbb{E}\Big[Y_n^k\Big(\frac{\overline{F}(Y_n)}{\overline{F}(2Y_n+S_{n-1})}\Big)^2;\, M_n = Y_n,\, Y_n > x\Big] \\ &= \int_{y_n=x}^\infty \int_{[0,y_n]^{n-1}} y_n^k \beta^n \prod_{i=1}^n y_i^{\beta-1} \exp\Big\{-2y_n^\beta + 2\Big(2y_n + \sum_{i=1}^{n-1} y_i\Big)^\beta - \sum_{i=1}^n y_i^\beta\Big\}\, \mathrm{d}\mathbf{y} \\ &\le (1+\epsilon)^{n-1} \int_{y_n=x}^\infty \beta y_n^{k+\beta-1} \exp\big\{-2y_n^\beta + 2\,(2y_n)^\beta - y_n^\beta\big\}\, \mathrm{d}y_n \\ &= (1+\epsilon)^{n-1} \int_{y_n=x}^\infty \beta y_n^{k+\beta-1} \exp\big\{-(3-2^{1+\beta})y_n^\beta\big\}\, \mathrm{d}y_n\end{aligned}$$

where the last integral is uniformly bounded in x since $3-2^{1+\beta}>0$ by assumption, and (2.2) follows.

Using the same arguments, we get that for $x>2nx_0$

$$\begin{aligned}
\mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^2;\, S_n>x/2\Big] &\leq \\
&\leq \mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^2;\, M_n>x/(2n)\Big] \\
&\leq n(1+\epsilon)^{n-1}\int_{x/(2n)}^{\infty}\beta y^{k+\beta-1}\exp\big\{-(3-2^{1+\beta})y^\beta\big\}\,\mathrm{d}y \\
&= \frac{n(1+\epsilon)^{n-1}}{(3-2^{1+\beta})^{1+k/\beta}}\Gamma\Big(1+\frac{k}{\beta},(3-2^{1+\beta})\frac{x}{2n}\Big),
\end{aligned}$$

where $\Gamma(\alpha,z)=\int_z^\infty x^{\alpha-1}\mathrm{e}^{-x}\mathrm{d}x$ is the incomplete Gamma function. Since $x/2n>x_0$ and for every $\gamma>0$ there exists an C_2 with $\mathrm{e}^{-x}<C_2x^{-\gamma-k/\beta}$, we have for some $C_3>0$

$$\frac{n(1+\epsilon)^{n-1}}{(3-2^{1+\beta})^{k+1}}\Gamma\Big(1+k,(3-2^{1+\beta})\frac{x}{2n}\Big)\leq C_3n^{\gamma+1}(1+\epsilon)^{n-1}x^{-\gamma}.$$

So (2.3) holds if $x>2nx_0$. If $x\leq 2nx_0$, then by (2.2)

$$\mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^2;\, S_n>x/2\Big]\leq C(1+\epsilon)^n\leq C(2nx_0)^\gamma(1+\epsilon)^nx^{-\gamma}$$

and the Lemma follows. □

4 The case of a random $n=N$

In practice one is often interested in a random $n=N$. The easiest way to get an estimator for random N is to first simulate N and then use the estimator Z_m which leads to the estimator

$$\begin{aligned}
Z_{m,N}(x) = N\overline{F}(x-S_{N-1}\vee u-S_{N-1}) - N\overline{F}(x) - N\sum_{k=1}^{m}\frac{(-1)^{k-1}S_{N-1}^k}{k!}f^{(k-1)}(x) \\
+\mathbb{E}N\overline{F}(x)+\sum_{k=1}^{m}\frac{(-1)^{k-1}}{k!}\mathbb{E}\big[NS_{N-1}^k\big]f^{(k-1)}(x).
\end{aligned}$$

Since in the proof of Theorem 2 we can bound all terms by $C_\epsilon(1+\epsilon)^n$ for all $\epsilon>0$ and some corresponding constant C_ϵ, we get by dominated convergence that

Theorem 13. *Assume that $\beta<\beta_0$ and for some $\epsilon>0$ $\mathbb{E}(1+\epsilon)^N<\infty$. Then the estimator $Z_{m,N}$ satisfies*

$$\mathbb{V}\mathrm{ar}\ Z_{m,N}\sim\frac{1}{(m+1)!^2}\,\mathbb{V}\mathrm{ar}(NS_{N-1}^{m+1})f^{(m)}(x)^2$$

5 Numerical examples

In this section we will provide some numerical examples. As distribution for Y_i we will use either a lognormal distribution (cf. Section 6) with parameters $\mu = 0$ and $\sigma = 1$ or a Weibull distribution with parameter $\beta \in \{0.25, 0.5\}$. For N we will use either a Poisson distribution with parameter $\lambda = 10$, a geometric distribution with parameter $p = 1/11$ or we take $N = 10$ constant. For these 9 examples we choose x such that (for the second order asymptotic cf. [1])

$$\mathbb{E}N\overline{F}\left(x - \left(\frac{\mathbb{E}N^2}{\mathbb{E}N} - 1\right)\mathbb{E}Y\right) = 10^{-k}, \quad k = 1, \ldots, 7$$

holds. In the Tables we present x, $z = \mathbb{P}(S_N > x)$ and the relative error $\mathbb{V}\text{ar}[Z_{i,N}]/(\mathbb{E}Z_{i,N})^2$ (compare Section 4 for the definition of the estimators).

The picture is that the higher order estimators provide a substantial improvement of the Asmussen-Kroese estimator $Z_{0,N}$ for large x. This was of course to be expected from the asymptotic results. However, one also sees that when x is fixed the higher order estimators can have a quite poor performance. This was somehow expected since for fixed x one can easily show that $\lim_{i\to\infty} \mathbb{V}\text{ar}\ Z_{i,N} = \infty$

Remark 14. We also see that for lognormal and Weibull with $\beta = 0.5$ and N geometric the estimators have a poor performance. In this case also the asymptotics provides poor estimates. A possible conclusion is that the estimators are not working well when the asymptotic approximation is not good. In principle one can understand this phenomenon when one convinces oneself that the estimators as well as the asymptotic approximation are not working well when there is a "high" probability that S_{N-1} is "large" and as was pointed out in Ghamami & Ross [20]. this will usually be the case when N is large. Therefore [20] suggests a stratification estimator which uses different estimators depending on the size of N. We want to add the following observation to the discussion that might be useful to construct future estimators. If we assume that $\overline{F}(x)$ is holomorphic for $\Re(x) > 0$ and for a fixed n we define $Y_i^x = Y_i | Y_i \leq x/n$, then

$$\mathbb{P}(S_n > u | M_{n-1} \leq x/n) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \mathbb{E}(Y_1^x + \cdots Y_{n-1}^x)^k \overline{F}^{(k)}(x).$$

So using the estimators discussed in this paper an efficient estimation of $\mathbb{P}(S_n > u | M_{n-1} \leq x/n)$ is possible and one has to find efficient estimators for $\mathbb{P}(S_n > u | M_{n-1} > x/n)$ of course this method is not easily applied to random n. One should note that it is also true that

$$\mathbb{P}(S_n > u | S_{n-1} \leq x/2) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \mathbb{E}\left[(S_{n-1})^k | S_{n-1} \leq x/2\right] \overline{F}^{(k)}(x).$$

but here the difficulty lies in evaluating $\mathbb{E}\left[(S_{n-1})^k | S_{n-1} \leq x/2\right]$ efficiently.

x	z	$Z_{0,N}$	$Z_{1,N}$	$Z_{2,N}$	$Z_{3,N}$	$Z_{4,N}$
27	0.11	2.83	2.7	2.47	3.65	688
38	0.021	6.21	5.98	5.49	5.28	24.3
58	0.0018	10.1	9.34	8.76	8.22	23.2
88	0.00013	4.9	4.45	3.56	3.41	4.2
132	1.1×10^{-5}	1.07	0.818	0.553	0.449	0.99
198	1×10^{-6}	0.303	0.152	0.0837	0.0569	0.0639
290	1×10^{-7}	0.109	0.0345	0.0149	0.00826	0.0113
419	1×10^{-8}	0.0497	0.00922	0.00362	0.00437	0.00114
596	1×10^{-9}	0.0242	0.00191	0.000656	0.000412	0.000173
834	1×10^{-10}	0.0127	0.000595	0.000653	1.06×10^{-5}	6.88×10^{-6}
1152	1×10^{-11}	0.0071	0.000162	8.06×10^{-6}	5.91×10^{-6}	5.02×10^{-6}
1571	1×10^{-12}	0.00407	5.72×10^{-5}	4.58×10^{-6}	4.02×10^{-6}	3.92×10^{-6}

Table 1: Lognormal Y with Poisson N.

x	z	$Z_{0,N}$	$Z_{1,N}$	$Z_{2,N}$	$Z_{3,N}$	$Z_{4,N}$
25	0.097	0.987	0.943	0.926	3.55	248
37	0.015	1.74	1.62	1.43	2.01	94.2
56	0.0013	1.86	1.68	1.38	1.59	14.7
86	0.00012	0.946	0.771	0.576	0.816	4.15
131	1.1×10^{-5}	0.325	0.23	0.144	0.137	0.415
196	1×10^{-6}	0.113	0.0595	0.0327	0.0284	0.07
289	1×10^{-7}	0.0387	0.0152	0.00648	0.00668	0.00881
417	1×10^{-8}	0.0151	0.00394	0.00218	0.000783	0.000861
594	1×10^{-9}	0.00688	0.00112	0.000981	0.000248	0.000384
832	1×10^{-10}	0.00345	0.000521	4.67×10^{-5}	9.07×10^{-5}	9.95×10^{-7}
1150	1×10^{-11}	0.00189	5.62×10^{-5}	2.45×10^{-6}	5.28×10^{-7}	5.97×10^{-8}
1569	1×10^{-12}	0.00106	1.72×10^{-5}	6.25×10^{-7}	4.68×10^{-7}	3.05×10^{-9}

Table 2: Lognormal Y with constant $N = 10$.

x	z	$Z_{0,N}$	$Z_{1,N}$	$Z_{2,N}$	$Z_{3,N}$	$Z_{4,N}$
43	0.087	13.1	12.8	12.1	17.5	399
55	0.047	23.8	23.9	23	22.1	50.7
74	0.017	62	61.8	61.9	59.6	59.1
104	0.0036	265	270	267	264	262
149	0.00035	2530	2390	2270	2290	2440
214	1.4×10^{-5}	45600	45900	58700	43000	30000
307	3.7×10^{-7}	699000	2100000	33300	123000	64800
436	1.2×10^{-8}	32.6	427	7.34	15.5	68.1
612	1.1×10^{-9}	2.11	0.965	0.369	0.239	0.166
850	1×10^{-10}	0.72	0.198	0.0712	0.0123	0.00319
1168	1×10^{-11}	0.343	0.0463	0.00676	0.00138	0.00058
1587	1×10^{-12}	0.179	0.0153	0.00168	0.000612	0.000547

Table 3: Lognormal Y with geometric N.

x	z	$Z_{0,N}$	$Z_{1,N}$	$Z_{2,N}$	$Z_{3,N}$	$Z_{4,N}$
690	0.075	0.168	1.35	758	84300000	1.35×10^{11}
2517	0.0099	0.127	0.207	18.1	17200	2.64×10^{8}
7436	0.001	0.0703	0.0444	0.875	85.6	42800
17809	1×10^{-4}	0.0312	0.0128	0.071	2.13	82.6
36671	1×10^{-5}	0.0133	0.00511	0.0105	0.125	2.15
67732	1×10^{-6}	0.00511	0.0021	0.00164	0.00463	0.11
115379	1×10^{-7}	0.00242	0.000949	0.000418	0.00127	0.000356
184671	1×10^{-8}	0.000796	0.000372	0.000193	5.91×10^{-5}	9.57×10^{-5}
281341	1×10^{-9}	0.000397	3.19×10^{-5}	$3.9\ \times10^{-5}$	1.38×10^{-5}	5.53×10^{-6}
411800	1×10^{-10}	0.000163	4.95×10^{-5}	8.06×10^{-6}	4.79×10^{-6}	4.53×10^{-6}
583132	1×10^{-11}	8.82×10^{-5}	6.41×10^{-6}	6.61×10^{-6}	4.51×10^{-5}	5.47×10^{-6}
803093	1×10^{-12}	5.54×10^{-5}	4.84×10^{-6}	4.61×10^{-6}	4.53×10^{-6}	4.54×10^{-6}

Table 4: Weibull Y with $\beta = 0.25$ and Poisson N.

x	z	$Z_{0,N}$	$Z_{1,N}$	$Z_{2,N}$	$Z_{3,N}$	$Z_{4,N}$
666	0.077	0.11	1.08	936	8480000	3.08×10^{11}
2493	0.0099	0.084	0.159	13.1	187000	1.51×10^{8}
7412	0.001	0.0485	0.0318	1.44	164	33500
17785	1×10^{-4}	0.0225	0.00943	0.041	1.62	26.9
36647	1×10^{-5}	0.00944	0.00329	0.0053	0.0654	0.699
67708	1×10^{-6}	0.00396	0.0012	0.00151	0.00446	0.0886
115355	1×10^{-7}	0.00158	0.000506	0.00049	0.000729	0.000247
184647	1×10^{-8}	0.000598	0.000216	0.00015	4.62×10^{-5}	2.94×10^{-5}
281317	1×10^{-9}	0.000286	9.41×10^{-5}	5.77×10^{-5}	2.29×10^{-5}	7.02×10^{-6}
411776	1×10^{-10}	0.000116	4.94×10^{-6}	1.09×10^{-6}	4.67×10^{-7}	8.14×10^{-7}
583108	1×10^{-11}	5.82×10^{-5}	1.84×10^{-6}	1.63×10^{-5}	1.59×10^{-8}	$4.4\ \times10^{-8}$
803069	1×10^{-12}	3.95×10^{-5}	6.26×10^{-7}	4.72×10^{-7}	1.25×10^{-9}	2.59×10^{-10}

Table 5: Weibull Y with $\beta = 0.25$ with constant $N = 10$.

x	z	$Z_{0,N}$	$Z_{1,N}$	$Z_{2,N}$	$Z_{3,N}$	$Z_{4,N}$
930	0.06	1.35	4.57	1530	4270000	2.23×10^{11}
2757	0.01	1.48	1.33	59.6	26300	$2.1\ \times10^{7}$
7676	0.001	0.735	0.399	4.32	1900	197000
18049	1×10^{-4}	0.252	0.106	0.281	9.9	823
36911	1×10^{-5}	0.0937	0.0335	0.053	0.292	67.3
67972	1×10^{-6}	0.0356	0.0115	0.00808	0.0253	0.326
115619	1×10^{-7}	0.0145	0.00448	0.00323	0.00363	0.0394
184911	1×10^{-8}	0.00583	0.00257	0.00161	0.00292	0.0764
281581	1×10^{-9}	0.00264	0.000937	0.000836	0.000885	0.000748
412040	1×10^{-10}	0.00172	0.000798	0.000742	0.000758	0.000742
583372	1×10^{-11}	0.00124	0.000824	0.000738	0.000738	0.00114
803333	1×10^{-12}	0.00102	0.000743	0.00074	0.000739	0.000785

Table 6: Weibull Y with $\beta = 0.25$ and geometric N.

x	z	$Z_{0,N}$	$Z_{1,N}$	$Z_{2,N}$	$Z_{3,N}$	$Z_{4,N}$
41	0.089	1.49	1.22	2.03	22.8	369
68	0.015	3.21	2.75	2.28	5.16	65.1
105	0.0017	6.3	5.6	4.56	4.27	13.5
153	0.00015	9.33	8.28	7.4	6.02	7.41
211	1.4×10^{-5}	9.76	9.3	8.78	7.65	7.26
280	1.2×10^{-6}	6.43	11.6	6.53	6.36	4.42
359	1.2×10^{-7}	5.17	6.13	4.82	3.25	5.46
449	1.1×10^{-8}	3.18	2.41	4.24	2.36	1.77
550	1.1×10^{-9}	1.82	2.02	2.45	2.03	1.51
662	1.1×10^{-10}	2.83	1.61	0.954	1.21	0.648
783	1.1×10^{-11}	0.546	0.257	0.873	0.131	2.5
916	1×10^{-12}	0.853	0.188	0.152	0.0437	0.0972

Table 7: Weibull Y with $\beta = 0.5$ and Poisson N.

x	z	$Z_{0,N}$	$Z_{1,N}$	$Z_{2,N}$	$Z_{3,N}$	$Z_{4,N}$
39	0.088	0.712	0.619	1.42	15.5	413
66	0.013	1.48	1.22	1.08	3.55	45.1
103	0.0014	2.59	2.21	1.69	2.04	9.34
151	0.00014	3.5	3.04	2.42	2	3.66
209	1.2×10^{-5}	3.75	3.12	2.57	2.12	2.21
278	1.2×10^{-6}	3.23	3.08	2.45	2.17	1.57
357	1.1×10^{-7}	2.16	2.02	1.75	1.82	1.19
447	1.1×10^{-8}	1.73	1.42	1.43	1.5	0.972
548	1.1×10^{-9}	1.24	1.18	0.692	0.958	0.779
660	1.1×10^{-10}	1.3	0.753	0.492	0.537	0.263
781	1×10^{-11}	0.415	0.704	1.23	1.98	0.113
914	1×10^{-12}	0.406	0.145	0.125	0.0677	0.0341

Table 8: Weibull Y with $\beta = 0.5$ and constant $N = 10$.

x	z	$Z_{0,N}$	$Z_{1,N}$	$Z_{2,N}$	$Z_{3,N}$	$Z_{4,N}$
61	0.072	10	8.95	9.85	140	5280
88	0.027	23.8	22.6	20.3	27.6	321
125	0.007	75.8	75	72.3	67.4	83
173	0.0013	335	324	321	320	308
231	0.00016	1990	2060	2090	1850	2010
300	1.5×10^{-5}	21000	14400	17700	17100	13200
379	8.3×10^{-7}	92800	266000	43900	220000	81400
469	3.5×10^{-8}	26000	35800	234000	603000	5190000
570	2.3×10^{-9}	21500	13900	395000	316000	45000
682	1.7×10^{-10}	945	614000	6610	3350	1030
803	1.7×10^{-11}	98500	691	249	521	1110
936	1.4×10^{-12}	400	87.2	53.8	418	950

Table 9: Weibull Y with $\beta = 0.5$ and geometric N.

6 Distributions with regularly varying hazard rate

In this section we assume that $\overline{F}(x) = \mathrm{e}^{-\Lambda(x)}$ where $\Lambda(x) = \int_0^x \lambda(y)\,\mathrm{d}y$ and $\lambda(x)$ is regularly varying with index $\beta - 1$ and $\beta < \beta_0 = \log(3/2)/\log(2)$. We further assume that $\lambda(x)$ is $m+1$ times differentiable and that $\lambda^{(m+1)}$ is regularly varying. It follows that the distribution of F is semiexponential (cf. [12, Definition 1.4]) and hence subexponential. To exclude regularly varying distribution we will assume that $\lim_{x\to\infty} \lambda(x)x = \infty$ (and hence $\overline{F}(x) = \mathrm{o}(x^{-\gamma})$ for all $\gamma > 0$). Using Karamata's Theorem (e.g. [13]) it is easy to see that

$$f^{(m)}(x) \sim (-1)^m \lambda(x)^{m+1}\overline{F}(x).$$

Remark 15. In [10] for the same class of distributions (without the bound on β) it is shown that the higher order asymptotic up to the term $f^{(m-1)}(x)$ holds if $\liminf_{x\to\infty} x\lambda(x)/\log(x) > 0$ and $\lim_{x\to\infty}\lambda(x) = 0$. So our result is a little bit more general for distributions close to the regularly varying distributions.

Theorem 16. *Assume that $\lambda(x)$ is regularly varying with index $\beta - 1$ and $\beta < \beta_0 = \log(3/2)/\log(2)$. Assume further that $\lambda(x)$ is $m+1$ times differentiable, that $\lambda^{(m+1)}$ is regularly varying and that $x\lambda(x) \to \infty$. If $\mathbb{E}(1+\epsilon)^N < \infty$ for some $\epsilon > 0$. Then the estimator $Z_{m,N}$ satisfies*

$$\mathbb{V}\mathrm{ar}\ Z_{m,N} \sim \frac{1}{(m+1)!^2}\,\mathbb{V}\mathrm{ar}(N(S_{N-1})^{m+1}) f^{(m)}(x)^2.$$

Proof. In the proof of Theorem 2, replace Lemmas 7 and 8 with Lemmas 17 and 18 below. The rest is obvious adaptations. □

Lemma 17. *Assume that $-\log(\overline{F}(x)) = \int_0^x \lambda(z)\mathrm{d}z$ with $\lambda(x) = L(x)x^{\beta-1}$ and $\beta < 1$. Then for every $\epsilon > 0$ there exists an $C_\epsilon > 1$ such that*

$$\frac{\overline{F}\big(M_{n-1} \vee (x - S_{n-1})\big)}{\overline{F}(x)} \le C_\epsilon n^{\alpha(1+\epsilon)} \left(\frac{\overline{F}(M_{n-1})}{\overline{F}(M_{n-1}+S_{n-1})}\right)^{1+\epsilon}.$$

Proof. Since $\lambda(x) \sim \sup_{z>x}\lambda(z)$ ($\lambda(x)$ is regularly varying) for every $\epsilon > 0$ there exists an x_0 such that for $x > x_0$ and $z > 0$ $\lambda(x+z) \le (1+\epsilon)\lambda(x)$.

The inequality is obvious if $x \le M_{n-1} + S_{n-1}$. Otherwise, let $z = x - M_{n-1} - S_{n-1}$. if $M_{n-1} > x_0$ then

$$\begin{aligned}
\log\frac{\overline{F}\big(M_{n-1} \vee (x - S_{n-1})\big)}{\overline{F}(x)} &= \log\frac{\overline{F}(x - S_{n-1})}{\overline{F}(x)} = \log\frac{\overline{F}(M_{n-1}+z)}{\overline{F}(z + M_{n-1} + S_{n-1})} \\
&= \int_{M_{n-1}+z}^{z+M_{n-1}+S_{n-1}} \lambda(u)\,\mathrm{d}u \le (1+\epsilon)\int_{M_{n-1}}^{M_{n-1}+S_{n-1}} \lambda(u)\,\mathrm{d}u \\
&= \log\left(\frac{\overline{F}(M_{n-1})}{\overline{F}(M_{n-1}+S_{n-1})}\right)^{1+\epsilon}.
\end{aligned}$$

If $M_{n-1} \le x_0$ then for $x > 2(n-1)x_0$ and some $K_1 > 0$.

$$\frac{\overline{F}(x - S_{n-1})}{\overline{F}(x)} \le \frac{\overline{F}(x-(n-1)x_0)}{\overline{F}(x)} \le \frac{\overline{F}(x/2)}{\overline{F}(x)} \le K_1$$

and for $x \le 2(n-1)x_0$ we get by the Potter bounds that

$$\frac{\overline{F}(x - S_{n-1})}{\overline{F}(x)} \le \frac{1}{\overline{F}(2(n-1)x_0)} \le K_2(n-1)^{\alpha(1+\epsilon)}.$$

The Lemma follows since

$$\frac{\overline{F}(M_{n-1})}{\overline{F}(M_{n-1}+S_{n-1})} > 1.$$

□

Lemma 18. *If $\beta < \beta_0$ then there exists a $\delta > 0$ such that for all $k > 0$, $\ell \in \{1,2\}$, $\gamma > 0$ and $\epsilon > 0$ there exist a C such that.*

$$\mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^{\ell+\delta}\Big] < C(1+\epsilon)^n\,, \tag{6.1}$$

$$\mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^{\ell+\delta}; \, S_n > x/2\Big] \le C(1+\epsilon)^n x^{-\gamma}. \tag{6.2}$$

Proof of Lemma 18. As in the proof of Lemma 8 it is enough to prove the Lemma for $\ell = 2$ and with $(1+\epsilon)^n$ replaced by $n^\tau(1+\epsilon)^n$ where τ might dependent on k. Since $\beta 2^\beta < 1$ and $3 - 2^{1+\beta} > 0$ for $\beta < \log(3/2)/\log(2)$, we can choose x_0 (bigger than the x_0 of Lemma 19), δ and γ such that for $x \ge x_0$

$$1 - \exp\big\{-(1-(\beta+\gamma)(2+\delta)^\beta)\Lambda(x)\big\} + C_{\delta,\gamma}\frac{\Lambda\,(x)}{x} \le 1+\epsilon, \tag{6.3}$$

$(1+\gamma)(\beta+\gamma)(2+\delta)^\beta < 1$ and $3 + \delta - (1+\gamma)(2+\delta)^{1+\beta} > 0$. First note that for every $\epsilon > 0$ there exists a C_1 with

$$\begin{aligned}\mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^{2+\delta}; \, M_n \le x_0\Big] &\le \frac{x_0^k}{\overline{F}\big((n+1)x_0\big)^{2+\delta}} \\ &= x_0^k\,\mathrm{e}^{(2+\delta)\Lambda(x_0(n+1))} \le C_1(1+\epsilon)^n\,.\end{aligned}$$

By the same exchangeability argument as for the Asmussen-Kroese estimator, we get that for every $x \ge 0$

$$\begin{aligned}&\mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^{2+\delta}; M_n > x\Big] \\ &\quad = n\,\mathbb{E}\Big[Y_n^k\Big(\frac{\overline{F}(Y_n)}{\overline{F}(2Y_n+S_{n-1})}\Big)^{2+\delta}; \, Y_n > x, \, M_n = Y_n\Big].\end{aligned}$$

If $x > x_0$ we get with an iterative application of Lemma 19 and (6.3) that

$$\begin{aligned}
&\mathbb{E}\Big[(Y_n^k\Big(\frac{\overline{F}(Y_n)}{\overline{F}(2Y_n+S_{n-1})}\Big)^{2+\delta};\, M_n = X_n,\, Y_n > x\Big] \\
&\quad = \int_{x_n=x}^{\infty}\int_{[0,y_n]^{n-1}} y_n^k \prod_{i=1}^{n}\lambda(y_i) \\
&\qquad\quad \exp\Big\{-(2+\delta)\Lambda(y_n) + (2+\delta)\Lambda\Big(2y_n + \sum_{i=1}^{n-1} y_i\Big) - \sum_{i=1}^{n}\Lambda(y_i)\Big\}\,\mathrm{d}\mathbf{y} \\
&\quad \le (1+\epsilon)^{n-1}\int_{y_n=x}^{\infty} y_n^k\lambda(y_n)\exp\big\{-(2+\delta)\Lambda(y_n) + (2+\delta)\Lambda\left(2y_n\right) - \Lambda(y_n)\big\}\,\mathrm{d}y_n \\
&\quad \le (1+\epsilon)^{n-1}\int_{y_n=x}^{\infty} y_n^k\lambda(y_n)\exp\big\{-(3+\delta-(1+\gamma)(2+\delta)^{1+\beta})\Lambda(y_n)\big\}\,\mathrm{d}y_n
\end{aligned}$$

where the last integral is uniformly bounded in x and (6.1) follows.

Since $x\lambda(x) \to \infty$ it follows that $\Lambda(x)/\log(x) \to \infty$ and hence for every $\gamma > 0$ we can find a C such that for all $x > x_0$

$$x^k\lambda(x)\exp\big\{-(3+\delta-(1+\gamma)(2+\delta)^{1+\beta})\Lambda(x)\big\} \le Cx^{-\gamma-1}.$$

Using the same arguments, we get that for $x > 2nx_0$

$$\begin{aligned}
&\mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^{2+\delta};\, S_n > x/2\Big] \le \mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^{2+\delta};\, M_n > x/(2n)\Big] \\
&\quad \le n(1+\epsilon)^{n-1}\int_{x_n=x/(2n)}^{\infty} x_n^k\lambda(x_n)\exp\big\{-(3+\delta-(1+\gamma)(2+\delta)^{1+\beta})\Lambda(x_n)\big\}\,\mathrm{d}x_n. \\
&\quad \le Cn(1+\epsilon)^{n-1}\int_{x_n=x/(2n)}^{\infty} x^{-\gamma-1}\mathrm{d}x_n = \frac{C}{\gamma}2^{\gamma}n^{\gamma+1}(1+\epsilon)^{n-1}.
\end{aligned}$$

So (6.2) holds if $x > 2nx_0$. If $x \le 2nx_0$, then by (6.1)

$$\mathbb{E}\Big[M_n^k\Big(\frac{\overline{F}(M_n)}{\overline{F}(M_n+S_n)}\Big)^{2};\, S_n > x/2\Big] \le C(1+\epsilon)^n \le C(2nx_0)^{\gamma}(1+\epsilon)^n x^{-\gamma}$$

and the Lemma follows. ☐

Lemma 19. *Let $c > 0$ and $(1+\gamma)(\beta+\gamma)(2+\delta)^{\beta} < 1$. Then there exists an $x > 0$ and constant $C_{\delta,\epsilon}$ such that for $x > x_0$*

$$\begin{aligned}
&\int_0^x \lambda(y)\exp\big\{(2+\delta)\Lambda\left(2x+c+y\right) - \Lambda(y)\big\}\,\mathrm{d}y \\
&\quad \le \exp\big\{(2+\delta)\Lambda\left(2x+c\right)\big\}\Big[1 - \exp\big\{-(1-(\beta+\gamma)(2+\delta)^{\beta})\Lambda(x)\big\} + C_{\delta,\epsilon}\frac{\Lambda\left(x\right)}{x}\Big].
\end{aligned}$$

Proof. Since $\lambda(x) \sim \sup_{x>z}\lambda(z)$, $\lambda(x) \sim \beta\Lambda(x)/x$ and $\lambda(x)$ is regularly varying, we get by Taylor's theorem that for some $0 < \xi_y < y$ and x large enough

$$\begin{aligned}
\Lambda\left(2x+c+y\right) &= \Lambda\left(2x+c\right) + y\lambda\left(2x+c+\xi_y\right) \\
&\le \Lambda\left(2x+c\right) + (\beta+\gamma)2^{\beta-1}\frac{y}{x}\Lambda\left(x\right).
\end{aligned}$$

Hence

$$\int_0^x \lambda(y) \exp\big\{(2+\delta)\Lambda\,(2x+c+y) - \Lambda(y)\big\}\,\mathrm{d}y$$

$$\leq \exp\big\{(2+\delta)\Lambda\,(2x+c)\big\} \int_0^x \lambda(y) \exp\left\{(\beta+\gamma)(2+\delta)^\beta \frac{y}{x}\Lambda\,(x) - \Lambda(y)\right\}\,\mathrm{d}y.$$

By partial integration

$$\int_0^x \lambda(y) \exp\left\{(\beta+\gamma)(2+\delta)^\beta \frac{y}{x}\Lambda\,(x) - \Lambda(y)\right\}\,\mathrm{d}y$$

$$= -\exp\left\{(\beta+\gamma)(2+\delta)^\beta \frac{y}{x}\Lambda\,(x) - \Lambda(y)\right\}\Bigg|_0^x$$

$$+ (\beta+\gamma)(2+\delta)^\beta \frac{\Lambda\,(x)}{x} \int_0^x \exp\left\{(\beta+\gamma)(2+\delta)^\beta \frac{y}{x}\Lambda\,(x) - \Lambda(y)\right\}\,\mathrm{d}y$$

$$\leq 1 - \exp\big\{-(1-(1+\epsilon)\beta(2+\delta)^\beta)\Lambda(x)\big\} + C_{\delta,\epsilon}\frac{\Lambda\,(x)}{x},$$

since for some x_1 and all $x_1 < y < x$ we have $\Lambda(x)/x \leq (1+\delta)\Lambda(y)/y$ and that

$$\int_0^{x_1} \exp\left\{(\beta+\gamma)(2+\delta)^\beta \frac{y}{x}\Lambda\,(x) - \Lambda(y)\right\}\,\mathrm{d}y$$

is uniformly bounded for $x > x_1$. □

References

[1] H. Albrecher, C. Hipp, and D. Kortschak, Higher-order expansions for compound distributions and ruin probabilities with subexponential claims, *Scand. Actuar. J.*, 2010, 105–135 (2010)

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

[3] S. Asmussen, and H. Albrecher, *Ruin Probabilities*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2nd ed., (2010)

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

[5] S. Asmussen, K. Binswanger, and B. Højgaard Rare events simulation for heavy–tailed distributions, *Bernoulli*, 6, 303–322 (2000)

[6] S. Asmussen, and P.W. Glynn, *Stochastic Simulation. Algorithms and Analysis*, Springer-Verlag (2007)

[7] S. Asmussen, and D. Kortschak, On error rates in rare event simulation with heavy tails, *Proceedings of the Winter Simulation Conference 2012.* Available from `www.thiele.au.dk/publications`.

[8] S. Asmussen, and D.P. Kroese, Improved algorithms for rare event simulation with heavy tails, *Adv. Appl. Probab.*, 38, 545–558 (2006)

[9] P. Barbe, and W. P. McCormick, Asymptotic expansions for infinite weighted convolutions of rapidly varying subexponential distributions, *Probab. Theory Related Fields*, 141, 155–180 (2008).

[10] P. Barbe, W. P. McCormick, and C. Zhang, Asymptotic expansions for distributions of compound sums of random variables with rapidly varying subexponential distribution, *J. Appl. Probab.*, 44, 670–684 (2007).

[11] A. Baltrūnas, and E. Omey, The rate of convergence for subexponential distributions, *Liet. Mat. Rink.*, 38, 1–18 (1998)

[12] A. A. Borovkov, On subexponential distributions and the asymptotics of the distribution of the maximum of sequential sums, *Sibirsk. Mat. Zh.*, 43, 1235–1264 (2002)

[13] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular Variation*, Cambridge University Press, Cambridge (1989)

[14] J. Blanchet, and P.W. Glynn, *Efficient rare-event simulation for the maximum of heavy-tailed random walks*, *Ann. Appl. Probab.*, 18, 1351–1378 (2008)

[15] J. Bucklew, *Large deviations Techniques in Decision, Simulation and Estimations*, Wiley (1990)

[16] P. Dupuis, K. Leder, and H. Wang, *Importance sampling for weighted-serve-the-longest-queue*, *Math. Oper. Res.*, 34, 642–660 (2009)

[17] P. Dupuis, K. Leder, and H. Wang, *Large deviations and importance sampling for a tandem network with slow-down*, *Queueing Syst.*, 57, 71–83 (2007)

[18] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extreme Events for Insurance and Finance*, Springer-Verlag (1997)

[19] S. Foss, D. Korshunov, and S. Zachary, *Introduction to Heavy-Tailed and Subexponential Distributions*, Springer-Verlag (2011).

[20] S. Ghamami, and S. M. Ross, Improving the Asmussen-Kroese type simulation estimators, *J. Appl. Probab.* (accepted for publication)

[21] J. Hartinger, and D. Kortschak, On the efficiency of the Asmussen-Kroese estimator and its applications to stop-loss transforms, *Bl. DGVFM*, 30, 363–377 (2009)

[22] P. Heidelberger, Fast simulation of rare events in queueing and reliability models, *ACM TOMACS*, 6, 43–85 (1995)

[23] S. Juneja, and P. Shahabuddin, Rare Event Simulation Techniques: An Introduction and Recent Advances, Handbook in Operations Research and Mangement Sciences, Volume 13: Simulation. Chapter 11, Elsevier, Editors: Shane Henderson and Barry Nelson, 291 – 350 (2006)

[24] E. Omey, and E. Willekens, Second-order behaviour of distributions subordinate to a distribution with finite mean, *Stoch. Models*, 3, 311–342 (1987)