J. Appl. Prob. **46,** 768–790 (2009)
Printed in England

IMPORTANCE SAMPLING FOR FAILURE PROBABILITIES IN COMPUTING AND DATA TRANSMISSION

SØREN ASMUSSEN,* *University of Aarhus*

Abstract

In this paper we study efficient simulation algorithms for estimating $\mathrm{P}(X > x)$, where X is the total time of a job with ideal time T that needs to be restarted after a failure. The main tool is importance sampling, where a good importance distribution is identified via an asymptotic description of the conditional distribution of T given $X > x$. If $T \equiv t$ is constant, the problem reduces to the efficient simulation of geometric sums, and a standard algorithm involving a Cramér-type root, $\gamma(t)$, is available. However, we also discuss an algorithm that avoids finding the root. If T is random, particular attention is given to T having either a gamma-like tail or a regularly varying tail, and to failures at Poisson times. Different types of conditional limit occur, in particular exponentially tilted Gumbel distributions and Pareto distributions. The algorithms based upon importance distributions for T using these asymptotic descriptions have bounded relative error as $x \to \infty$ when combined with the ideas used for a fixed t. Nevertheless, we give examples of algorithms carefully designed to enjoy bounded relative error that may provide little or no asymptotic improvement over crude Monte Carlo simulation when the computational effort is taken into account. To resolve this problem, an alternative algorithm using two-sided Lundberg bounds is suggested.

Keywords: Communications engineering; compound sum; computer reliability; conditioned limit theorem; Cramér root; exponential tilting; geometric sum; Gumbel distribution; integral asymptotics; Lundberg's inequality; rare event simulation; regular variation; RESTART

2000 Mathematics Subject Classification: Primary 65C05; 68M15
Secondary 60F05; 68O20

1. Introduction

Consider a task of length T that is subject to failures and must be restarted if a failure occurs before completion. For example, the task may be the execution of a computer program, the transmission of a file on a communications channel, or a conversation with a call center.

The distribution of the (ideal) task time T is denoted throughout by F and the distribution of the failure time U is denoted by G. For convenience, the densities f and g are assumed to exist, except when $T \equiv t$ is constant. Owing to the possibility of (multiple) failures, the total task time X can possibly be large (certainly, we always have $X \geq T$). Here we are interested in the distribution H of X or, more specifically, in its tail $\bar{H}(x) = \mathrm{P}(X > x)$.

This problem has a long history in computer science, where the model goes under the name of RESTART (see [6] for references). Nevertheless, a comprehensive description of the tail asymptotics of X was only recently provided by Sheahan *et al.* [17] and Asmussen *et al.* [6].

Received 10 November 2008; revision received 8 June 2009. Accepted by Peter Jagers, Coordinating Editor.
* Postal address: Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark. Email address: asmus@imf.au.dk

At about the same time (in part independently), Jelenković and Tan [15], [16] performed a related study in the communications engineering context; the main difference from [6] is an on–off assumption on the channel, which in the computer reliability context corresponds to incorporating repair times. Further aspects involve parallel computing [2] and checkpointing (fragmentation) [5].

In the early work of Sheahan *et al.* [17], a numerical comparison of approximations and simulated values was performed. This turned out to be an extremely demanding task computationally, since 10^8 independent copies of X were needed to be generated to obtain sufficiently precise estimates of $\mathrm{P}(X > x)$ in the range of x values under study. (In [17], 10^6 independent copies were quoted, but this was a typo.)

In the present paper we suggest and analyze some more sophisticated algorithms designed to reduce the computational effort. Given the literature on rare event simulation (surveyed in, e.g. [4, Chapter VI]), it is not unexpected that importance sampling is the main tool (though other ideas like conditional Monte Carlo and splitting have been used for specific purposes; see again *loc. cit.*). The classical idea when using importance sampling is to look for an asymptotic description of the conditional distribution given the rare event and use this as an importance distribution. This is also the path we follow here and it leads to some additional theoretical problems on the model, since we must analyze such problems as how failures accumulate within a long but fixed time horizon and what the asymptotics as $x \to \infty$ of T are given that $X > x$. We will see some rather nonstandard limit distributions arise.

For most applications, it would be of particular interest to assume that G is exponential, say at rate μ, and that F is either degenerate (say at t), gamma-like, or of power form. Here by gamma-like we mean

$$f(t) \sim ct^{\alpha-1}\mathrm{e}^{-\lambda t}, \qquad t \to \infty, \tag{1.1}$$

for suitable constants $c > 0$, $\alpha \in \mathbb{R}$, and $\lambda > 0$, where $f(t) \sim g(t)$ means that $f(t)/g(t) \to 1$ as $t \to \infty$; (1.1) incorporates as a special case the three distributions in the numerical example of [17]. By power form we mean $\log f(t)/t \to -\alpha - 1$; this covers as a special case a regularly varying f,

$$f(t) \sim \frac{L(t)}{t^{\alpha+1}}, \qquad t \to \infty, \tag{1.2}$$

with L slowly varying (since $\log L(t)/t \to 0$ for any slowly varying L). We shall therefore pay particular attention to these three specific cases.

The paper is organized as follows. In Section 2 we give the relevant preliminaries both on RESTART and rare event simulation. In particular, a crucial quantity for the rest of the paper is introduced, a Cramér-type root $\gamma(t)$. In Section 3 and Appendix B we study the simulation problem when $T \equiv t$ is deterministic. This is fairly standard in its simplest formulation since, as surveyed in Section 2, $X - T = X - t$ then admits a geometric sum representation, and it is folklore that the simulation of tails of light-tailed geometric sums is most efficiently carried out by exponential tilting; in the RESTART setting, this means involving $\gamma(t)$. However, we also discuss to what extent the evaluation of $\gamma(t)$ can be avoided.

The rest of the paper deals with the case of a random T. The asymptotic results of [6] exhibit great diversity depending on the specific form of the tails of F and G, and, for this reason, we have to expect the same to be true for the form of efficient rare event simulation algorithms. We consider two cases, in both of which G is taken to be exponential(μ). In Section 4 we study the gamma-like case, (1.1). Motivated by general principles for rare event simulation, the asymptotic behavior of T given $X > x$ is studied, and after appropriate centering, we obtain a nonstandard limit, the exponentially tilted Gumbel distribution Q_β. The use of this as an

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

importance distribution is discussed, and an important message is that importance sampling on T alone is only modestly efficient—to do better, we have to combine the importance sampling algorithm with the more sophisticated algorithms for geometric sums.

In Section 5, a similar discussion is carried out for the regularly varying case, (1.2). Here T, given $X > x$, needs to be both centered and scaled (not just centered), and the limit is Pareto. However, using the Pareto distribution (shifted and scaled back to T) as an importance distribution, we encounter an absolute continuity problem. This is resolved by combining this algorithm with another importance sampling algorithm.

Perhaps the most surprising feature of these algorithms is that even if the distribution of X is always heavy tailed when T has unbounded support, the ideas all come from the light-tailed area; in general, the methodologies for simulation of light versus heavy tails are intrinsically different (cf. [4, Chapter VI]).

The algorithms outlined above enjoy bounded relative error, a concept at the center of the rare event simulation literature (for a definition of rare event simulation, see Section 2.2) and generally considered to represent the ultimate improvement over crude Monte Carlo simulation one can hope for. However, focusing solely on the bounded relative error as the efficiency measure is misleading—we also need to consider the computational effort. This is done in Section 6, and a considerably more diverse picture emerges. A partial solution to the problem based upon two-sided Lundberg-type bounds is suggested in Section 7. Finally, Section 8 contains some numerical examples.

2. Preliminaries

2.1. The RESTART model

Consider a deterministic $T \equiv t$, and let $X(t)$ be the corresponding simple RESTART total time, $H_t(x) = \mathrm{P}(X(t) \le x)$.

As in [6], we can write $X(t) = t + S(t)$, where $S(t) = \sum_1^N U_i(t)$ is a geometric sum, $N, U_1(t), U_2(t), \ldots$ are independent such that $\mathrm{P}(N = n) = (1-\rho)\rho^n$ with $\rho = G(t)$, and the $U_i(t)$ have distribution $G_{|t}$, defined as G conditioned to $(0, t)$. That is, the cumulative distribution function is $\mathrm{P}(U_i(t) \le s) = G(s)/G(t)$ for $s \le t$ and $\mathrm{P}(U_i(t) \le s) = 1$ for $s > t$, and the density is $g(s)/G(t)$ for $s \le t$ and 0 for $s > t$. By the general theory for geometric sums [19] (see also [6]) we know that

$$\mathrm{P}(S(t) > x) \sim C_1(t)\mathrm{e}^{-\gamma(t)x},$$

where $\gamma(t)$ is the solution of

$$1 = \int_0^t \mathrm{e}^{\gamma u} g(u)\,\mathrm{d}u \tag{2.1}$$

and

$$C_1(t) = \frac{\bar{G}(t)}{\gamma(t)m(t)}, \quad \text{where} \quad m(t) = \int_0^t u\mathrm{e}^{\gamma(t)u} g(u)\,\mathrm{d}u.$$

Since $\mathrm{P}(X(t) > x) = \mathrm{P}(S(t) > x - t)$, we therefore have

$$\bar{H}_t(x) = \mathrm{P}(X(t) > x) \sim C_2(t)\mathrm{e}^{-\gamma(t)x}, \quad \text{where} \quad C_2(t) = \mathrm{e}^{\gamma(t)t}C_1(t).$$

From [6] we also quote the two-sided Lundberg inequality:

$$\mathrm{e}^{-\gamma(t)x} \le \bar{H}_t(x) \le \mathrm{e}^{\gamma(t)t}\mathrm{e}^{-\gamma(t)x}. \tag{2.2}$$

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

It was shown in [6] that, for a general G, $\gamma(t) \sim \mu\bar{G}(t)$ as $t \to \infty$, where $1/\mu$ is the mean of G. For the exponential case, we shall need certain refinements and related results that are proved/collected in Appendix A. In particular,

$$\mu e^{-\mu t} \le \gamma(t) = \mu e^{-\mu t} + \mu^2 t e^{-2\mu t} + o(t e^{-2\mu t}) \quad \text{as } t \to \infty, \tag{2.3}$$

$$\gamma(t) = -\frac{\mu \log t}{t(1 + o(1))} \quad \text{as } t \downarrow 0. \tag{2.4}$$

For a random T, we write the total task time as $X = X(T)$, with the understanding that N and the $U_i(t)$ have the same distributions as above given $T = t$. Thus, the distribution H is given by $H(x) = \int_0^\infty H_t(x) f(t)\, dt$.

2.2. Rare event simulation

Consider the probability $z(x)$ of an event $A(x)$ (in our case, $\{X > x\}$) that is rare in the sense that $z(x) \to 0$ as $x \to \infty$. As in [4], we call a random variable $Z(x)$ an estimator for $z(x)$ if $Z(x)$ can be generated by simulation and is unbiased, i.e. $E\, Z(x) = z(x)$. A family $\{Z(x)\}_{x>0}$ of such estimators (or just $Z(x)$) is said to have bounded relative error if $\operatorname{var} Z(x) = O(z(x)^2)$ as $x \to \infty$, and to be logarithmically efficient if $\operatorname{var} Z(x) = O(z(x)^{2-\varepsilon})$ for all $\varepsilon > 0$ (cf. [4, p. 159]). In practice, the estimate of $z(x)$ for a given x is obtained by averaging R replications of $Z(x)$, and Gaussian confidence intervals can be produced in a standard way by computing the empirical variance.

If we (in a nonstandard terminology) define the *logarithmic efficiency factor* of an estimator $Z(x)$ as

$$\sup\left\{p > 0 \colon \frac{\operatorname{var} Z(x)}{z(x)^p} \to 0\right\},$$

then the crude Monte Carlo method has logarithmic efficiency factor 1 and an estimator that is logarithmically efficient or has bounded relative error has logarithmic efficiency factor at least 2.

The traditional approach to exhibiting estimators with logarithmic efficiency factor greater than 1 via importance sampling is to provide an asymptotic description of the conditional distribution $P(\cdot \mid A(x))$ given the rare event $A(x)$, and to use this as an importance distribution. The philosophy is that sampling from $P(\cdot \mid A(x))$ yields a zero-variance estimator, so that an importance distribution that is close hopefully has a small variance.

As already touched upon in Section 1, the computational effort also needs to be taken into account; this is often neglected in the rare event simulation literature. We defer the discussion of this to Section 6.

3. Simulation algorithms for a deterministic $T \equiv t$

In this section we discuss efficient algorithms for the simulation of $z(x) = P(S(t) > x)$ for a fixed t. One of them (Algorithm 3.1, below) has bounded relative error. Replacing x by $x - t$ gives algorithms with bounded relative error for the simulation of $\bar{H}_t(x)$ (the case of a random T is the subject of the rest of the paper and requires more work). The other approach, Algorithm 3.2, below, is conceptually simpler and reduces variance by an exponential factor, but does not have bounded relative error.

We will allow G to be general, not necessarily exponential. We write

$$S_n = U_1 + \cdots + U_n, \qquad \tau(x) = \inf\{n \colon S_n > x\}.$$

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

Recall from Section 2 that $G_{|t}$ denotes G conditioned to $(0,t)$, and define $G_{\gamma(t)}$ as the distribution on $(0,t)$ with density $g_{\gamma(t)}(y) = \mathrm{e}^{\gamma(t)y} g(y),\ 0 < y < t$.

The following algorithm is a special case of the one given in [4, Exercise 2.3, p. 172] for general geometric sums (see also [10]). An outline of the approach is given in Appendix B. We need to determine a certain root and to define a corresponding exponentially tilted distribution. When specialized to the RESTART setting, it is easy to see that the root is precisely $\gamma(t)$ and that the exponentially tilted distribution becomes $G_{\gamma(t)}$ (see Remark B.1). This yields the following algorithm.

Algorithm 3.1. *Generate $U_1, U_2, \ldots$ from $G_{\gamma(t)}$. Stop the simulation at $\tau(x)$ and return the estimator $Z_1(x) = \exp\{-\gamma(t) S_{\tau(x)}\}$.*

Noting that $G_{\gamma(t)}$ has finite mean because the support is finite, the results of Appendix B at once give the following result.

Theorem 3.1. *The estimator $Z_1(x)$ is unbiased for $z(x)$ and has bounded relative error. That is, $\mathrm{var}_{\gamma(t)}\, Z_1(x) = \mathrm{O}(z(x)^2)$ as $x \to \infty$.*

Random variate generation from $G_{\gamma(t)}$ as well as root finding may sometimes be tedious. A simpler idea is to take advantage of the special feature of bounded support (which is not available for general geometric sums) and simulate using the distribution $G_{|t}$. This leads to the following algorithm.

Algorithm 3.2. *Generate $U_1(t), U_2(t), \ldots$ from $G_{|t}$. Stop the simulation at $\tau(x)$ and return the estimator $Z_2(x) = G(t)^{\tau(x)}$.*

Proposition 3.1. *The estimator $Z_2(x)$ is unbiased for $z(x)$. Furthermore, $\mathrm{var}_{|t}\, Z_2(x)$ is of order $\mathrm{e}^{-(\gamma(t)+\xi(t))x}$, where $\xi(t)$ is the solution of*

$$1 = G(t) \int_0^t \mathrm{e}^{(\gamma(t)+\xi(t))u} g(u)\, \mathrm{d}u \tag{3.1}$$

and satisfies $0 < \xi(t) < \gamma(t)$. That is, the logarithmic efficiency factor is $1+\xi(t)/\gamma(t) \in (1,2)$.

Proof. For $u < t$, we have

$$\mathrm{P}_{|t}(U_1 \in \mathrm{d}u) = \frac{g(u)\,\mathrm{d}u}{G(t)} = \frac{\mathrm{e}^{-\gamma(t)u}}{G(t)}\, \mathrm{P}_{\gamma(t)}(U_1 \in \mathrm{d}u),$$

and it follows by a standard extension to stopping times (see, e.g. [4, pp. 131–132]) that

$$\begin{aligned} \mathrm{E}_{|t}\, Z_2(x) &= \mathrm{E}_{\gamma(t)} \left[\frac{\exp\{-\gamma(t) S_{\tau(x)}\}}{G(t)^{\tau(x)}} Z_2(x) \right] \\ &= \mathrm{E}_{\gamma(t)} \exp\{-\gamma(t) S_{\tau(x)}\} \\ &= \mathrm{E}_{\gamma(t)}\, Z_1(x) \\ &= z(x), \end{aligned}$$

showing unbiasedness.

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

Since (3.1) can be rewritten as $1 = G(t)\,\mathrm{E}_{\gamma(t)} \exp\{\xi(t)U_1\}$, it follows in a similar way that

$$\begin{aligned}\mathrm{E}_{|t}\, Z_2(x)^2 &= \mathrm{E}_{\gamma(t)}\left[\frac{\exp\{-\gamma(t)S_{\tau(x)}\}}{G(t)^{\tau(x)}} Z_2(x)^2\right] \\ &= \mathrm{E}_{\gamma(t)}[\exp\{-\gamma(t)S_{\tau(x)}\}G(t)^{\tau(x)}] \\ &= \mathrm{E}_{\gamma(t)}\left[\frac{\exp\{-(\gamma(t)+\xi(t))S_{\tau(x)}\}\exp\{\xi(t)S_{\tau(x)}\}}{(\mathrm{E}_{\gamma(t)}\exp\{\xi(t)U_1\})^{\tau(x)}}\right].\end{aligned}$$

Using $|S_{\tau(x)} - x| \le t$, we show that this expression is bounded above and below by a constant times

$$\mathrm{e}^{-(\gamma(t)+\xi(t))x}\,\mathrm{E}_{\gamma(t)}\left[\frac{\exp\{\xi(t)S_{\tau(x)}\}}{(\mathrm{E}_{\gamma(t)}\exp\{\xi(t)U_1\})^{\tau(x)}}\right].$$

But the expectation is the expectation of an exponential Wald martingale stopped at $\tau(x)$. The condition for optional stopping (see [3, p. 362]) is trivially satisfied because, by positivity, $\tau(x)$ is automatically finite for any exponential tilting of $\mathrm{P}_{\gamma(t)}$. Thus, the expectation is indeed 1, and so the order of $\mathrm{var}_{|t}\, Z_2(x)$ is as asserted.

To complete the proof, it remains to show that $0 < \xi(t) < \gamma(t)$. Clearly, the right-hand side of (3.1) is increasing in $\xi(t)$. The value at $\xi(t) = 0$ is $G(t) < 1$ because of the definition of $\gamma(t)$. This implies that $\xi(t) > 0$. Similarly, $\xi(t) < \gamma(t)$ will follow if we can show that the value at $\gamma(t)$ is greater than 1. But this value is

$$G(t)\int_0^t \mathrm{e}^{2\gamma(t)u} g(u)\,\mathrm{d}u = G(t)^2\,\mathrm{E}_{|t} \exp\{2\gamma(t)U_1\} > G(t)^2(\mathrm{E}_{|t}\exp\{\gamma(t)U_1\})^2 = 1,$$

where in the last step we have used the fact that the definition of $\gamma(t)$ can be rewritten as

$$1 = \int_0^t \mathrm{e}^{\gamma(t)u} g(u)\,\mathrm{d}u = G(t)\,\mathrm{E}_{|t} \exp\{\gamma(t)U_1\}.$$

The last part of Proposition 3.1 shows that Algorithm 3.2 does indeed provide exponential variance reduction (at rate $\xi(t)$), but does not have bounded relative error (for this, $\xi(t) \ge \gamma(t)$ would be necessary). However, the loss of efficiency vanishes as $t \to \infty$.

Proposition 3.2. *Assume that* $\hat{G}[\varepsilon] = \int_0^\infty \mathrm{e}^{\varepsilon t} G(\mathrm{d}t) < \infty$ *for some* $\varepsilon > 0$. *Then* $\xi(t) \sim \gamma(t) \sim \mu\mathrm{e}^{-\mu t}$ *as* $t \to \infty$. *That is, the logarithmic efficiency factor of Algorithm 3.2 goes to* 2 *as* $t \to \infty$.

Proof. See Appendix A.

Algorithm 3.2 is simpler than Algorithm 3.1 as it avoids finding the root and exponential tilting. However, for G exponential, the exponentially tilted distribution is truncated exponential. So, both algorithms require simulation from an exponential distribution truncated to $(0, t)$ (but with different parameters, $\mu_1 = \mu - \gamma(t)$ for Algorithm 3.1 and $\mu_2 = \mu$ for Algorithm 3.2). This can easily be done by inversion: generate the random variable as $-\log((1 - \exp\{-\mu_i t\})V/\mu_i)$ with V uniform on $(0, 1)$; cf. [4, Remark 2.4, p. 39]. Another way is acceptance–rejection: use the exponential(μ_i) distribution as proposed and reject values greater than t.

4. Simulation algorithms for a gamma-like T

If T is random, we expect a large X to occur as a consequence of a large T. Thus, the general principles of importance sampling surveyed in Section 2.2 suggest looking for the conditional distribution of T given $X > x$. Our result is as follows.

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

Theorem 4.1. *Assume that F is gamma-like, as in (1.1). Then the conditional distribution of $Y = Y(x) = \mu T - \log x - \log \mu$ given $X > x$ has a limit in distribution as $x \to \infty$, namely, the distribution Q_β with density*

$$q(y) = \frac{\exp\{-e^{-y} - \beta y\}}{\Gamma(\beta)}, \quad -\infty < y < \infty, \quad \text{where} \quad \beta = \frac{\lambda}{\mu}. \tag{4.1}$$

One simple message is that T given $X > x$ is of order $\log x/\mu$. When $\lambda = \mu$, $Q = Q_1$ is the Gumbel distribution familiar from extreme value theory. The cumulative distribution function at y is $\exp\{-e^{-y}\}$. When $\lambda \neq \mu$, Q_β is an exponentially tilted Gumbel distribution, and the properties are less standard. We return to this at the end of the section.

Proof of Theorem 4.1. Specializing Corollary 1.1 (or Theorem 2.2) of [6], we obtain

$$\mathrm{P}(X > x) \sim \frac{c\Gamma(\beta)}{\mu^{\alpha+\beta}} \frac{\log^{\alpha-1} x}{x^\beta}. \tag{4.2}$$

Let $f(t; x)$ be the density of T on the event $X > x$, that is, $f(t; x)\,\mathrm{d}t = \mathrm{P}(T \in \mathrm{d}t,\ X > x)$, and let $t(x, y) = (\log x + \log \mu + y)/\mu$. Then the density $q(y \mid x)$ at y of Y given $X > x$ is $f(t(x, y); x)/\mu\,\mathrm{P}(X > x)$. Using (2.3) gives

$$\gamma(t(x, y)) = \frac{e^{-y}}{x} + \mathrm{O}\left(\frac{\log x}{x^2}\right).$$

It then follows from the two-sided Lundberg inequality, (2.2), that

$$\mathrm{P}(X > x \mid T = t(x, y)) \sim \exp\{-e^{-y}\},$$

and so

$$\begin{aligned}
q(y \mid x) &= \frac{1}{\mu\,\mathrm{P}(X > x)} f(t(x, y); x) \\
&= \frac{1}{\mu\,\mathrm{P}(X > x)} f(t(x, y))\,\mathrm{P}(X > x \mid T = t(x, y)) \\
&\sim \frac{\mu^{\alpha+\beta-1}}{c\Gamma(\beta)} \frac{x^\beta}{\log^{\alpha-1} x} c t(x, y)^{\alpha-1} e^{-\lambda t(x,y)} \exp\{-e^{-y}\} \\
&\sim \frac{\mu^{\alpha+\beta-1}}{c\Gamma(\beta)} \frac{x^\beta}{\log^{\alpha-1} x} c \frac{\log^{\alpha-1} x}{\mu^{\alpha-1}} x^{-\beta} \mu^{-\beta} e^{-\beta y} \exp\{-e^{-y}\} \\
&= q(y).
\end{aligned}$$

But Scheffé's theorem (see [7, p. 224]) states that convergence of densities implies convergence in distribution.

Theorem 4.1 suggests that in the case of a gamma-like F as in (1.1), we should proceed as follows in order to simulate $z(x) = \mathrm{P}(X > x)$.

Algorithm 4.1. *Generate Y from the density q in (4.1), and let $T = t = (\log x + \log \mu + Y)/\mu$. If $T \leq 0$, return the estimator $Z_3(x) = 0$. Otherwise, calculate the likelihood ratio*

$$W = \frac{f(T)}{\mu q(\mu T - \log x - \log \mu)} = f(T) x^{-\beta} \mu^{-1-\beta} \Gamma(\beta) \exp\{\mu e^{-\mu T} x + \lambda T\},$$

compute the crude Monte Carlo estimator $Z_0(x - t)$ for $\mathrm{P}(S(t) > x - t) = \mathrm{P}(X(t) > x)$, and return the estimator $Z_3(x) = W Z_0(x - t)$.

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

The algorithm is motivated by the general principle of rare event simulation, i.e. that we should use a distribution close to the conditional distribution given the rare event (here $X > x$) as the importance distribution; cf. [4, Example 1.3, p. 128]. Indeed, the suggested importance distribution for T corresponds to the asymptotic description of the conditional distribution provided by Theorem 4.1, and the event $X > x$ is not rare when T is simulated from q. The following result shows that the algorithm does indeed have a substantially smaller asymptotic variance than the crude Monte Carlo method, but does not get close to bounded relative error or logarithmic efficiency.

Proposition 4.1. *The estimator $Z_3(x)$ has logarithmic efficiency factor at most $\frac{3}{2}$, and exactly equal to $\frac{3}{2}$ provided $\int_0^{t_0} f(t)^2\,\mathrm{d}t < \infty$ for all $t_0 < \infty$.*

In the proof, we shall need the following analytical result.

Lemma 4.1. *For any $t_0 > 0$,*

$$\int_{t_0}^{\infty} \exp\{-k\mathrm{e}^{-\eta t}x\}ct^{\delta-1}\mathrm{e}^{-\lambda t}\,\mathrm{d}t \sim \frac{\Gamma(\lambda/\eta)}{\eta^\delta k^{\lambda/\eta}}\frac{\log^{\delta-1}x}{x^{\lambda/\eta}} \quad \textit{as } x \to \infty.$$

Proof. This lemma is of the same type as a crucial step in the proof of Equation (4.2) of [6], but since its proof is short, we reproduce it here: substituting $s = \mathrm{e}^{-\eta t}$, the integral becomes

$$\int_0^{\mathrm{e}^{-\eta t_0}} \mathrm{e}^{-ksx}\frac{(-\log s)^{\delta-1}}{\eta^\delta}s^{\lambda/\eta-1}\,\mathrm{d}s,$$

and Karamata's Tauberian theorem (see [8, Theorems 1.5.11 and 1.7.1]) implies that this has the asserted asymptotics.

Proof of Proposition 4.1. From $\mathrm{E}\,Z_0(x-t)^2 = \mathrm{P}(S(t) > x - t) = \bar{H}_t(x)$, we obtain, by conditioning upon $T = t$, for a given $\varepsilon > 0$,

$$\begin{aligned}
\mathrm{E}\,Z_3(x)^2 &= \int_0^\infty \bar{H}_t(x)\frac{f(t)^2}{\mu^2 q(\mu t - \log x - \log\mu)^2}\frac{\mu q(\mu t - \log x - \log\mu)}{\mu}\,\mathrm{d}t \\
&= x^{-\beta}\int_0^\infty \bar{H}_t(x)f(t)^2\Gamma(\beta)\mu^{-1-\beta}\exp\{\mu\mathrm{e}^{-\mu t}x + \lambda t\}\,\mathrm{d}t \qquad (4.3)\\
&\geq k_1 x^{-\beta}\int_{t_0}^\infty \mathrm{e}^{-\gamma(t)x}t^{2\alpha-2}\mathrm{e}^{-2\lambda t}\exp\{\mu\mathrm{e}^{-\mu t}x + \lambda t\}\,\mathrm{d}t \\
&\geq k_1 x^{-\beta}\int_{t_0}^\infty t^{2\alpha-2}\exp\{-\mathrm{O}(t\mathrm{e}^{-2\mu t})x - \lambda t\}\,\mathrm{d}t \\
&\geq k_1 x^{-\beta}\int_{t_0}^\infty t^{2\alpha-2}\exp\{-k_2(\varepsilon)\mathrm{e}^{-(2-\varepsilon)\mu t}x - \lambda t\}\,\mathrm{d}t \\
&\sim k_3(\varepsilon)x^{-\beta}\frac{\log^{2\alpha-2}x}{x^{\lambda/(2-\varepsilon)\mu}} \\
&= k_3(\varepsilon)\frac{\log^{2\alpha-2}x}{x^{\beta(1+1/(2-\varepsilon))}},
\end{aligned}$$

where we have used the lower Lundberg bound in (2.2), the right-hand side inequality in (2.3), and Lemma 4.1. Combining this with (4.2) shows that the logarithmic efficiency factor is at most $1 + 1/(2-\varepsilon)$ and, therefore, at most $\frac{3}{2}$.

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

For the lower bound, first note that the upper Lundberg bound implies that (4.3) can be bounded by

$$k_5 x^{-\beta} \int_0^\infty f(t)^2 \exp\{\psi(t,x) - \lambda t\}\,\mathrm{d}t,$$

where $\psi(t,x) = \gamma(t)t - \gamma(t)x + \mu \mathrm{e}^{-\mu t}x$. Let I_1 and I_2 denote the contributions to this integral from the intervals $0 < t \le t_0$ and $t > t_1$, respectively, where t_0 and t_1 will be specified later. Then, with $k_7 = \sup_{t>t_0} \gamma(t)t$, we have, by the right-hand side of (2.3),

$$\begin{aligned} I_2 &\le k_6 \int_{t_1}^\infty t^{2\alpha-2} \exp\{k_7 - k_8 t \mathrm{e}^{-2\mu t}x - \lambda t\}\,\mathrm{d}t \\ &\le k_9 \int_{t_1}^\infty t^{2\alpha-2} \exp\{-k_8 t_1 \mathrm{e}^{-2\mu t}x - \lambda t\}\,\mathrm{d}t \\ &\sim k_{10} \frac{\log^{2\alpha-2} x}{x^{\lambda/2\mu}}. \end{aligned}$$

For $x \ge 1$, we can bound I_1 by

$$\int_0^{t_0} f(t)^2 \mathrm{e}^{\psi(t)x}\,\mathrm{d}t,$$

where $\psi(t) = \gamma(t)t - \gamma(t) + \mu \mathrm{e}^{-\mu t}$. Using (2.4) yields

$$\psi(t) \le -\mu \log t \left(1 - \frac{1}{t}\right)(1 + \mathrm{O}(1)) + \mu \quad \text{as } t \downarrow 0.$$

This shows that if t_0 is small enough then $\psi(t) < 0$ uniformly in $0 < t \le t_0$. Hence, using the assumption on f^2 shows that I_1 goes to 0 exponentially fast as $x \to \infty$.

Replacing t_1 by a smaller value, we may assume that $t_1 \le t_0$ and then (4.3) is bounded by $x^{-\beta}(I_1 + I_2)$, which in turn, by the above estimates, is $\mathrm{O}(x^{-\delta})$ for all $\delta < 3\beta/2$. This completes the proof.

Remark 4.1. An essential ingredient of the proof of Proposition 4.1 is informally to replace $\mathrm{P}(S(t-x) > x)$ for a large t by its Cramér–Lundberg approximation $C_2(t)\mathrm{e}^{-\gamma(t)x}$ (note that $C_2(t) \sim 1$ and $\gamma(t) \sim \mu \mathrm{e}^{-\mu t}$ as $t \to \infty$, so that the final approximation is $\exp\{-\mu \mathrm{e}^{-\mu t}x\}$); to justify this, Lundberg's inequality (and in part more refined estimates like (2.4)) was used. The same procedure will be used in Section 5, but since we have carefully given the details for the present case, we will not do so there.

To improve Algorithm 4.1, we involve further properties of the conditional distribution given the rare event, namely, the behaviour of $U_1(t), \ldots, U_{N(t)}(t)$ as used in Algorithms 3.1 and 3.2 (it is not a priori obvious that this will help since the event $X > x$ is not rare when T is simulated from q).

Algorithm 4.2. *Generate Y from the density q in (4.1), and let $T = t = (\log x + \log \mu + Y)/\mu$. If $T \le 0$, return $Z_4(x) = 0$. Otherwise, calculate the likelihood ratio*

$$W = \frac{f(t)}{\mu q(\mu t - \log x - \log \mu)},$$

compute one of the two estimators $Z_i(x-t)$ of Section 3 ($i = 1$ or 2), and return $Z_4(x) = W Z_i(x)$.

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

Theorem 4.2. *The estimator $Z_4(x)$ has bounded relative error provided $\int_0^{t_0} f(t)^2\,\mathrm{d}t < \infty$ for all $t_0 < \infty$.*

Proof. The proof is a slight variant of the last part of the proof of Proposition 4.1. First let $i = 1$. From $\mathrm{E}\,Z_1(x-t)^2 \le \mathrm{e}^{-2\gamma(t)(x-t)}$, we obtain, by conditioning upon $T = t$ and replacing $\bar{H}(t)$ by $\mathrm{e}^{-2\gamma(t)(x-t)}$ in (4.3),

$$\mathrm{E}\,Z_4(x)^2 \le x^{-\beta}\int_0^\infty \mathrm{e}^{2\gamma(t)t} f(t)^2 \exp\{-2\gamma(t)x + \mu\mathrm{e}^{-\mu t}x + \lambda t\}\,\mathrm{d}t.$$

Again, let I_1 and I_2 denote the contributions to this integral from the intervals $0 < t \le t_0$ and $t > t_1$, respectively. The proof that I_1 goes to 0 exponentially fast follows the same lines as above. Furthermore,

$$I_2 \le k_{12}\int_0^\infty \exp\{2k_7\}t^{2\alpha-2}\exp\{-\mu\mathrm{e}^{-\mu t}x - \lambda t\}\,\mathrm{d}t \sim k_{13}\frac{\log^{2\alpha-2}x}{x^\beta}.$$

This shows the assertion for $i = 1$. For $i = 2$, we have

$$\mathrm{E}\,Z_4(x)^2 \le k_{14}\int_0^\infty \mathrm{e}^{\gamma(t)t+\xi(t)t} f(t)^2 \exp\{-\gamma(t)x - \xi(t)x + \mu\mathrm{e}^{-\mu t}x + \lambda t\}\,\mathrm{d}t.$$

For I_1, we insert $\xi(t) \ge 0$, and are then back to the same integral as above. For I_2, we use $\xi(t) \ge k_{15}\gamma(t)$ for $t \ge t_1$ and can then use just the same estimates.

For the implementation of Algorithms 4.1 and 4.2, we note the following results.

Proposition 4.2. *The distribution Q_β in (4.1) has cumulative distribution function*

$$Q_\beta(y) = \frac{1}{\Gamma(\beta)}\int_{\mathrm{e}^{-y}}^\infty u^{\beta-1}\mathrm{e}^{-u}\,\mathrm{d}u.$$

Proof. In the identity $Q_\beta(y) = \int_{-\infty}^y q(v)\,\mathrm{d}v$, substitute $u = \mathrm{e}^{-v}$.

Corollary 4.1. *Assume that $\beta > 1$. Then a random variable Y with distribution Q_β can be generated as $Y = -\log Z_\beta$, where Z_β is gamma with density $z^{\beta-1}\mathrm{e}^{-z}/\Gamma(\beta)$.*

Proof. We have $\mathrm{P}(-\log Z_\beta \le y) = \mathrm{P}(Z_\beta \ge \mathrm{e}^{-y}) = \mathrm{P}(Y \le y)$.

5. Simulation algorithms for heavy-tailed F

In this section we assume that F is regularly varying; cf. (1.2). As in Section 4, the first step in the design of simulation algorithms is to look for the conditional distribution of T given $X > x$, that is, for an analogue of Theorem 4.1. We then face the difficulty that the results of [6] (more precisely part (2:1) of Theorem 2.1 of [6]) gives only logarithmic asymptotics. Part (i) of the following result improves this to sharp asymptotics.

Theorem 5.1. *Assume that G is exponential with rate μ, and that $f(t) = L(t)/t^{\alpha+1}$ with $\alpha > 0$ and $L(x)$ slowly varying as $t \to \infty$. Then,*

(i) $\bar{H}(x) \sim L(\log x)\mu^\alpha/\alpha\log^\alpha x$;

(ii) $\mathrm{P}(X > x,\ T > \log x/\mu) \sim L(\log x)\mu^\alpha/\alpha\log^\alpha x$;

(iii) $\mathrm{P}(X > x,\ T \le \log x/\mu) \sim L(\log x)\mu^\alpha E_1(\mu)/\log^{\alpha+1} x$.

Here $E_1(z) = \int_z^\infty v^{-1}\mathrm{e}^v\,\mathrm{d}v$ denotes the exponential integral; cf. [1, p. 228].

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

Note that the asymptotics in (i) and (ii) are the same, whereas the one in (iii) exhibits a lighter tail. Thus, the main contribution to $\mathrm{P}(X > x)$ comes from the event $T > \log x/\mu$.

Proof of Theorem 5.1. Obviously, (i) is a trivial consequence of (ii) and (iii), so it suffices to prove (ii) and (iii).

First consider (ii). Appealing to Remark 4.1 and substituting $t = \log x/\mu + y \log x/\mu$, we obtain

$$\begin{aligned}
&\frac{1}{L(\log x)} \mathrm{P}\left(X > x,\ T > \frac{\log x}{\mu} \right) \\
&\quad \sim \frac{1}{L(\log x)} \int_{\log x/\mu}^{\infty} \exp\{-\mu \mathrm{e}^{-\mu t} x\} \frac{L(t)}{t^{\alpha+1}}\, \mathrm{d}t \\
&\quad = \frac{1}{L(\log x)} \int_0^{\infty} \exp\{-\mu \mathrm{e}^{-y \log x}\} \frac{L(\log x(1/\mu + y/\mu))}{\log x(1/\mu + y/\mu)^{\alpha+1}} \frac{\log x}{\mu}\, \mathrm{d}y \\
&\quad \sim \frac{\mu^{\alpha}}{\log^{\alpha} x} \int_0^{\infty} \frac{R(x, y)}{(1+y)^{\alpha+1}}\, \mathrm{d}y, \qquad (5.1)
\end{aligned}$$

where $R(x, y) = L(\log x(1/\mu + y/\mu))/L(\log x)$. Choose $0 < \delta < \alpha$. By the Potter bounds (see [8, p. 25]), there exist k and y_0 such that $R(x, y) \le k y^{\delta}$ for all $y > y_0$, and by the uniform convergence theorem for slowly varying functions (see [8, p. 22] with $\rho = 0$ and $a = 1/\mu$), $R(x, y) \to 1$ uniformly on $(0, y_0)$. Since $R(x, y) \to 1$ on (y_0, ∞) also, dominated convergence applies to the integral over this interval, and we conclude that (5.1) asymptotically behaves like

$$\frac{\mu^{\alpha}}{\log^{\alpha} x} \int_0^{\infty} \frac{1}{(1+y)^{\alpha+1}}\, \mathrm{d}y = \frac{\mu^{\alpha}}{\alpha \log^{\alpha} x},$$

as claimed.

For (iii), $\mathrm{P}(X > x,\ T \le t_0)$ goes to 0 exponentially fast (at rate at least $\gamma(t_0)$) and can be neglected. Furthermore (cf. Remark 4.1 again),

$$\begin{aligned}
\mathrm{P}\left(X > x,\ t_0 \le T \le \frac{\log x}{\mu} \right) &\sim \int_{t_0}^{\log x/\mu} \exp\{-\mu \mathrm{e}^{-\mu t} x\} \frac{L(t)}{t^{\alpha+1}}\, \mathrm{d}t \\
&= \int_0^{\log x/\mu - t_0} \exp\{-\mu \mathrm{e}^{y}\} \frac{L(\log x/\mu - y)}{(\log x/\mu - y)^{\alpha+1}}\, \mathrm{d}y \\
&\sim \frac{L(\log x)\mu^{\alpha+1}}{\log^{\alpha+1} x} \int_0^{\infty} \exp\{-\mu \mathrm{e}^{y}\}\, \mathrm{d}y,
\end{aligned}$$

where in the last step we have used similar arguments as in the proof of (ii). But substituting $v = \mathrm{e}^{y}$, the integral becomes $E_1(\mu)/\mu$.

Theorem 5.2. *Assume that F is regularly varying, as in (1.2). Then the conditional distribution of $Y = \mu T/\log x - 1$ given $X > x$ has a limit in distribution as $x \to \infty$, namely, the Pareto(α) distribution P_α with density $p_\alpha(y) = \alpha/(1+y)^{\alpha+1}$, $y > 0$.*

It follows that, given $X > x$, the order of T is again $\log x/\mu$. However, whereas the deviation of T from $\log x/\mu$ remained of constant order in the gamma case, it now has to be scaled by $\log x$.

Proof of Theorem 5.2. We recall from Theorem 5.1(i) that

$$\mathrm{P}(X > x) \sim \frac{L(\log x)\mu^{\alpha}}{\alpha \log^{\alpha} x}.$$

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

Let $t(x, y) = \log x(1 + y)/\mu$. Then $\gamma(t(x, y)) \sim \mu e^{-\mu t(x,y)} = \mu e^{-y \log x}/x$, and, therefore (cf. Remark 4.1), $\mathrm{P}(S(t(x, y)) > x - t(x, y)) \sim \exp\{-\gamma(t(x, y))x\} \to 1$. With $f(t; x)$ as in the proof of Theorem 4.1, it follows that the density of Y given $X > x$ is

$$\begin{aligned}
&\frac{\log x}{\mu \,\mathrm{P}(X > x)} f(t(x, y); x)\, \mathrm{P}(S(t(x, y)) > x - t(x, y)) \\
&\quad \sim \frac{\alpha \log^{\alpha+1} x}{\mu^{\alpha+1} L(\log x)} \frac{L(\log x(1 + y)/\mu)\mu^{\alpha+1}}{(1 + y)^{\alpha+1} \log^{\alpha+1} x} \\
&\quad \sim \frac{\alpha}{L(\log x)} \frac{L(\log x)}{(1 + y)^{\alpha+1}} \\
&\quad = p_\alpha(y).
\end{aligned}$$

For simulation of $\mathrm{P}(X > x)$, Theorem 5.2 suggests using the distribution of $T(Y) = (Y \log x + \log x)/\mu$ as the importance distribution for T. This choice meets the difficulty that the support of $T(Y)$ is $(\log x/\mu, \infty)$, so that absolute continuity fails and the algorithm can only estimate $\mathrm{P}(X > x,\ T > \log x/\mu)$.

Algorithm 5.1. *Generate Y from the Pareto density p_α, and let $T = t = (Y \log x + \log x)/\mu$. Calculate the likelihood ratio*

$$W = \frac{\mu f(t)}{\log x\, p_\alpha(\mu t/\log x - 1)} = \frac{f(t)\mu^{\alpha+2} t^{\alpha+1}}{\alpha \log^{\alpha+2} x}.$$

Compute the crude Monte Carlo estimator $Z_0(x - t)$ for $\mathrm{P}(S(t) > x - t)$. Return the estimator $Z_5(x) = W Z_0(x - t)$ for $\mathrm{P}(X > x,\ T > \log x/\mu)$.

That only the crude Monte Carlo estimator of $\mathrm{P}(S(t) > x - t)$ needs to be used comes of course from the fact that the event $X > x$ is not rare even in the whole support of $T(Y)$.

Theorem 5.3. *Algorithm 5.1 has bounded relative error for estimating*

$$\mathrm{P}\left(X > x,\ T > \frac{\log x}{\mu}\right).$$

Proof. Appealing to Remark 4.1, we obtain

$$\begin{aligned}
\mathrm{E}\, Z_5(x)^2 &\sim \int_{\log x/\mu}^{\infty} \mathrm{P}(S(t) > x - t) \frac{f(t)^2 \mu^\alpha t^{\alpha+1}}{\alpha \log^\alpha x}\, \mathrm{d}t \\
&\le \frac{k_{15}}{\log^\alpha x} \int_{\log x/\mu}^{\infty} \exp\{-\mu e^{-\mu t} x\} \frac{L(t)^2}{t^{2\alpha+2}} t^{\alpha+1}\, \mathrm{d}t \\
&\le \frac{k_{15}}{\log^\alpha x} \int_{\log x/\mu}^{\infty} \frac{L(t)^2}{t^{\alpha+1}}\, \mathrm{d}t \\
&\sim \frac{k_{15} L(\log x/\mu)^2}{\log^{2\alpha} x} \\
&\sim \frac{k_{15} L(\log x)^2}{\log^{2\alpha} x} \\
&\sim k_{16}\, \mathrm{P}\left(X > x,\ T > \frac{\log x}{\mu}\right)^2,
\end{aligned}$$

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

where we have used Karamata's theorem for the integral asymptotics and (in the last step) Theorem 5.1(i).

To provide an unbiased estimate of $\mathrm{P}(X > x)$, we thus need an estimator of $\mathrm{P}(X > x,\ T \le \log x/\mu)$. We first note the following result.

Theorem 5.4. *The conditional distribution of $Y = \log x - \mu T$ given $X > x$ and $T \le \log x/\mu$ has a limit in distribution as $x \to \infty$, namely, the distribution R_μ with density $r_\mu(y) = \exp\{-\mu \mathrm{e}^y\}/E_1(\mu),\ y > 0$.*

It follows that, given $X > x$ and $T < \log x/\mu$, the order of T is again $\log x/\mu$. However, whereas the deviation of T from $\log x/\mu$ had to be scaled by $\log x$ when T was unrestricted as in Theorem 5.2, it now remains constant.

Proof of Theorem 5.4. Let $t(x, y) = \log x/\mu - y/\mu$. Since $T = \log x/\mu - Y/\mu$, it follows that the density of Y given $X > x$ and $T \le \log x/\mu$ is asymptotically

$$\begin{aligned}
&\frac{1}{\mu\,\mathrm{P}(X > x,\ T \le \log x/\mu)} f(t(x, y))\,\mathrm{P}(S(t(x, y)) > x - t(x, y)) \\
&\quad\sim \frac{\log^{\alpha+1} x}{\mu^{\alpha+1} L(\log x) E_1(\mu)} f\left(\frac{\log x}{\mu} - \frac{y}{\mu}\right) \exp\{-\mu \mathrm{e}^{-\mu(\log x/\mu - y/\mu)} x\} \\
&\quad\sim \frac{\exp\{-\mu \mathrm{e}^y\}}{E_1(\mu)}.
\end{aligned}$$

We are now led to the following algorithm for estimating $\mathrm{P}(X > x,\ T \le \log x/\mu)$.

Algorithm 5.2. *Generate Y from the density r_μ, and let $T = t = \log x/\mu - Y/\mu$. Calculate the likelihood ratio*

$$W = \frac{f(t)}{\mu r_\mu(\log x - \mu t)} = \frac{E_1(\mu) f(t) \exp\{\mu \mathrm{e}^{-\mu t} x\}}{\mu}.$$

Compute one of the two estimators $Z_i(x - t)$ of Section 3 ($i = 1$ or 2). Return the estimator $Z_6(x) = W Z_i(x - t)$ of $\mathrm{P}(X > x,\ T \le \log x/\mu)$.

Theorem 5.5. *Algorithm 5.2 has bounded relative error for estimating*

$$\mathrm{P}\left(X > x,\ T \le \frac{\log x}{\mu}\right).$$

Proof. First let $i = 1$. Then

$$\begin{aligned}
\mathrm{E}\, Z_6(x)^2 &= \int_0^{\log x/\mu} \frac{\mathrm{E}\, Z_1(x - t)^2 E_1(\mu) f(t)^2 \exp\{\mu \mathrm{e}^{-\mu t} x\}}{\mu}\,\mathrm{d}t \\
&\le k_{17} \int_0^{\log x/\mu} \frac{\mathrm{e}^{-2\gamma(t)x} f(t)^2 \exp\{\mu \mathrm{e}^{-\mu t} x\}}{\mu}\,\mathrm{d}t.
\end{aligned}$$

A similar argument as in the proof of Theorem 5.3, together with bound (2.3) for $\gamma(t)$, shows that this is asymptotically bounded by

$$\begin{aligned}
&k_{17} \int_{t_0}^{\log x/\mu} \mathrm{e}^{-2\gamma(t)x} \frac{L(t)^2}{t^{2\alpha+2}} \exp\{\mu \mathrm{e}^{-\mu t} x\}\,\mathrm{d}t \\
&\quad\le k_{17} \int_{t_0}^{\log x/\mu} \exp\{-\mu \mathrm{e}^{-\mu t} x\} \frac{L(t)^2}{t^{2\alpha+2}}\,\mathrm{d}t
\end{aligned}$$

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

$$= k_{18} \int_0^{\log x - \mu t_0} \exp\{-\mu e^y\} \frac{L(\log x/\mu - y/\mu)^2}{(1 + \log x/\mu - y/\mu)^{2\alpha+2}} \, dy$$
$$\sim k_{19} \frac{L(\log x)^2}{\log^{2\alpha+2} x}$$
$$\sim k_{20} \, \mathrm{P}(X > x)^2.$$

We omit the details for $i = 2$.

6. Computational effort

As already noted by Hammersley and Handscombe [14], considering variance alone as the performance measure of an algorithm may be misleading: we also need to consider the computational effort. They even quantified this effect in the statement

> The efficiency of a Monte Carlo process may be taken as inversely proportional to the product of the sampling variance and the amount of labor expended in obtaining this estimate.

The philosophy behind this is the fact that the 'inverse efficiency', var Z time Z, of a simulation estimator Z can be identified with the variance per unit computer time; here time Z is the expected computer time to generate Z. See [4, III.10] and [13]. The quantity time Z is of course hard, if not impossible, to identify in a mathematically rigorous way and time Z is, obviously, also highly implementation dependent, but in many situations a natural definition (up to a constant) suggests itself. For example, when simulating a random walk with positive drift up to its first passage of level x, it seems reasonable to take time $Z = x$. A good example of this is the Siegmund algorithm for simulating the probability $\mathrm{P}(M > x)$ that the maximum, M, of a random walk with light tails exceeds x; see [4, Chapter VI]. Here var Z decays asymptotically exponential, var $Z \sim D_1 e^{-\theta x}$ for some D_1, and $\theta > 0$ so that var Z and var Z time $Z = D_1 x e^{-\theta x}$ do not differ much in order (the same is true for many other standard rare event algorithms in the presence of light tails and the running time issue is therefore often ignored in the literature). However, with power tails, more disturbing examples exist. For example, for a random walk with increments with tail (say) $c/(1 + x)^\alpha$, $\alpha > 1$, Blanchet and Glynn [9] gave an algorithm with var $Z \sim D_2/x^{2\alpha-2}$ and time $Z \approx x$. Thus, var Z time Z is one power larger than var Z.

Crude Monte Carlo simulation of $\mathrm{P}(X > x)$ was implemented in [17] by generating X and returning $Z(x) = \mathbf{1}_{\{X>x\}}$. The effort in generating X is roughly proportional to the number of restarts, which in turn is roughly proportional to X. Thus, we take time $Z(x) = \mathrm{E}\, X$ and obtain

$$\text{var } Z(x) \text{ time } Z(x) = \mathrm{P}(X > x)(1 - \mathrm{P}(X > x)) \, \mathrm{E}\, X \approx \mathrm{P}(X > x) \, \mathrm{E}\, X. \tag{6.1}$$

For the algorithms considered so far in this paper, we can take time $Z(x) \approx x$ and the bounded relative error property implies that

$$\text{var } Z(x) \text{ time } Z(x) \approx \mathrm{P}(X > x)^2 x. \tag{6.2}$$

To compare these two expressions, we consider the case of F being exponential(λ) and G being exponential(μ). With $\beta = \lambda/\mu$, the probability of no restarts is $\int_0^\infty \lambda e^{-(\mu+\lambda)t} \, dt = \beta/(1+\beta)$. Thus, we have many restarts for small β and few restarts for large β. Furthermore, $\mathrm{P}(X > x)$ is of order $x^{-\beta}$ by (4.2). In particular, $\mathrm{E}\, X = \infty$ when $\beta \le 1$, and then the advantage of (6.2) over (6.1) is of course enormous. However, for $\beta > 1$, (6.1) is of order $x^{-\beta}$ and (6.2) is of order $x^{1-2\beta}$, which is only notably better if β is large.

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

The calculation does not, however, pay full justice to the crude Monte Carlo simulation, because in order to simulate $\mathrm{P}(X > x)$, it is not necessary to generate X, only $X\,\mathbf{1}_{\{X\le x\}}$. Thus, when $\beta < 1$, (6.1) has to be replaced by

$$\text{var } Z(x) \text{ time } Z(x) \approx \mathrm{P}(X > x)\,\mathrm{E}[X\,\mathbf{1}_{\{X\le x\}}] \approx x^{1-2\beta},$$

where in the last step we have used

$$\mathrm{E}[X\,\mathbf{1}_{\{X\le x\}}] = \int_0^x \mathrm{P}(X > s)\,\mathrm{d}s \approx \int_{x_0}^x s^{-\beta}\,\mathrm{d}s \approx x^{1-\beta}.$$

Thus, the order is of the same magnitude as (6.2). In order words, the importance sampling algorithm does not lead to any asymptotic improvement in the work-corrected variance.

This raises the problem of finding a complexity O(1) but still efficient estimator of $\mathrm{P}(S(t) > x - t)$. We may note that this probability is simply the probability that the largest interevent time of a Poisson(μ) process M on the interval $[0, x - t]$ is at most t (counting 0 and $x - t$ as epochs). An explicit expression for this is known (see [12], and also [11] and [18]), given the number, $m = M(x - t)$, of Poisson epochs, but is an alternating series with order m terms, so using this formula would not reduce the complexity from $\mathrm{O}(x)$ and could potentially be numerically unstable. We have therefore not pursued this approach, but suggest a different solution in the next section.

7. An algorithm exploiting Lundberg's inequality

The problem in the analysis of Section 6 is the order of increase in time $Z(x)$ in x. We now suggest an alternative estimator having the property time $Z(x) = \mathrm{O}(1)$. The estimator may lead to increased confidence bands, in particular for small x, but the problem vanishes as $x \to \infty$.

The idea is to avoid the $\mathrm{O}(x)$ simulation of $\mathrm{P}(S(t) > x - t)$ by just replacing this probability by its upper and lower Lundberg bounds. For example, in the gamma-exponential setting of Section 4, we have the following algorithm.

Algorithm 7.1. *Generate Y from the density q in (4.1), and let $T = t = (\log x + \log\mu + Y)/\mu$. If $T \le 0$, return the estimator $Z_3(x) = 0$. Otherwise, calculate $\gamma(t)$ and the likelihood ratio*

$$W = \frac{f(T)}{\mu q(\mu T - \log x - \log\mu)} = f(T)x^{-\beta}\mu^{-1-\beta}\Gamma(\beta)\exp\{\mu \mathrm{e}^{-\mu T}x + \lambda T\},$$

and let $Z_8'(x) = W\mathrm{e}^{-\gamma(t)x}$ and $Z_8''(x) = W\mathrm{e}^{-\gamma(t)(x-t)}$. Repeat R times and compute the empirical means $z_8'(x)$ and $z_8''(x)$, and the variances $s_7'(x)^2$ and $s_7''(x)^2$. Return the interval

$$\left(z_8'(x) - \frac{1.96 s_8'(x)}{R^{1/2}}, z_8''(x) + \frac{1.96 s_8''(x)}{R^{1/2}}\right). \tag{7.1}$$

We immediately obtain the following theorem.

Theorem 7.1. *Interval (7.1) is an asymptotic 95% confidence interval for $\mathrm{P}(X > x)$. That is, as $R \to \infty$, it contains $\mathrm{P}(X > x)$ with probability at least 95%.*

The limiting probability that (7.1) contains $\mathrm{P}(X > x)$ is of course somewhat larger than 95%. How much larger depends on how tight the Lundberg bounds are, but as noted above, these bounds are asymptotically tight as $t \to \infty$.

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

8. Numerical examples

We took F as exponential(1) and G as exponential(0.8). Thus, we are in the setting of Section 4 with $\alpha = 1$ and $\beta = 1.25$. We consider 10 x values, $10^{i/2}$, $i = 1, \ldots, 10$; in this range, $z(x) = \mathrm{P}(X > x)$ varies approximately from 10^{-1} to 10^{-6}.

We first implemented Algorithm 4.2 with $R = 1000$ replications.

Figure 1 shows the 95% two-sided confidence band (the scale is $\log_{10}$–$\log_{10}$, as for all figures except Figure 4). As is seen, the precision is excellent even with the modest value $R = 1000$, except for small values of x. The error appears to be decreasing in x, and this is further confirmed by Figure 2 which gives the relative error of the algorithm, as defined by the halfwidth of the confidence band divided by the simulated values. It may even look as if the relative error goes to 0, even if our theoretical analysis suggests it has a limit. This could be explained by Algorithm 3.2 becoming more and more efficient as $t \to \infty$ because $\xi(t) \uparrow \gamma(t)$

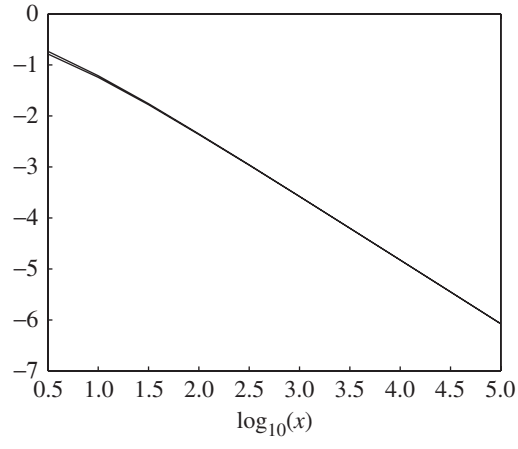

FIGURE 1: Confidence bands for Algorithm 4.2.

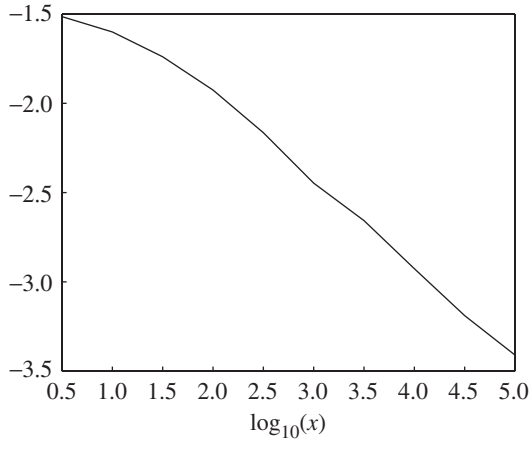

FIGURE 2: Relative precision of Algorithm 4.2.

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

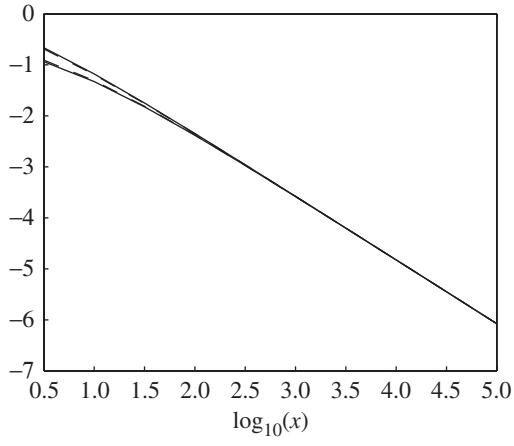

FIGURE 3: Confidence bands for Algorithm 7.1.

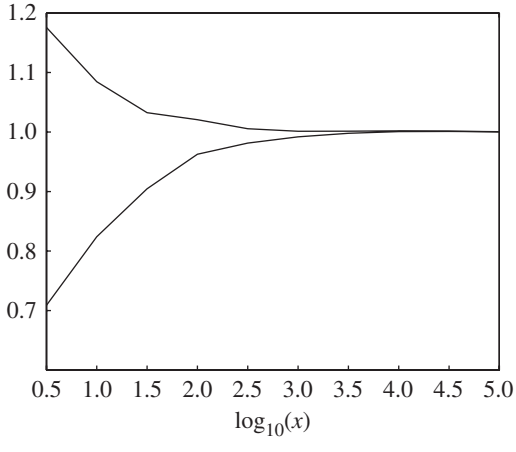

FIGURE 4: Lundberg bounds.

as $t \to \infty$ (cf. Proposition 3.2) and by the fact that the limit $\gamma(t)$ is not attained in the range of x values under consideration.

In comparison to Figure 1, the confidence bands produced by Algorithm 7.1 are given in Figure 3. The precision is comparable to Algorithm 4.2 except for the smallest values of x. Of course, we expect this to be due to the inaccuracy of the Lundberg bounds for small x, and this is confirmed by Figure 4, which shows the upper and lower Lundberg bounds divided by the simulated values.

Table 1 gives a comparison of the running times for Algorithms 4.2 and 7.1, more precisely, the ratio between the running time for Algorithm 7.1 and the running time for Algorithm 4.2 as produced by the MATLAB® commands 'tic' and 'toc'.

It is seen that the root finding in Algorithm 7.1 (implemented using the MATLAB routine 'fsolve') is indeed much more expensive than the $O(x)$ complexity of Algorithm 4.2 for small

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

TABLE 1.

x	$10^{1/2}$	10^{1}	$10^{3/2}$	10^{2}	$10^{5/2}$	10^{3}	$10^{7/2}$	10^{4}	$10^{9/2}$	10^{5}
	407	377	228	157	50	18	4.8	1.2	0.35	0.12

or moderate x. The overall picture when comparing this with the precision discussed above is that Algorithm 4.2 is preferable for small or moderate x, but Algorithm 7.1 is preferable for large x.

Finally, we took the opportunity to use our MATLAB program to check the accuracy of the approximations of [6], more precisely, (4.2) in the present paper. Figure 5 shows the simulated values versus the approximations, and Figure 6 shows the relative error of the

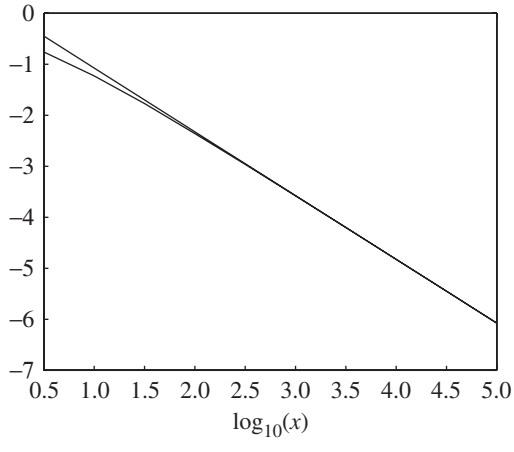

FIGURE 5: Simulated values versus approximations.

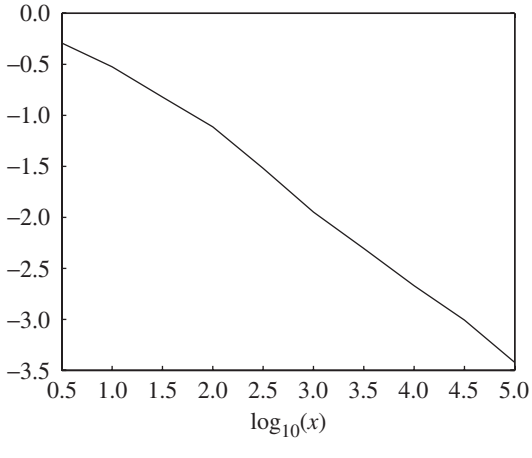

FIGURE 6: Relative error of approximation.

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

approximations, as defined by the absolute value of the difference between the simulated value and the approximation divided by the simulated value. The relative error indeed appears to go to 0, as expected, and the roughly linear shape of Figure 6 (cf. the log-log scale) suggests a roughly power-like rate of decrease.

Appendix A. Root properties

It was shown in [6] that, for a general G, $\gamma(t) \sim \mu\bar{G}(t)$ as $t \to \infty$. If G is exponential(μ), as assumed in the following, we shall need certain refinements and related results. First note that the defining equation (2.1) for $\gamma(t)$ means that

$$1 = \varphi(\gamma(t)), \quad \text{where} \quad \varphi(\gamma) = \frac{\mu}{\gamma - \mu}(\mathrm{e}^{(\gamma-\mu)t} - 1) \tag{A.1}$$

(note that t is fixed but suppressed in the definition of φ).

Proof of (2.3). The right-hand side of (2.1) (or, equivalently, of $\varphi(\gamma)$) is an increasing function of γ. Taking $\gamma = \mu\mathrm{e}^{-\mu t}$, this right-hand side becomes

$$\frac{\mu}{\mu - \mu\mathrm{e}^{-\mu t}}(1 - \exp\{(\mu\mathrm{e}^{-\mu t}t - \mu t)\}) < \frac{\mu}{\mu - \mu\mathrm{e}^{-\mu t}}(1 - \mathrm{e}^{-\mu t}) = 1.$$

Therefore, the desired solution $\gamma(t)$ must be greater than $\mu\mathrm{e}^{-\mu t}$.

Since

$$\begin{aligned}
\sum_{k=2}^{\infty} \frac{\gamma(t)^k}{k!} \int_0^t y^k \mu\mathrm{e}^{-\mu y}\,\mathrm{d}y &\le \gamma(t)^2 \sum_{k=0}^{\infty} \frac{\gamma(t)^k}{k!} \int_0^{\infty} y^{k+2}\mu\mathrm{e}^{-\mu y}\,\mathrm{d}y \\
&= \gamma(t)^2 \int_0^{\infty} \mu y^2 \mathrm{e}^{\gamma(t)y-\mu y}\,\mathrm{d}y \\
&\sim \gamma(t)^2 \int_0^{\infty} \mu y^2 \mathrm{e}^{-\mu y}\,\mathrm{d}y \\
&= \mathrm{O}(\gamma(t)^2)
\end{aligned}$$

as $t \to \infty$, we further obtain

$$\begin{aligned}
1 &= \int_0^t (1 + \gamma(t)y)\mu\mathrm{e}^{-\mu y}\,\mathrm{d}y + \mathrm{O}(\gamma(t)^2) \\
&= 1 - \mathrm{e}^{-\mu t} + \gamma(t)\frac{1}{\mu} - \gamma(t)\int_t^{\infty} \mu y\mathrm{e}^{-\mu y}\,\mathrm{d}y + \mathrm{O}(\gamma(t)^2) \\
&= 1 - \mathrm{e}^{-\mu t} + \gamma(t)\frac{1}{\mu} - \gamma(t)t\mathrm{e}^{-\mu t} - \frac{\gamma(t)}{\mu}\mathrm{e}^{-\mu t} + \mathrm{O}(\gamma(t)^2) \\
&= 1 - \mathrm{e}^{-\mu t} + \gamma(t)\frac{1}{\mu} - \gamma(t)t\mathrm{e}^{-\mu t}(1 + \mathrm{o}(1)) + \mathrm{O}(\gamma(t)^2).
\end{aligned}$$

This implies the right-hand side inequality in (2.3).

Equivalent forms of (A.1) are

$$\gamma(t) = \mu\mathrm{e}^{\gamma(t)t-\mu t}, \tag{A.2}$$

$$\gamma(t) = \mu + \frac{\log\gamma(t) - \log\mu}{t} \tag{A.3}$$

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

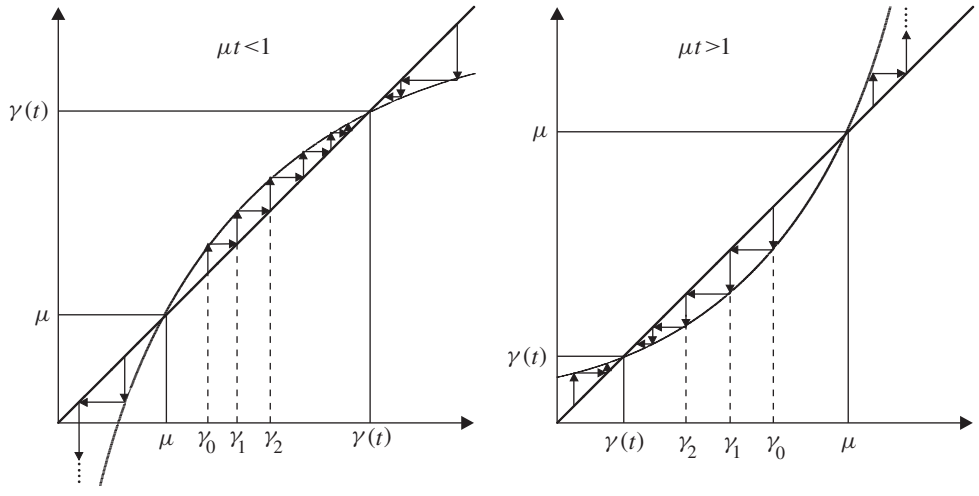

FIGURE 7: Simulated values versus approximations.

(indeed, (A.2) follows from (A.1) by trivial algebra, and (A.3) follows from (A.2) by taking logarithms). Obviously, there is no explicit solution. Since some of our algorithms require computation of $\gamma(t)$ for a large number of t, an efficient numerical scheme is needed. In our numerical examples, we used the MATLAB routine 'fsolve'. Another possibility is involving the Lambert W function (the root of $\theta e^{-\theta} = y$, in terms of which the solution of (A.2) can be expressed). In software environments, where general root-finding algorithms are unavailable, we may use traditional Newton–Raphson iteration, $\gamma_{n+1} = \gamma_n - \varphi(\gamma_n)/\varphi'(\gamma_n)$, or iterative schemes based upon (A.2) and (A.3).

Proposition A.1. *We have $\gamma(t) > \mu$, $\gamma(t) = \mu$, or $\gamma(t) < \mu$ according to whether $\mu t < 1$, $\mu t = 1$, or $\mu t > 1$, respectively. Furthermore, $\gamma = \gamma(t)$ can be computed as $\gamma = \lim_{n\to\infty} \gamma_n$, where in the case $\mu t > 1$,*

$$\gamma_{n+1} = \mu \exp\{\gamma_n t - \mu t\},$$

and the initial value γ_0 is chosen with $\gamma_0 < \mu$, and in the case $\mu t < 1$,

$$\gamma_{n+1} = \mu + \frac{\log \gamma_n - \log \mu}{t},$$

and the initial value γ_0 is chosen with $\gamma_0 > \mu$.

The need to distinguish between the cases $\mu t < 1$ and $\mu t > 1$ arises because (A.2) and (A.3) have the additional fixed point μ, and $\gamma(t)$ in (A.2) is attractive when $\mu t > 1$, but repulsive when $\mu t > 1$ (similar remarks apply to (A.3)); see Figure 7.

Proof of Proposition A.1. The first statement follows immediately since the right-hand side of (2.1) equals μt when $\gamma = \mu$ and is increasing in γ.

The convergence properties follow by standard arguments based upon convexity and concavity; see Figure 7.

Proof of (2.4): $\gamma(t) = -\mu \log t/t(1 + o(1))$ *as* $t \downarrow 0$. Define $\gamma_\delta = \mu - \delta \log t/t - \log \mu$. Then the right-hand side of (A.2) is of order $t^{-\delta}$ for $\gamma = \gamma_\delta$, whereas the left-hand side is of

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

order $|\log t|/t$. If $\delta > 1$, $t^{-\delta}$ increases faster than $|\log t/t|$, and since the right-hand side of (A.2) is convex and the left-hand side is affine, the desired solution, $\gamma(t)$, must be less than γ_δ. A similar argument shows that $\gamma(t) > \gamma_\delta$ when $\delta < 1$.

Proof of Proposition 3.2. That $\gamma(t) \sim \mu\bar{G}(t)$ is shown in [6]. For $\xi(t) \sim \gamma(t)$, note that the definition of $\xi(t)$ means that

$$\begin{aligned} 1 &= G(t)\int_0^t \mathrm{e}^{(\gamma(t)+\xi(t))u} g(u)\,\mathrm{d}u \\ &= (1-\bar{G}(t))\int_0^t \mathrm{e}^{\gamma(t)u}[1+\xi(t)u+\xi(t)\mathrm{O}(t^2\xi(t))]g(u)\,\mathrm{d}u \\ &= (1-\bar{G}(t))\left(1+\frac{\xi(t)}{\mu}+\mathrm{o}(\xi(t))\right), \end{aligned} \tag{A.4}$$

where we have used $\xi(t) < \gamma(t) \sim \mu\bar{G}(t)$ together with $\hat{G}[\varepsilon] < \infty$ to infer that $t^2\xi(t) \to 0$, and $\hat{G}[\varepsilon] < \infty$ and dominated convergence to infer that

$$\int_0^t u\mathrm{e}^{\gamma(t)u} g(u)\,\mathrm{d}u = \frac{1}{\mu} + \mathrm{o}(1).$$

However, (A.4) is only possible if $\xi(t) \sim \mu\bar{G}(t)$.

Appendix B. Simulation of geometric sums

Let $U_1^*, U_2^*, \ldots$ be i.i.d. with common distribution G^* concentrated on $(0,\infty)$, and let N be an independent geometric random variable, $\mathrm{P}(N = n) = (1-\rho)\rho^n$ for $n = 0, 1, \ldots$. Furthermore, define

$$S_n^* = U_1^* + \cdots + U_n^*, \qquad z(x) = \mathrm{P}(S_N^* > x), \qquad \tau^*(x) = \inf\{n : S_n^* > x\}.$$

In [4, Exercise 2.3, p. 172] (see also [10]), the following algorithm is suggested for simulation of $z(x)$ and it is claimed that it has bounded relative error as $x \to \infty$. (Note that the expression for the estimator in *loc. cit.* contains typos, corrected here.) As a preliminary, compute γ^*, the solution of

$$1 = \rho\int_0^\infty \mathrm{e}^{\gamma^* y} G^*(\mathrm{d}y). \tag{B.1}$$

Let G^* be the distribution defined by $G_{\gamma^*}(\mathrm{d}y)/G^*(\mathrm{d}y) = \rho\mathrm{e}^{\gamma^* y}$. To generate one replication of the estimator, proceed as follows.

Algorithm B.1. *Generate $U_1^*, U_2^*, \ldots$ from G_{γ^*}. Stop the simulation at $\tau^*(x)$ and return the estimator $Z^*(x) = \exp\{-\gamma^* S_{\tau^*(x)}\}$.*

To understand the algorithm, first note that $z(x) = \mathrm{P}(\tau^*(x) \le N)$. Next, let P_{γ^*} be the probability measure such that the U_i^* are i.i.d. with distribution G_{γ^*} and N remains independent and geometric. Then, by the definition of G_{γ^*},

$$\mathrm{P}(U_1^* \in \mathrm{d}u) = \frac{1}{\rho}\,\mathrm{E}_{\gamma^*}[\exp\{-\gamma^* U_1^*\}; U_1^* \in \mathrm{d}u].$$

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press

By a standard extension to stopping times (see, e.g. [4, pp. 131–132]), this implies that

$$z(x) = \mathrm{E}_{\gamma^*}\left[\frac{1}{\rho^{\tau^*(x)}}\exp\{-\gamma^* S_{\tau^*(x)}\};\, \tau^*(x) \le N\right] = \mathrm{E}_{\gamma^*}\exp\{-\gamma^* S^*_{\tau(x)}\},$$

where we have used the fact that N remains geometric and independent of the U_i^* under P_{γ^*}, i.e. the estimator $Z^*(x)$ is unbiased.

Furthermore,

$$\mathrm{E}_{\gamma^*} Z^*(x)^2 = \mathrm{E}_{\gamma^*}\exp\{-2\gamma^* S_{\tau(x)}\} \le \mathrm{e}^{-2\gamma^* x} = \mathrm{O}(z(x)^2),$$

where in the last step we have used the standard Cramér–Lundberg asymptotics $z(x) \sim C^*\mathrm{e}^{-2\gamma^* x}$ valid with $0 < C^* < \infty$ provided that G_{γ^*} has finite mean. This shows that $Z^*(x)$ has bounded relative error.

Remark B.1. For the geometric sum occurring in RESTART with $T \equiv t$, as discussed in Section 2, we have $\rho = G(t)$ and G^* is the distribution with density $g(y)/G(t)$, $0 < y < t$. Therefore, γ^* is the root $\gamma(t)$ defined in (2.1), and G_{γ^*} is the distribution with density $\mathrm{e}^{\gamma(t)y}g(y)$, $0 < y < t$.

Remark B.2. Algorithm B.1 may appear rather different from the best algorithm known for Poisson (rather than geometric) sums discussed in [4, Section VI.2d], where one exponentially tilts the whole distribution of S_N^*, leading to a new compound sum with changed Poisson parameter and exponentially tilted increment distribution. The tilting parameter, $\theta = \theta(x)$, is determined by $\mathrm{E}\, S_N^*\exp\{\theta S_N^*\}/\mathrm{E}\exp\{\theta S_N^*\} = x$. Performing the same operation for a geometric sum S_N^*, we can easily check that the relevant θ has limit γ^* so that the two algorithms asymptotically coincide.

References

[1] Abramowitz, M. and Stegun, I. A. (eds) (1972). *Handbook of Mathematical Functions*. Dover, New York.

[2] Andersen, L. N. and Asmussen, S. (2008). Parallel computing, failure recovery and extreme values. *J. Statist. Theory Pract.* **2,** 279–292.

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

[4] Asmussen, S. and Glynn, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer, New York.

[5] Asmussen, S. and Lipsky, L. (2008). Failure recovery in computing and data transmission: limit theorems for checkpointing. Working paper. Aarhus University.

[6] Asmussen, S. *et al.* (2008). Asymptotic behavior of total times for jobs that must start over if a failure occurs. *Math. Operat. Res.* **33,** 932–944.

[7] Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.

[8] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation*. Cambridge University Press.

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

[10] Blanchet, J. H. and Li, C. (2006). Efficient rare-event simulation for geometric sums. In *Proc. RESIM*, Bamberg, Germany.

[11] David, H. A. (1970). *Order Statistics*. John Wiley, New York.

[12] Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proc. R. Soc. London A* **125,** 54–59.

[13] Glynn, P. W. and Whitt, W. (1992). The asymptotic efficiency of simulation estimators. *Operat. Res.* **40,** 505–520.

[14] Hammersley, J. M. and Handscomb, D. C. (1964). *Monte Carlo Methods.* Methuen, London.

[15] Jelenković, P. and Tan, J. (2007). Can retransmissions of superexponential documents cause subexponential delays? In *Proc. IEEE INFOCOM* (Anchorage, May 2007), pp. 892–900.

[16] Jelenković, P. and Tan, J. (2007). Characterizing heavy-tailed distributions induced by retransmissions. Tech. Rep. EE2007-09-07, Columbia University.

[17] Sheahan, R., Lipsky, L., Fiorini, P. and Asmussen, S. (2006). On the distribution of task completion times for tasks that must restart from the beginning if failure occurs. In *ACM SIGMETRICS Performance Evaluation Review* Association for Computing Machinery, New York, pp. 24–26.
[18] Van Leeuwaarden, J. S. H., Löpker, A. H. and Janssen, A. J. E. M. (2008). Connecting renewal age processes and M/D/1 processor sharing queues through stickbreaking. EURANDOM Report 2008-17.
[19] Willmot, G. E. and Lin, S. X. (2001). *Lundberg Approximations for Compound Distributions with Insurance Applications* (Lecture Notes Statist. **156**), Springer, New York.

https://doi.org/10.1239/jap/1253279851 Published online by Cambridge University Press