Estimating Tail Probabilities of Heavy Tailed Distributions with Asymptotically Zero Relative Error

S. Juneja
Tata Institute of Fundamental Research, Mumbai
juneja@tifr.res.in

February 28, 2007

Abstract

Efficient estimation of tail probabilities involving heavy tailed random variables is amongst the most challenging problems in Monte-Carlo simulation. In the last few years, applied probabilists have achieved considerable success in developing efficient algorithms for some such simple but fundamental tail probabilities. Usually, unbiased importance sampling estimators of such tail probabilities are developed and it is proved that these estimators are asymptotically efficient or even possess the desirable bounded relative error property. In this paper, as an illustration, we consider a simple tail probability involving geometric sums of heavy tailed random variables. This is useful in estimating probability of large delays in $M/G/1$ queues. In this setting we develop an unbiased estimator whose relative error decreases to zero asymptotically. The key idea is to decompose the probability of interest into a known dominant component and an unknown small component. Simulation then focuses on estimating the latter 'residual' probability. Here we show that the existing conditioning methods or importance sampling methods are not effective in estimating the residual probability while an appropriate combination of the two estimates it with bounded relative error. As a further illustration of the proposed ideas, we apply them to develop an estimator for large delays in stochastic activity networks that has an asymptotically zero relative error.

1 Introduction

Consider non-negative, independent, identically distributed (i.i.d.) regularly varying heavy-tailed random variables $(X_i : i \geq 1)$. Let $S_n = \sum_{i=1}^{n} X_i$ and consider the problem of efficiently estimating, $P(S_n > u)$ and $P(S_N > u)$ for large u, where throughout this paper N denotes a geometrically distributed random variable. These are amongst the simplest problems involving heavy-tailed variables studied in the literature. One reason for importance of the probability $P(S_N > u)$ is that it equals the level crossing probability of a random walk when X_i's denote the ladder heights associated with the random walk (see, e.g., Feller 1971). The level crossing probabilities are in turn useful as the steady state waiting time tail probabilities in $GI/GI/1$ queues can be expressed as level crossing probabilities using Lindley's recursion. In practice, these transformations are particularly useful for $M/G/1$ queue as then the distribution of the ladder-heights X_i's is explicitly known (see, e.g., Asmussen 2003). Similarly, under certain conditions S_N has the same distribution as losses in an insurance model (see, e.g., Asmussen 2000).

These probabilities were first studied by Asmussen and Binswanger (1997) where they proposed a conditioning method to asymptotically efficiently estimate these. Asmussen, Binswanger and Hojgaard (2000) and Juneja and Shahabuddin (2002) proposed importance sampling techniques for estimating $P(S_n > u)$ and $P(S_N > u)$ asymptotically efficiently. Asmussen and Kroese (2006) propose algorithms for estimating $P(S_n > u)$ and $P(S_N > u)$ that have a stronger bounded relative error property. Blanchet and Liu (2006) consider the problem $P(S_n > nu)$ as $n \to \infty$ and $u > 0$ is fixed. For this and related problems they develop a novel state dependent change of measure that has the bounded relative error property.

Our interest in this paper is restricted to state-independent changes of measure. One reason for this is that often they are easier to implement. We refer the reader to Embrechts, Kluppelberg and Mikosch (1997) and Juneja and Shahabuddin (2006) for some finance and insurance applications of rare event probabilities involving heavy-tailed random variables.

In this paper, we note the following decomposition:

$$P(S_n > u) = P(\max_{i \le n} X_i > u) + P(S_n > u, \max_{i \le n} X_i \le u).$$

The probability $P(\max_{i \le n} X_i > u)$ is easily evaluated in closed form, and it is well known that

$$P(S_n > u) \sim P(\max_{i \le n} X_i > u)$$

when X_i have regularly varying heavy-tailed distributions. (Above discussion also holds for n replaced by N.) Therefore, we focus on developing efficient simulation techniques for estimating the residual probability $P(R_n(u))$ (also $P(R_N(u))$) where $R_n(u) = \{S_n > u, \max_{i \le n} X_i \le u\}$. Note that the above decomposition may also be viewed as using the indicator of the event $\{\max_{i \le n} X_i > u\}$ as a control variate in estimating the probability $P(S_n > u)$.

We focus on efficiently estimating the residual probability for two broad purposes:

1. Many general probabilities involving heavy tailed random variables allow for such a decomposition. For instance, in a single-server queueing set-up one may be interested in the probability of large delays in a queue during a busy cycle. This may be decomposed into two constituent components: a) when a service time in a busy cycle takes a large value, and b) the residual term. Here, even if the dominant term then does not have a simple closed form, it may be estimated via an appropriate set of combined analytical/simulation techniques. Then, the issue of efficient estimation of the residual probability is also of importance.

2. Typically, for the same order of magnitude of the rare event probabilities involving sums of independent random variables, the computational effort needed in the light tailed setting is orders of magnitude less than in the heavy tailed settings, even though in both the cases the proposed techniques are provably efficient (see, e.g., Juneja, Karandikar and Shahabuddin 2007). Therefore, in heavy tailed settings, there is a need for better methods to further speed-up simulations. As we see later in Section 5, the numerical results validate this as the proposed method perform orders of magnitude better than an existing asymptotically efficient scheme.

In our analysis of $P(R_n(u))$, we allow the distribution of X_i to be a function of i. This helps in generalizing results for $P(R_n(u))$ to the stochastic activity network settings where our interest is in the probability that the overall all network time exceeds a large threshold.

The main contributions of the paper are:

1. Through simple examples we illustrate that common importance sampling techniques and conditioning based techniques do not estimate residual probabilities asymptotically efficiently.

2. We first assume that $(X_i : i \geq 1)$ are i.i.d. and have regularly varying tail distributions. In this setting we develop a sharp asymptotic for $P(R_N(u))$.

3. We develop a combined importance sampling and conditioning based approach to develop an estimator for $P(R_N(u))$ with bounded relative error. This amounts to developing an unbiased estimator of $P(S_N > u)$ with a relative error that asymptotically (as $u \to \infty$) converges to zero. In a recent work Hartinger and Kortschak (2006) report an estimator for $P(S_N > u)$ that has asymptotically zero relative error*. The estimator proposed by them is based on the estimator considered in Asmussen and Kroese (2006) and is different from the one we propose.

4. We then develop sharp asymptotic for the probability $P(R_n(u))$ when $(X_i : i \leq n)$ are independent and have regularly varying tail distributions. Here we allow them to have different distributions. This generalizes the sharp asymptotic developed by Omey (1994, 2002) where they consider the case where $n = 2$ under less restrictive distributional assumptions.

5. We also develop an estimator for $P(R_n(u))$ using a combined importance sampling and conditioning based approach and show that it has bounded relative error.

6. As an application of our methodology, we use it to efficiently estimate the probability of large delays in stochastic activity networks where individual activity times have a regularly varying tail distributions. As discussed in Juneja et. al. (2007), stochastic activity networks are useful in modelling composite service requests on the web, and heavy tailed distributions are particularly useful in modelling web based processes.

In Section 2 we develop the mathematical framework. We also review some of the conditioning and importance sampling algorithms proposed in the literature. In Section 3 we show that the existing conditioning and importance sampling methods do not work well in efficiently estimating the residual probability. Here we also develop a sharp asymptotic for $P(R_N(u))$. We then develop a combined importance sampling and conditioning approach and show that the resultant estimator estimates $P(R_N(u))$ with bounded relative error. In Section 4, we develop a sharp asymptotic and an efficient estimation scheme for $P(R_n(u))$, when the individual X_i' are allowed to have different regularly varying heavy-tailed distributions. We also generalize these results to stochastic activity networks in this section. Finally, in Section 5, we illustrate the efficacy of the proposed methodology on a small numerical example. Some of the more technical proofs are relegated to the appendix.

2 Preliminaries

In this section we develop a preliminary mathematical framework useful to our analysis. We also discuss some existing conditioning and importance sampling based estimation approaches for $P(S_n > u)$ and $P(S_N > u)$. This is useful in positioning the proposed algorithms in relation to the existing literature.

As is well known, a function L is slowly varying if

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1$$

*We thank one of the editors for bringing this reference to our notice for all $t > 0$ (see Feller 1971). In our analysis for $(X_i : i \geq 1)$ i.i.d., we make the following assumption:

Assumption 1 *The pdf f of X_i exists and is regularly varying, i.e., it has the form*

$$f(x) = L(x)/x^{\alpha}$$

for $x \geq 0$ (it is zero otherwise), $\alpha > 2$, where L is a slowly varying function. Furthermore, the pdf is eventually non-increasing.

Let F denote the cdf of X_i and $\bar{F}$ denote its tail cdf (i.e., $\bar{F}(x) = 1 - F(x)$). It is well known that under Assumption 1,

$$\bar{F}(x) \sim L(x)/x^{\alpha-1}$$

where we say $a_x \sim b_x$ for non-negative sequences $\{a_x\}$ and $\{b_x\}$ if $\lim_{x\to\infty} a_x/b_x = 1$. Let P denote the resultant probability measure. Throughout this paper we assume that X has the same distribution as X_i whenever $(X_i : i \geq 1)$ are i.i.d.

Suppose that Z_u is an unbiased estimator of $\gamma(u) > 0$ for each $u > 0$. Let $V(Z_u)$ denote its variance. Note that $EZ_u^2 \geq \gamma(u)^2$ and $\log\gamma(u) < 0$. Therefore,

$$\limsup_{u\to\infty} \frac{\log EZ_u^2}{\log\gamma(u)} \leq 2.$$

We say that the collection of estimators $(Z_u : u > 0)$ asymptotically efficiently estimate $(\gamma(u) : u > 0)$ if

$$\liminf_{u\to\infty} \frac{\log V(Z_u)}{\log\gamma(u)} \geq 2.$$

This is equivalent to the condition that

$$\lim_{u\to\infty} \frac{\log EZ_u^2}{\log\gamma(u)}$$

exists and equals 2.

As is well known, relative error of a rv is defined as the ratio of its standard deviation and its mean. The collection of estimators $(Z_u : u > 0)$ has the bounded relative error property if

$$\limsup_{u\to\infty} \frac{V(Z_u)}{\gamma(u)^2} < \infty.$$

This is equivalent to the condition

$$\limsup_{u\to\infty} \frac{EZ_u^2}{\gamma(u)^2} < \infty.$$

The collection of estimators $(Z_u : u > 0)$ have asymptotically zero relative error if

$$\limsup_{u\to\infty} \frac{V(Z_u)}{\gamma(u)^2} = 0.$$

2.1 Conditioning algorithm for $P(S_n > u)$

We now outline the conditioning algorithm as proposed by Asmussen and Binswanger (1997) for estimating $P(S_n > u)$.

1. Generate samples of $X_1, X_2, \ldots, X_n$.

2. From these samples determine the order statistics $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$.

3. Discard the order statistic $X_{(n)}$. From the remaining generated data compute the conditional probability

$$Z_u = P(X > u - \sum_{i=1}^{n-1} X_{(i)} | X > X_{(n-1)}).$$

   This is a sample output from the simulation.

4. Average of many such independent samples provides an estimator of $P(S_n > u)$.

Asmussen and Binswanger (1997) show that when $(X_i : i \leq n)$ have an identical regularly varying distribution the above estimator Z_u is asymptotically efficient.

2.2 Importance sampling algorithms for $P(S_n > u)$

Recall that under P, $(X_i : i \leq n)$ are i.i.d. with pdf f. Suppose that under the importance sampling measure P^*, $(X_i : i \leq n)$ remain i.i.d. and each X_i has pdf f^*. Furthermore, $f^*(x) > 0$ whenever $f(x) > 0$. Then, the average of independent samples of

$$\frac{f(X_1) \ldots f(X_n)}{f^*(X_1) \ldots f^*(X_n)} I(S_n > u)$$

generated using P^* provides an unbiased estimator of $P(S_n > u)$. (Here, and in the remaining paper $I(A)$ denotes the indicator of event A. Thus, it equals 1 if A occurs, and zero otherwise.)

Asmussen, Binswanger and Hojgaard (2000) propose an asymptotically efficient importance sampling estimator in this setting. Specifically, they propose an importance sampling density $f^*(x) \sim \frac{c}{x \log x}$, for an appropriate normalization constant c.

Juneja and Shahabuddin (2002) propose a hazard rate twisting based algorithm to asymptotically efficiently estimate $P(S_n > u)$ when $(X_i : i \leq n)$ are i.i.d. Some notation is needed for this. Let $\lambda(x) = \frac{f(x)}{\bar{F}(x)}$ and $\Lambda(x) = -\log \bar{F}(x)$ denote respectively, the hazard rate and hazard function of X_i. Then, $f(x) = \lambda(x) \exp(-\Lambda(x))$. Given functions $u : \Re^+ \to \Re^+$ and $v : \Re^+ \to \Re^+$, we say that $u(x)$ is $\Theta(v(x))$ if the relation $c_1 v(x) \leq u(x) \leq c_2 v(x)$ holds for some positive constants c_1 and c_2, for all x sufficiently large. Function $u(x)$ is $O(v(x))$ if $\limsup_{x\to\infty} \frac{u(x)}{v(x)} < \infty$, and it is $o(v(x))$ if $\limsup_{x\to\infty} \frac{u(x)}{v(x)} = 0$. Juneja and Shahabuddin (2002) show that the hazard rate twisted distribution with density function

$$f_\theta(x) = \lambda(x)(1-\theta) \exp(-(1-\theta)\Lambda(x))$$

and $\theta = 1 - 1/g(u)$, asymptotically efficiently estimates $P(S_n > u)$ when the function $g(x)$ is $\Theta(\Lambda(x))$. Juneja and Shahabuddin (2002) also propose a modification 'delayed hazard rate twisting' to estimate $P(S_N > u)$ efficiently. We now discuss this as it proves useful in efficient estimation of $P(R_N(u))$.

For a positive constant x^*, suppose that under $\hat{P}_{\theta,x^*}$, rv $(X_i : i \geq 1)$ are i.i.d. with density function

$$\hat{f}_{\theta,x^*}(x) = f(x)$$

for $x \leq x^*$ and

$$\hat{f}_{\theta,x^*}(x) = \frac{\bar{F}(x^*)}{\bar{F}_\theta(x^*)} f_\theta(x)$$

for $x > x^*$ (here $\bar{F}_\theta$ denotes the tail cdf corresponding to f_θ). They show that for x^* an appropriate poly-log increasing function of u, $\hat{P}_{\theta,x^*}$ asymptotically efficiently estimates $P(S_N > u)$.

3 Efficient Estimation of Residual Probability

In this section we consider the problem of efficient estimation of the residual probability $P(R_N(u))$. This may be motivated from control variate viewpoint as follows: Note that naive estimation of $P(S_N > u)$ involves generating i.i.d. samples of $I(S_N > u)$. Since, in heavy-tailed settings, $I(\max_{i \leq N} X_i > u)$ is highly correlated with $I(S_N > u)$ and since $P(\max_{i \leq N} X_i > u)$ is easily computed, we may consider using $I(\max_{i \leq N} X_i > u) - P(\max_{i \leq N} X_i > u)$ as a control variate in estimating $P(S_N > u)$. If $P(N = n) = \rho^n(1-\rho)$ for $n = 0, 1, 2, \ldots$ and $0 < \rho < 1$, then

$$P(\max_{i \leq N} X_i > u) = \sum_{n=1}^{\infty} \rho^n (1-\rho)(1 - F(u)^n),$$

and this equals $\frac{\rho}{1-\rho F(u)} \bar{F}(u)$. Our estimator is the average of i.i.d. samples of

$$I(S_N > u) - b(I(\max_{i \leq N} X_i > u) - P(\max_{i \leq N} X_i > u)).$$

It is easy to see that the optimal b that minimizes the variance of these samples equals

$$\text{covariance}(I(S_N > u), I(\max_{i \leq N} X_i > u)) / \text{ variance}(I(\max_{i \leq N} X_i > u))$$

(see, e.g., Glasserman 2004) and this asymptotically converges to 1. Therefore, a good estimator of $P(S_N > u)$ is the average of i.i.d. samples of

$$P(\max_{i \leq N} X_i > u)) + I(S_N > u : \max_{i \leq N} X_i \leq u).$$

Hence we focus on efficient estimation of the residual probability $P(R_N(u))$. We first note that the previously proposed conditioning and importance sampling methods do not work very well in estimating $P(R_n(u))$. Our arguments can be easily extended to establish this for estimation of $P(R_N(u))$ as well. We then develop a sharp asymptotic and a combined importance sampling and conditioning estimator for $P(R_N(u))$ that has bounded relative error. Similar results for $P(R_n(u))$ essentially follow from this analysis. However, we directly address the latter problem in a more general setting in Section 4.

3.1 Conditioning

We now show that the conditioning algorithm as proposed by Asmussen and Binswanger (1997) may not be effective in estimating $P(R_n(u))$. For this and the next subsection, it suffices to keep in mind that under Assumption 1, $P(R_n(u))$ is $\Theta(f(u))$. This is proved later in Section 4. To keep the notation simple we consider the case of estimating $P(X_1+X_2>u, X_1, X_2 \leq u)$, i.e., $P(R_2(u))$ via conditioning when X_1 and X_2 have identical pdf

$$f(x) = \frac{\alpha-1}{x^\alpha}$$

for $x \geq 1$, and $f(x)=0$ otherwise. Recall that X has the same distribution as X_i. The conditioning algorithm is straightforward. It involves generating a sample of $X_{(1)}$ as discussed in Section 2.1. The resulting estimator then is

$$Z_u = P(u - X_{(1)} < X \leq u | X > X_{(1)}) = \frac{P(\max(X_{(1)}, u - X_{(1)}) < X \leq u)}{P(X > X_{(1)})}.$$

We now argue that

$$E(Z_u^2) \geq \frac{\text{constant}}{u^{2(\alpha-1)}}, \tag{1}$$

so that this algorithm does not asymptotically efficiently estimate $R_2(u)$. Since $P(S_2>u) \sim \frac{2}{u^{\alpha-1}}$, it then follows that the resulting estimator $P(\max(X_1,X_2)>u)+Z_u$ for $P(S_2>u)$ does not have a relative error that converges to zero as $u \to \infty$.

To see (1), note that

$$E(Z_u^2 I(X_{(1)} \in [u/4, u/3])) = \int_{x \in [u/4,u/3]} \frac{P(u-x<X<u)}{P(X>x)} f_{X_{(1)}}(x)dx,$$

where $f_{X_{(1)}}(\cdot)$ denotes the pdf of $X_{(1)}$. This is lower bounded by

$$P(X_{(1)} \in [u/4, u/3]) \frac{\bar{F}(3u/4) - \bar{F}(u)}{\bar{F}(u/3)}.$$

The result now follows as $P(X_{(1)} \in [u/4, u/3])$ equals

$$P(X_1 \in [u/4, u/3])^2 + 2P(X_1 \in [u/4, u/3]) P(X_1 \geq u/3),$$

and this in turn equals $c/u^{2(\alpha-1)}$ for an appropriate positive constant c. Furthermore,

$$\frac{\bar{F}(3u/4) - \bar{F}(u)}{\bar{F}(u/3)}$$

converges to a positive constant since $\bar{F}$ is regularly varying.

3.2 Importance Sampling

We now show that even importance sampling may not be effective in efficiently estimating $P(R_n(u))$. Again, we focus on $P(R_2(u))$ and consider the case where X_1 and X_2 have identical pdf f specified in Section 3.1. Suppose that under importance sampling, the pdf of each X_i is f^*, and the resulting probability measure is P^*. Then, the simulation output involves average of samples of

$$\hat{Z}_u = \frac{f(X_1)f(X_2)}{f^*(X_1)f^*(X_2)} I(X_1+X_2>u, X_1, X_2 \leq u).$$

Let P_K denote a probability measure under which $(X_i : i \leq n)$ are i.i.d. and have pdf

$$k(x) = \frac{f^2(x)}{\tilde{c} f^*(x)}$$

with an appropriate normalization constant $\tilde{c}$. If f^* is chosen so that k is regularly varying, then, in view of Proposition 2 proved later, it is easily checked that

$$E_{P^*} \hat{Z}_u^{\,2} = \tilde{c}^2 P_K(X_1 + X_2 > u, X_1, X_2 \leq u) \sim 2\tilde{c}^2 E_{P_K}(X_1) k(u).$$

(Here and elsewhere, the subscript to the expectation operator denotes the associated probability measure.) Thus, to minimize this, the pdf f^* needs to be chosen to have a sufficiently heavy tail so that $k(u)$ is of small order of magnitude. Note that even if we select $f^*(x) \sim \frac{c}{x \log x}$ (for some constant c), amongst the heaviest tails feasible, we get

$$k(u) = \Theta(\frac{\log u}{u^{2\alpha-1}}),$$

so that $P(X_1+X_2 > u, X_1, X_2 \leq u) = \Theta(u^{-\alpha})$ is not asymptotically efficiently estimated. Although interestingly, since,

$$P(S_2 > u) = \Theta(u^{-(\alpha-1)}),$$

even in this setting, the unbiased estimator $P(\max(X_1, X_2) > u) + \hat{Z}_u$ for $P(S_2 > u)$ has a relative error that converges to zero as $u \to \infty$.

3.3 Estimating $P(R_N(u))$

We first develop a sharp asymptotic for $P(R_N(u))$ where N is independent of $(X_i : i \geq 1)$ and is geometrically distributed with parameter $\rho \in (0,1)$. Recall that $P(N = n) = \rho^n(1-\rho)$ for $n = 0, 1, 2, \ldots$

Proposition 1 *Under Assumption 1,*

$$P(R_N(u)) \sim 2EX \frac{\rho^2}{(1-\rho)^2} f(u).$$

Proof: Note that

$$P(R_N(u)) = \sum_{n=2}^{\infty} \rho^n (1-\rho) P(R_n(u)).$$

We can select $g(u) = \Theta(\log u)$ so that $\rho^{g(u)}$ is $o(f(u))$. Therefore, we focus on

$$\sum_{n=2}^{g(u)-1} \rho^n (1-\rho) P(R_n(u)),$$

Select $\beta < 1$ such that $\bar{F}(u^\beta)^2$ is $o(f(u))$, i.e., $\beta \in (\frac{\alpha}{2(\alpha-1)}, 1)$. Recall that for each n, $X_{(n-1)}$ denotes the order statistic $n-1$ of $(X_1, \ldots, X_n)$, i.e., the second largest element in this vector of n random variables. Consider the decomposition

$$\sum_{n=2}^{g(u)-1} \rho^n (1-\rho) P(R_n(u)) = \sum_{n=2}^{g(u)-1} \rho^n (1-\rho) P(R_n, X_{(n-1)} < u^\beta) + \sum_{n=2}^{g(u)-1} \rho^n (1-\rho) P(R_n X_{(n-1)} \geq u^\beta).$$

First we argue that

$$\sum_{n=2}^{g(u)-1} \rho^n (1-\rho) P(R_n(u), X_{(n-1)} \geq u^\beta)$$

is $o(f(u))$. This follows from the observation that

$$P(R_n(u), X_{(n-1)} \geq u^\beta) \leq P(X_{(n-1)} \geq u^\beta) \leq \frac{n(n-1)}{2} \bar{F}(u^\beta)^2,$$

since $\bar{F}(u^\beta)^2$ is $o(f(u))$ and

$$\sum_{n=2}^{\infty} \rho^n (1-\rho) \frac{n(n-1)}{2}$$

is finite.

Now consider

$$\sum_{n=2}^{g(u)-1} \rho^n (1-\rho) P(R_n(u), X_{(n-1)} < u^\beta).$$

This may be re-expressed as

$$(1-\rho) f(u) \sum_{n=2}^{g(u)-1} n\rho^n \int_{x_i \in [0,u^\beta), i \leq n-1} f(x_1) \dots f(x_{n-1}) \left(\frac{F(u) - F(u - \sum_{i=1}^{n-1} x_i)}{f(u)} \right) dx_1 \dots dx_{n-1}. \tag{2}$$

Note that $n\rho^n$ is bounded above for all n. Furthermore, since f is eventually decreasing, for u sufficiently large

$$\sum_{i=1}^{n-1} x_i \leq \left(\frac{F(u) - F(u - \sum_{i=1}^{n-1} x_i)}{f(u)} \right) \leq \left(\frac{f(u - \sum_{i=1}^{n-1} x_i)}{f(u)} \right) \sum_{i=1}^{n-1} x_i.$$

Since f has a regularly varying tail, it follows that $\lim_{x\to\infty} \frac{f(x-y)}{f(x)} = 1$ for all y (see, e.g., Embrechts et. al. 1997). Hence,

$$\frac{F(u) - F(u - \sum_{i=1}^{n-1} x_i)}{f(u)} \to \sum_{i=1}^{n-1} x_i$$

as $u \to \infty$. Note that for $n \leq g(u)$, the left hand side is upper bounded by

$$\sum_{i=1}^{n-1} x_i \left(\frac{f(u - g(u)u^\beta)}{f(u)} \right) \leq \sum_{i=1}^{n-1} x_i \times \text{constant}.$$

(since, $\frac{f(u-g(u)u^\beta)}{f(u)} \to 1$ as $u \to \infty$.) Hence, by the dominated convergence theorem (2) is asymptotically similar to

$$f(u)(1-\rho) EX \sum_{n=2}^{\infty} n(n-1)\rho^n.$$

From this the result follows. □

3.4 Combined conditioning and importance sampling

Select a non-negative integer valued function $h(u) = \Theta(\log u)$ so that $\rho^{h(u)}$ is $o(f(u)^2)$. Consider the pdf

$$k_{\theta,x^*}(x) = \frac{f^2(x)}{c_{\theta,x^*} \hat{f}_{\theta,x^*}(x)}$$

for $x \geq 0$ (and zero elsewhere), where c_{θ,x^*} is the normalization constant $\int_{x=0}^{\infty} \frac{f^2(x)}{\hat{f}_{\theta,x^*}(x)} dx$. Let $\bar{K}_{\theta,x^*}$ denote the associated tail distribution function. It is easy to see that

$$\bar{K}_{\theta,x^*}(x) = \frac{\tilde{L}(x)}{x^{(\alpha-1)(1+\theta)}}$$

for some slowly varying function $\tilde{L}$. Through simple calculations, c_{θ,x^*} can be seen to equal

$$F(x^*) + \frac{\bar{F}(x^*)}{1-\theta^2}.$$

Note that for a given $\theta < 1$ and $0 < \rho < 1$, we can always find an x^* so that $c_{\theta,x^*} < 1/\rho$. We now consider the following simulation algorithm to estimate $P(R_N(u))$. Here we assume that x^* is sufficiently large so that $c_{\theta,x^*} < 1/\rho$.

1. Generate a sample N of a geometrically distributed random variable with parameter ρ.

2. If $N > h(u)$, generate samples of $(X_1, \ldots, X_N)$ using the original distribution of X_i's. The unbiased sample estimator of $P(R_N(u))$ in this case equals $I(R_N(u))$.

3. If $N \leq h(u)$, generate samples $(X_1, \ldots, X_N)$ using the probability associated with delayed hazard rate twisted pdf $\hat{f}_{\theta,x^*}$. From these samples determine the order statistics $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(N)}$.

4. Discard the order statistic $X_{(N)}$. From the remaining generated data compute the conditional probability

$$P(u - \sum_{i=1}^{N-1} X_{(i)} < X < u | X \geq X_{(N-1)}) = \frac{P(\max(u - \sum_{i=1}^{N-1} X_{(i)}, X_{(N-1)}) < X < u)}{P(X \geq X_{(N-1)})}.$$

   (This equals zero if $X_{(N-1)} \geq u$.) Note that along the set $\{N = n\}$, the probability density of observing $(X_{(i)} : i \leq n-1)$ under the original measure equals

$$n! \left(\prod_{i \leq n-1} f(X_{(i)}) \right) f(X_{(n-1)}) \bar{F}(X_{(n-1)}).$$

   It equals $n! \left(\prod_{i \leq n-1} \hat{f}_{\theta,x^*}(X_{(i)}) \right) \hat{f}_{\theta,x^*}(X_{(n-1)}) \bar{\hat{F}}_{\theta,x^*}(X_{(n-1)})$ under the probability measure $\hat{P}_{\theta,x^*}$. Hence, the simulation output $Z_1(u)$ equals

$$\frac{\left(\prod_{i \leq N-1} f(X_{(i)}) \right) \bar{F}(X_{(N-1)})}{\left(\prod_{i \leq N-1} \hat{f}_{\theta,x^*}(X_{(i)}) \right) \bar{\hat{F}}_{\theta,x^*}(X_{(N-1)})} P(u - \sum_{i=1}^{N-1} X_{(i)} < X < u | X \geq X_{(N-1)}).$$

5. Average of many such independent samples provides an estimator of $P(R_N(u))$.

Theorem 1 *Under Assumption 1, for $\theta \in (\frac{1}{\alpha-1}, 1)$, the estimator $Z_1(u)$ has the bounded relative error property, i.e.,*

$$\limsup_{u\to\infty} \frac{E_{\hat{P}_{\theta,x^*}} Z_1(u)^2}{f(u)^2} < \infty.$$

The proof is given in the appendix.

4 Heterogeneous Component Distributions

We now consider the probability $P(R_n(u))$ when each X_i is allowed to have a different pdf. As mentioned earlier, this is useful in many applications including estimating tail probabilities of stochastic activity networks. We make the following assumption for this section:

Assumption 2 *For each $i \le n$, the pdf f_i of X_i exists and is regularly varying, i.e., it has the form*

$$f_i(x) = L_i(x)/x^{\alpha_i}$$

for $x \ge 0$ (it is zero otherwise), $\alpha_i > 2$, where L_i is a slowly varying function. Furthermore, the pdf is eventually non-increasing.

We append the subscript i to all the remaining previously introduced notation to generalize to this setting. Thus, for instance, F_i denote the cdf of X_i, $\bar{F}_i$ denotes its tail cdf and $f_{\theta,i}$ denotes the pdf obtained by hazard rate twisting the pdf f_i by amount θ.

Proposition 2 *Under Assumption 2,*

$$P(R_n(u)) \sim \sum_{i=1}^{n} f_i(u) \left(\sum_{j\neq i} EX_j \right). \tag{3}$$

We need some definitions before proving Proposition 2. For each i,

$$A_i = \{X_i > u/n, \max_{j\neq i} X_j \le u/n\} \cap R_n(u)\}.$$

Furthermore, for each $i \neq j$,

$$A_{ij} = \{X_i > u/n, X_j > u/n\}.$$

Proof: Note that,

$$\sum_{i\le n} P(A_i) \le P(R_n(u)) \le \sum_{i\le n} P(A_i) + \sum_{i\neq j} P(A_{ij}).$$

Further note that

$$P(A_{ij}) \sim L_i(u/n)L_j(u/n)n^{\alpha_i+\alpha_j-2}/u^{\alpha_i+\alpha_j-2},$$

and this is $o(\sum_{k=1}^{n} f_k(u))$ for each i, j. Hence, to prove the result, it suffices to show that

$$P(A_i) \sim f_i(u) \left(\sum_{j\neq i} EX_j \right). \tag{4}$$

For notational simplicity, we establish the above for $P(A_n)$ and the rest then follows. Note that

$$P(A_n) = f_n(u) \int_{x_i \le u/n, i \le n-1} f_1(x_1) \dots f_{n-1}(x_{n-1}) \left(\frac{F_n(u) - F_n(u - \sum_{i=1}^{n-1} x_i)}{f_n(u)} \right) dx_1 \dots dx_{n-1}. \quad (5)$$

As in the proof of Proposition 1, we have

$$I(x_i \le u/n, i \le n-1) \frac{F_n(u) - F_n(u - \sum_{i=1}^{n-1} x_i)}{f_n(u)} \to \sum_{i=1}^{n-1} x_i$$

as $u \to \infty$. Furthermore, the LHS is upper bounded by $\frac{f_n(u/n)}{f_n(u)} \sum_{i=1}^{n-1} x_i$ which in turn is upper bounded by $n^\alpha(1+\epsilon)\sum_{i=1}^{n-1} x_i$ for $\epsilon > 0$ and u sufficiently large. Therefore, by the dominated convergence theorem, (4) follows from (5). □

4.1 Combined conditioning and importance sampling

We now show that by appropriately combining conditioning with importance sampling, an estimator for $P(R_n(u))$ can be developed that has a bounded relative error property. Let $f_{\theta,i}(\cdot)$ denote the pdf obtained by hazard rate twisting the pdf $f_i(\cdot)$ by an amount θ. Let P_θ denote the associated probability measure.

The proposed combined algorithm is as follows:

1. Generate independent samples of $X_1, X_2, \dots, X_n$ using the hazard rate twisted pdfs $(f_{\theta,i}(\cdot) : i \le n)$.

2. From these samples determine the order statistics $X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$.

3. Let k denote the index associated with $X_{(n)}$ (i.e., $X_k = X_{(n)}$). Let j denote the index associated with $X_{(n-1)}$. Discard the order statistic $X_{(n)}$. From the remaining generated data compute the conditional probability

$$P(u - \sum_{i=1}^{n-1} X_{(i)} < X_k < u | X_k > X_{(n-1)}).$$

Therefore, an unbiased sample output from this simulation $Z_2(u)$ equals

$$\frac{\left(\prod_{i \ne j,k} f_i(X_i)\right) f_j(X_j)\bar{F}_j(X_j)}{\left(\prod_{i \ne j,k} f_{\theta,i}(X_i)\right) f_{\theta,j}(X_j)\bar{F}_{\theta,j}(X_j)}$$

$$\times\ P(u - \sum_{i=1}^{n-1} X_{(i)} < X_k < u | X_k > X_{(n-1)}).$$

4. Average of many such independent samples provides an estimator of $P(R_n(u))$.

Let $\alpha_{min} = \min_{i \le n} \alpha_i$.

Theorem 2 *Under Assumption 2, for $\theta \in [\frac{1}{(\alpha_{min}-1)}, 1)$, the estimator $Z_2(u)$ from the above algorithm has bounded relative error so that*

$$E_{P_\theta} Z_2(u)^2 = \Theta(P(R_n(u))^2) = \Theta[\sum_{i=1}^{n} f_i(u)^2].$$

The proof of Theorem is given in the appendix.

Remark 1 It is easy to generalize the results in Proposition 2 and Theorem 2 to cover the case where some of the X_i's are light tailed or have tails decaying at a faster than polynomial rate (e.g., Weibull distribution for any value of shape parameter). Since such random variables have a decay rate of tail probability that is faster than polynomial, they contribute negligibly to the tail asymptotic (3). Similarly, in the combined conditioning and importance sampling technique, even if no importance sampling is performed on these random variables, the effect on the second moment of the resultant estimator is insignificant. This is easily formalized. We avoid this to maintain the simplicity of the analysis.

4.2 Stochastic activity networks

Again consider independent random variables $(X_1, \ldots, X_n)$ and let $(\mathcal{P}_j : j \le \tau)$ denote a collection of τ subsets of $\{1, 2, \ldots, n\}$ such that $\cup_{i=1}^{\tau} \mathcal{P}_j = \{1, 2, \ldots, n\}$. Set

$$T = \max_{j \le \tau} \sum_{i \in \mathcal{P}_j} X_i.$$

Note that the sets $(\mathcal{P}_j : j \le \tau)$ need not be disjoint so that the associated sums may be dependent.

One application of such probabilities arises in stochastic activity or PERT networks (Project Evaluation and Review Technique; see e.g., Elmagharaby 1977, Adalakha and Kulkarni 1989). These networks consist of many tasks or activities with stochastic durations that need to be executed under specified precedence constraints. The random variables $(X_1, \ldots, X_n)$ may be used to model these task durations and the subsets $(\mathcal{P}_j : j \le \tau)$ may be chosen to reflect the precedence constraints so that T denotes the overall duration of the project. Our interest then is in efficient estimation of the probability of large delay, $P(T > u)$, in such networks. These probabilities are of enormous interest in project management as the costs associated with large delays can be prohibitive.

4.2.1 Asymptotic result for residual probability

As before, we may re-express $P(T > u)$ as

$$P(\max_{i \le n} X_i > u) + P(T > u, \max_{i \le n} X_i \le u).$$

Denote by $TR_n(u)$ the event $\{T > u, \max_{i \le n} X_i \le u\}$. In this section, we focus on efficient estimation of $P(TR_n(u))$. Juneja, Karandikar and Shahabuddin (2007) show that under Assumption 2, $P(T > u) \sim P(S_n > u)$. The following proposition is also easily seen:

Proposition 3 *Under Assumption 2,*

$$P(TR_n(u)) = \Theta(P[R_n(u)]).$$

Proof: The proof follows from the following two observations and Proposition 2:

$$P(TR_n(u)) \le P(R_n(u)),$$

and

$$\max_{j \le \tau} P(\sum_{i \in \mathcal{P}_j} X_i > u, \max_{i \in \mathcal{P}_j} X_i \le u) \le P(TR_n(u)).$$

□

4.2.2 Combined conditioning and importance sampling

The combined algorithm to efficiently estimate $P(TR_n(u))$ is straightforward.

1. Generate independent samples of $X_1, X_2, \ldots, X_n$ using the hazard rate twisted pdfs $(f_{\theta,i}(\cdot) : i \leq n)$.

2. From these samples determine the order statistics $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$.

3. Let k denote the index associated with $X_{(n)}$ (i.e., $X_k = X_{(n)}$). Let j denote the index associated with $X_{(n-1)}$. Discard the order statistic $X_{(n)}$. From the remaining generated data compute the conditional probability

$$P(u - Y < X_k < u | X_k > X_{(n-1)})$$

where Y denotes the maximum of the sums of generated times along the paths that contain activity k. Clearly $Y \leq \sum_{i=1}^{n-1} X_{(i)}$. An unbiased sample output from this simulation equals

$$Z_3(u) = \frac{\left(\prod_{i \neq j,k} f_i(X_i)\right) f_j(X_j) \bar{F}_j(X_j)}{\left(\prod_{i \neq j,k} f_{\theta,i}(X_i)\right) f_{\theta,j}(X_j) \bar{F}_{\theta,j}(X_j)} P(u - Y < X_k < u | X_k > X_{(n-1)}).$$

4. Average of many such independent samples provides an estimator of $P(TR_n(u))$.

Theorem 3 *Under Assumption 2, for $\theta \in (\frac{1}{(\alpha_{min}-1)}, 1)$,*

$$E_{P_\theta} Z_3(u)^2 = \Theta(P(TR_n(u))^2),$$

i.e., the combined algorithm estimator for $P(TR_n(u))$ has bounded relative error.

The proof simply follows from Theorem 2 and Proposition 3 by noting that $Y \leq \sum_{i=1}^{n-1} X_{(i)}$ implies that $Z_3(u) \leq Z_2(u)$.

5 Numerical Results

We conduct a small experiment where we estimate $P(S_n > u)$ using hazard rate twisting and estimate $P(R_n(u))$ and hence $P(S_n > u)$ using the proposed combined algorithm. We consider the case where $(X_i : i \leq n)$ are i.i.d. While, in the existing literature, there are algorithms that perform better than hazard rate twisting, It remains worthwhile to compare that with the proposed combined algorithm simply because the proposed algorithm essentially builds upon hazard rate twisting idea and combines it with conditioning. A better algorithm to estimate $P(S_n > u)$ can also be suitably modified to estimate $P(R_n(u))$ by combining it with conditioning.

In our experiment we set $n = 5$. The pdf of $X_i's$ is

$$f(x) = \frac{2}{x^3}$$

for $x \geq 1$ and is zero otherwise. Hence, $\bar{F}(x) = 1/x^2$ for $x \geq 1$. In both the cases the number of trials is set to 100,000. In hazard rate twisting, θ is selected as suggested in Juneja and Shahabuddin (2002) to equal $1 - n/\Lambda(u)$ where Λ denotes the hazard function of X_i and $\Lambda(x) = 2\log(x)$. For

u	Est. $P(S_n > u)$	VR
30	$1.07 \times 10^{-2} \pm 2.92\%$	4.2
50	$2.79 \times 10^{-3} \pm 4.02\%$	8.51
70	$1.28 \times 10^{-3} \pm 4.76\%$	13.27
90	$7.45 \times 10^{-3} \pm 5.24\%$	18.75

Table 1: Variance reduction over naive simulation using hazard rate twisting based importance sampling. VR denotes the ratio of the estimated variance using naive simulation over estimated variance using importance sampling.

u	$P(\max_{i\le n} X_i > u)$	Est. $P(R_n(u))$	Est. $P(S_n > u)$	VR $P(R_n(u))$	VR $P(S_n > u)$
30	5.54×10^{-3}	$4.93 \times 10^{-3} \pm 0.93\%$	$1.05 \times 10^{-2} \pm 0.45\%$	89.32	189.7
50	2.00×10^{-3}	$8.77 \times 10^{-4} \pm 0.78\%$	$2.88 \times 10^{-3} \pm 0.24\%$	715.5	2345
70	1.02×10^{-3}	$2.95 \times 10^{-4} \pm 0.71\%$	$1.31 \times 10^{-3} \pm 0.16\%$	2596	11,582
90	6.17×10^{-4}	$1.31 \times 10^{-4} \pm 0.64\%$	$7.48 \times 10^{-4} \pm 0.11\%$	7120	40,634

Table 2: Variance reduction over naive simulation using combined importance sampling and conditioning. VR $P(R_n(u))$ denotes the ratio of the estimated variance using naive simulation over estimated variance using the combined approach in estimating $P(R_n(u))$. VR $P(S_n > u)$ denotes the same ratio in estimating $P(S_n > u)$

simplicity, we select the same θ in the combined algorithm as well. We get similar orders of variance reduction even with other values of $\theta \ge 1/(\alpha - 1)$. The results of the simulation are reported in Tables 1 and 2. In these tables the variance under naive simulation is estimated from the estimation of the probability using the variance reduction technique. (Recall that the variance of an indicator function of event A equals $P(A)(1 - P(A))$.) As can be seen, the proposed algorithm provides dramatic orders of magnitude improvement in estimating $P(R_n(u))$ as well as $P(S_n > u)$.

6 Appendix

6.1 Proof of Theorem 1

The following observation is useful in proving Theorem 1:

$$\bar{K}_{\theta,x^*}(x) \ge \frac{\bar{F}(x)^2}{c^\theta \bar{\hat{F}}_{\theta,x^*}(x)} \tag{6}$$

for all x. To see this note that

$$\bar{K}_{\theta,x^*}(x) = \frac{\bar{\hat{F}}_{\theta,x^*}(x)}{c_{\theta,x^*}} \int_{(x,\infty)} \frac{f(y)^2}{\hat{f}_{\theta,x^*}(y)^2} \frac{\hat{f}_{\theta,x^*}(y)}{\bar{\hat{F}}_{\theta,x^*}(x)} dy.$$

By Jensen's inequality, the RHS is lower bounded by

$$\frac{\bar{\hat{F}}_{\theta,x^*}(x)}{c_{\theta,x^*}} \left(\int_{(x,\infty)} \frac{f(y)}{\hat{f}_{\theta,x^*}(y)} \frac{\hat{f}_{\theta,x^*}(y)}{\bar{\hat{F}}_{\theta,x^*}(x)} dy \right)^2.$$

From this (6) follows.

Proof of Theorem 1: We perform the proof in two steps.
Step 1: We first show that

$$\frac{E_{\hat{P}_{\theta,x^*}}[Z_1(u)^2 I(X_{(N-1)} < u^\beta, N < h(u))]}{f(u)^2} \tag{7}$$

is $O(1)$ for any $\beta \in (0,1)$. To see this, first consider

$$\frac{E_{\hat{P}_{\theta,x^*}}[Z_1(u)^2 I(X_{(n-1)} < u^\beta)]}{f(u)^2}.$$

For u sufficiently large, this equals

$$n(n-1)\int_{A,x_{n-1}<u^\beta} \frac{\left(\prod_{i\le n-1} f(x_i)^2\right)\bar{F}(x_{n-1})^2}{\left(\prod_{i\le n-1}\hat{f}_{\theta,x^*}(x_i)\right)\hat{\bar{F}}_{\theta,x^*}(x_{n-1})} \frac{P(u-\sum_{i\le n-1} x_i < X < u | X > x_{n-1})^2}{f(u)^2} dx_1\ldots dx_{n-1}.$$

where $A = \{(x_1,\ldots,x_n) : x_n = \max_{i\le n} x_i, x_{n-1} = \max_{i\le n-1} x_i\}$.

Using (6), the above may be upper bounded by

$$c^n_{\theta,x^*} n(n-1)\int_{A,x_{n-1}<u^\beta}\left(\prod_{i\le n-1} k_{\theta,x^*}(x_i)\right)\bar{K}_{\theta,x^*}(x_{n-1})\frac{P(u-\sum_{i\le n-1} x_i < X < u)^2}{f(u)^2\bar{F}(x_{n-1})^2} dx_1\ldots dx_{n-1}.$$

This in turn may be upper bounded by

$$c^n_{\theta,x^*} n(n-1)\int_{x_{n-1}<u^\beta} k_{\theta,x^*}(x_{n-1})\bar{K}_{\theta,x^*}(x_{n-1})\frac{P(u-(n-1)x_{n-1} < X < u)^2}{f(u)^2\bar{F}(x_{n-1})^2} dx_{n-1}.$$

Now for u sufficiently large, for $n \le h(u)$, $x_{n-1} < u^\beta$, the term

$$\frac{P(u-(n-1)x_{n-1} < X < u)^2}{f(u)^2\bar{F}(x_{n-1})^2}$$

is upper bounded by

$$\left(\frac{f(u-u^\beta h(u))}{f(u)}\right)^2\left(\frac{(n-1)x_{n-1}}{\bar{F}(x_{n-1})}\right)^2.$$

Since, $\beta < 1$, and f is regularly varying, $\frac{f(u-u^\beta h(u))}{f(u)} \to 1$ as $u \to \infty$. Therefore, for $\epsilon > 0$, there exists u_ϵ so that for $u > u_\epsilon$,

$$I(n \le h(u), x_{n-1} < u^\beta)\frac{P(u-(n-1)x_{n-1} < X < u)^2}{f(u)^2\bar{F}(x_{n-1})^2} \le \left(\frac{(n-1)x_{n-1}}{\bar{F}(x_{n-1})}\right)^2(1+\epsilon)$$

Recall that $\bar{F}(x) = \frac{L(x)}{x^{\alpha-1}}$, $k_{\theta,x^*}(x) = \frac{\tilde{L}(x)}{x^{1+(\alpha-1)(1+\theta)}}$, $\bar{K}_{\theta,x^*}(x) = \frac{\tilde{L}(x)}{x^{(\alpha-1)(1+\theta)}}$, where $\tilde{L}$ is a slowly varying function, and $\theta > \frac{1}{\alpha-1}$. Therefore,

$$\int_{x_{n-1}\ge 0}\left(\frac{x_{n-1}}{\bar{F}(x_{n-1})}\right)^2 k^\theta(x_{n-1})\bar{K}^\theta(x_{n-1})dx_{n-1}$$

is bounded by some positive constant H. In particular, (7) is upper bounded by

$$H(1+\epsilon)(1-\rho)\sum_{n=2}^{h(u)}\rho^n c_{\theta,x^*}^n n(n-1)^3$$

and hence is $O(1)$.

Step 2: We now show that

$$\frac{E_{\hat{P}_{\theta,x^*}}[Z_1(u)^2 I(X_{(N-1)} \geq u^\beta, N < h(u))]}{f(u)^2}$$

is $o(1)$. Note that $E_{\hat{P}_{\theta,x^*}}[Z_1(u)^2 I(X_{(n-1)} \geq u^\beta)]$ equals

$$n(n-1)\int_{A,x_{n-1}\geq u^\beta} \frac{\left(\prod_{i\leq n-1} f(x_i)^2\right)\bar{F}(x_{n-1})^2}{\left(\prod_{i\leq n-1}\hat{f}_{\theta,x^*}(x_i)\right)\hat{\bar{F}}_{\theta,x^*}(x_{n-1})}$$
$$\times P(u-\sum_{i\leq n-1} x_i < X < u | X > x_{n-1})^2 dx_1 \ldots dx_{n-1}.$$

This is upper bounded by

$$c_{\theta,x^*}^n n(n-1)\int_{A,x_{n-1}\geq u^\beta}\left(\prod_{i\leq n-1} k_{\theta,x^*}(x_i)\right)\bar{K}_{\theta,x^*}(x_{n-1})dx_1\ldots dx_{n-1}.$$

This in turn is dominated by,

$$n(n-1)c_{\theta,x^*}^n \bar{K}_{\theta,x^*}(u^\beta)^2.$$

Recall that

$$\bar{K}_{\theta,x^*}(u^\beta) = \frac{\tilde{L}(u^\beta)}{u^{\beta(\alpha-1)(1+\theta)}}$$

where $\tilde{L}$ is a slowly varying function. Since, $(\alpha-1)(1+\theta) > \alpha$, hence, for β sufficiently close to 1 $\bar{K}_{\theta,x^*}(u^\beta)$ is $o(u^{-\alpha})$. The result now follows as $\bar{K}_{\theta,x^*}(u^\beta)^2$ is $o(f(u)^2)$ and

$$(1-\rho)\sum_{n=2}^{h(u)}\rho^n c_{\theta,x^*}^n n(n-1)$$

is bounded by a constant.

□

6.2 Proof of Theorem 2

Some notation is useful for proof of Theorem 2. For each i, consider the pdf

$$k_{\theta,i}(x) = \frac{f_i(x)^2}{d_\theta f_{\theta,i}(x)}$$

where d_θ is the normalization constant that can be seen to equal $1/(1-\theta^2)$. Let $P_{K_{\theta,i}}$ be the associated probability measure. Then,

$$P_{K_{\theta,i}}(X_i > u/n) = \frac{\tilde{L}(u/n)n^{(1+\theta)(\alpha_i-1)}}{u^{(1+\theta)(\alpha_i-1)}}$$

for some slowly varying function $\tilde{L}$. Therefore, for $\theta > 1/(\alpha_{min} - 1)$,

$$1 + \theta > \frac{\alpha_{min}}{\alpha_{min} - 1} \geq \frac{\alpha_i}{\alpha_i - 1}.$$

Hence,

$$P_{K_{\theta,i}}(X_i > u/n) = o(u^{-\alpha_i}). \tag{8}$$

Recall that $A = \{(x_1, \ldots, x_n) : x_n = \max_{i \leq n} x_i, x_{n-1} = \max_{i \leq n-1} x_i\}$. For notational convenience, we let $I(A)$ denote the indicator function of the event

$$(X_1, \ldots, X_n) \in A.$$

Proof of Theorem 2: Note that $E_{P_\theta}[Z_2(u)^2 I(X_j \leq u/n) I(A)]$ equals

$$\int_{x_{n-1} \leq u/n, A} \frac{\prod_{i \leq n-1} f_i(x_i)^2}{\prod_{i \leq n-1} f_{\theta,i}(x_i)} \frac{\bar{F}_{n-1}(x_{n-1})^2}{\bar{F}_{\theta,n-1}(x_{n-1})} P(u - \sum_{i \leq n-1} x_i < X_n < u | X_{n-1} > x_{n-1})^2 dx_1 ... dx_{n-1}.$$

Using the fact that $\frac{\bar{F}_{n-1}(x)^2}{d_\theta \bar{F}_{\theta,n-1}(x)} \leq \bar{K}_{\theta,n-1}(x)$, this may be upper-bounded by

$$f_n(u)^2 {d_\theta}^n \int_{x_{n-1} \leq u/n, A} \prod_{i \leq n-1} k_{\theta,i}(x_i) \bar{K}_{\theta,n-1}(x_{n-1}) \frac{P(u - (n-1)x_{n-1} < X_n < u)^2}{f_n(u)^2 \bar{F}_{n-1}(x_{n-1})^2} dx_1 ... dx_{n-1},$$

which in turn may be upper bounded by

$$f_n(u)^2 {d_\theta}^n \int_{x_{n-1} \leq u/n} k_{\theta,n-1}(x_{n-1}) \bar{K}_{\theta,n-1}(x_{n-1}) \frac{P(u - (n-1)x_{n-1} < X_n < u)^2}{f_n(u)^2 \bar{F}_{n-1}(x_{n-1})^2} dx_{n-1},$$

Note that $\frac{P(u-(n-1)x_{n-1} < X_n < u)^2}{f_n(u)^2}$ is upper bounded by $\frac{f_n(u/n)^2}{f_n(u)^2}(n-1)x_{n-1}$. Also, $\frac{f_n(u/n)^2}{f_n(u)^2}$ converges to a constant as $u \to \infty$. Furthermore, as in the proof of Theorem 1,

$$(n-1)^2 \int_{x \in (0,\infty)} \frac{x^2}{\bar{F}_{n-1}(x)^2} k^\theta_{n-1}(x) \bar{K^\theta}_{n-1}(x) dx$$

is finite. Therefore, $E_{P_\theta}[Z_2(u)^2 I(X_{n-1} \leq u/n) I(A)]$ is $O(f_n(u)^2)$ and hence $E_{P_\theta}[Z_2(u)^2 I(X_j \leq u/n)]$ is $O([\sum_{i \leq n} f_i(u)]^2)$.

We conclude the proof by showing that $E_{P_\theta}[Z_2(u)^2 I(X_j > u/n) I(A)]$ is $o(u^{-(\alpha_{n-1} + \alpha_n)})$. To see this, note that $E_{P_\theta}[Z_2(u)^2 I(X_j > u/n) I(A)]$ may be upper bounded by

$${d_\theta}^n \int_{x_{n-1} > u/n, A} \prod_{i \leq n-1} k_{\theta,i}(x_i) \bar{K}_{\theta,n-1}(x_{n-1}) dx_1 ... dx_{n-1}.$$

This upper bounded by

$${d_\theta}^n P_{K_{\theta,n-1}}(X_{n-1} > u/n) P_{K_{\theta,n}}(X_n > u/n).$$

Recall that

$$P_{K_{\theta,i}}(X_i > u/n) = \frac{\tilde{L}_i(u/n) n^{(1+\theta)(\alpha_i - 1)}}{u^{(1+\theta)(\alpha_i - 1)}}$$

Therefore, for $\theta > 1/(\alpha_{min} - 1)$, $P_{K_{\theta,i}}(X_i > u/n)$ is $o(u^{-\alpha_i})$ and the result follows. □

Acknowledgement: *Some of the initial ideas in this paper arose in discussions with late Perwez Shahabuddin.*

REFERENCES

Adalakha, V. G. and V. G. Kulkarni, A Classified Bibliography of Research on Stochastic PERT Networks, INFOR, **27**, 3, 1989, 272-296.

Asmussen, S. 2000. *Ruin Probabilities,* World Scientific, London.

Asmussen, S. 2003. *Applied Probabilities and Queues*, 2nd edition. Springer-Verlag, New York.

Asmussen, S., and K. Binswanger. 1997. Simulation of Ruin Probabilities for Subexponential Claims. *ASTIN BULLETIN* **27**, 2, 297-318.

Asmussen, S., K. Binswanger, and B. Hojgaard. 2000. Rare Event Simulation for Heavy Tailed Distributions. *Bernoulli* **6** (2), 303-322.

Asmussen, S., Kroese, D.P. 2006. Improved Algorithms for Rare Event Simulation with Heavy Tails. Advances in Applied Probability 38 (2), 545-558.

Blanchet, J., and J. Liu. 2006. Efficient Simulation for Large Deviation Probabilities of Sums of Heavy-tailed Random Variables. *Proceedings of the 2006 WSC Conference*, 757-764.

Elmaghraby, S. E.. 1977. *Activity Networks: Project Planning and Control by Network Models*, Wiley, New York, N. Y.

Embrechts, P., C. Kluppelberg, and T. Mikosch 1997. *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin, Heidelberg.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, New York.

Feller, W. 1971. *An Introduction to Probability Theory and Its Applications,* Vol. II. Wiley, New York.

Hartinger, J. and D. Kortschak. 2006. On the efficiency of Asmussen-Kroese-estimator and its applications to stop-loss transforms. Proceedings of RESIM 2006 Conference held in Bamberg, Germany.

Juneja, S., R.L. Karandikar, and P. Shahabuddin. 2007. Asymptotics and Fast Simulation for Tail Probabilities of Maximum of Sums of Few Random Variables. To appear in *ACM TOMACS.*

Juneja, S., and P. Shahabuddin. 2002. Simulating Heavy Tailed Processes using Delayed Hazard Rate Twisting. *ACM TOMACS*, **12**, 94-118.

Juneja, S. and P. Shahabuddin. 2006. Rare-Event Simulation Techniques: An Introduction and Recent Advances. *Handbook in OR & MS, Vol. 13*, S. G. Henderson and B. L. Nelson (Eds.), Elsevier, 291-350.

Omey, E. 1994. On the Difference Between the Product and the Convolution Product of Distribution Functions. *Publications De L'Institut Mathematique* **55**, (69), 111-145.

Omey, E. 2002. On the Difference Between the Distribution Function of the Sum and the Maximum of Real Random Variables. *Publications De L'Institut Mathematique* **71**, (85), 63-77.