Asymptotics and Fast Simulation for Tail Probabilities of Maximum of Sums of Few Random Variables

S. JUNEJA
Tata Institute of Fundamental Research, Mumbai
R. L. KARANDIKAR
Indian Statistical Institute, Delhi
and
P. SHAHABUDDIN
Columbia University, New York

We derive tail asymptotics for the probability that the maximum of sums of a few random variables exceeds an increasing threshold, when the random variables may be light as well as heavy tailed. These probabilities arise in many applications including in PERT networks where our interest may be in measuring the probability of large project delays. We also develop provably asymptotically optimal importance sampling techniques to efficiently estimate these probabilities. In the light-tailed settings we show that an appropriate mixture of exponentially twisted distributions efficiently estimates these probabilities. As is well known, exponential twisting based methods are not applicable in the heavy-tailed settings. To remedy this, we develop techniques that rely on "asymptotic hazard rate twisting" and prove their effectiveness in both light and heavy-tailed settings. We show that in many cases the latter may have implementation advantages over exponential twisting based methods in the light-tailed settings. However, our experiments suggest that when easily implementable, the exponential twisting based methods significantly outperform asymptotic hazard rate twisting based methods.

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]: Probabilistic Algorithms (including Monte Carlo); I.6.1 [**Simulation and Modeling**]: Simulation Theory

General Terms: Algorithms, Performance, Theory

Additional Key Words and Phrases: PERT networks, importance sampling, rare event simulation, tail asymptotics

1. INTRODUCTION

Consider independent random variables $(X_1, \ldots, X_n)$ and let $(\mathcal{P}_j : j \leq \tau)$ 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

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

In this article, we develop asymptotics and importance sampling techniques for $P(T > u)$ as $u \to \infty$ for n fixed. When $\tau = 1$, $P(T > u)$ reduces to $P(S_n > u)$ where $S_n = \sum_{i=1}^{n} X_i$. Also note that the sets $(\mathcal{P}_j : j \leq \tau)$ need not be disjoint so that the associated sums may be dependent.

One application of such probabilities arises in stochastic PERT networks (Project Evaluation and Review Technique; see Elmaghraby [1977] and 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 $(\mathcal{P}_j : j \leq \tau)$ 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.

Many other applications fit our framework. For example, consider the performance evaluation problem faced by a composite web service provider. As the name suggests, the composite service providers provide a number of services or activities. As in a typical project, these activities may be performed under prespecified precedence constraints. A simple example is the case where a service provider is asked to provide an address of persons A and B and the shortest path connecting the two addresses. Another example may be a request to determine the credit rating of a customer where the service provider may need to contact many agencies to establish the customer's credit. Composite service providers in the Internet settings typically provide guarantees to customers against large delays (e.g., delay is less than 5 minutes 95% of times). Thus, accurate estimation of large delay probabilities is particularly important to them. Our framework may also be used to model grid computing networks where the overall computation consists of many sub-computational processes that may be executed in parallel in different processors but that need to satisfy certain prespecified precedence constraints. Similarly, we may also model transportation schedules in this framework, (e.g., for trains, planes or buses) where precedence constraints are induced by the requirement that a vehicle at a particular location may take off only after the arrival of certain other vehicles.

Note that when $\tau = n$ and each $\mathcal{P}_j = (1, 2, \ldots, j)$, then $T = \max_{k \leq n} S_k$ so that $P(T > u)$ corresponds to a probability of considerable interest in the applied probability literature.

In our analysis we allow activity durations (i.e., X_i's) to have both light and heavy-tailed distributions (a random variable is said to have a light-tailed distribution if its tail distribution function decays at least at an exponential rate; when the tail distribution function decays at a subexponential rate, under mild regularity conditions, the random variable is said to be subexponentially distributed or heavy tailed; precise definition in Section 2.2). Typically, light-tailed distributions may be used to model activity duration in project management or transportation settings (see Adalakha and Kulkarni [1989]). In the web-services and grid computing settings, heavy-tailed distributions may often be used (see, e.g., Crovella et al. [1998] and Leland et al. [1994]). We further assume that the duration of all the activities are mutually independent, that is, the random variables $(X_1, \ldots, X_n)$ are mutually independent.

When the underlying random variables are light tailed, importance sampling probability measure obtained by appropriate exponential twisting of the underlying distributions has been found to be useful for efficient simulation of many performance measures (see survey articles, e.g., Heidelberger [1995] and Juneja and Shahabuddin [2006]; this technique is reviewed in Section 2). In this article, we use and generalize some of these ideas to develop exponential twisting based asymptotically optimal importance sampling techniques to efficiently estimate $P(S_n > u)$ and $P(T > u)$ when the $(X_i : i \leq n)$ are light tailed (asymptotic optimality is a standard criteria for measuring effectiveness of importance sampling distribution and is reviewed in Section 2).

As is well known, an essential requirement to obtain an exponentially twisted distribution from the original distribution of a random variable (rv; rvs when plural) is that it be light tailed (more specifically, its moment generating function should exist in a neighborhood of the origin). When the distribution of the random variable does not satisfy this property (roughly speaking, this is true when the rv is subexponentially distributed) exponential twisting is not feasible. Juneja and Shahabuddin [2002] develop hazard rate twisting to efficiently estimate $P(S_n > u)$ when the $(X_i : i \leq n)$ are independent and identically distributed (iid), and have a subexponential distribution (we review this technique in Section 3.3). In this article, we introduce 'asymptotic hazard rate twisting' in light and heavy-tailed settings. This generalizes the hazard rate twisting proposed for heavy-tailed rvs. In many cases plain hazard rate twisting or exponential twisting (in light-tailed settings) may be difficult to implement, however it may be feasible to successfully implement asymptotic hazard rate twisting.

The specific contributions of this article include:

(1) We develop a logarithmic tail asymptotic for $P(T > u)$ when all the underlying activity durations X_i are light tailed, n is fixed and $u \to \infty$. Enroute, we develop logarithmic asymptotics for the probability $P(S_n > u)$ when the X_i's are light-tailed random variables. Here again, n is fixed and $u \to \infty$. While there exists enormous literature on logarithmic asymptotics for the

large deviations probability $P(S_n/n > u)$ as $n \to \infty$ and u is fixed (see, for example, Dembo and Zeitouni [1998] and Petrov [1975]), to the best of our knowledge, little is documented for the case that we consider (in Jureckova [1981], this problem is considered when $(X_i : i \leq n)$ are iid).

(2) We develop an exact asymptotic for $P(T > u)$ when one or more X_i are subexponentially distributed.

(3) We develop exponential twisting based asymptotically optimal importance sampling techniques to efficiently estimate $P(S_n > u)$ when $(X_i : i \leq n)$ are light-tailed. We note through a simple example that straightforward application of exponential twisting may not effectively estimate $P(T > u)$. We then show that an importance sampling distribution that is a convex combination of many appropriately exponentially twisted distributions, asymptotically optimally estimates this probability.

(4) We develop asymptotically optimal importance sampling techniques to estimate $P(T > u)$ (optimal as $u \to \infty$) using asymptotic hazard rate twisting in light and heavy-tailed settings.

We restrict our analysis to small networks. Thus, in our analysis of $P(T > u)$, n and hence τ remain fixed while $u \to \infty$. This is relevant as in many applications the number of activities may be small (e.g., for composite web services) or an aggregate level analysis may be conducted so that many activities are grouped under a single head resulting in a network consisting of a few aggregated activities. From technical viewpoint, it may be straightforward to generalize exponential-twisting-based methodologies to large n in light-tailed settings. The analysis in Sadowsky and Bucklew [1990] becomes useful here as they consider development of importance sampling techniques for $P(S_n/n > a)$ for $(X_i : i \leq n)$ iid and light tailed, fixed $a > EX_i$ and $n \to \infty$. However, when subexponential random variables are involved, effective importance sampling techniques for estimating $P(S_n/n > a)$ as $n \to \infty$ do not exist. Also, as we verify through experiments, the hazard twisting techniques discussed here may not be effective for large n. Therefore, to maintain focus in this paper we restrict our analysis to fixed n case.

In Section 2, we develop the mathematical background needed to analyze $P(T > u)$. In particular, we review the importance sampling technique and the classification of distributions as light and heavy tailed. In Section 3, we develop logarithmic asymptotics and exponential-twisting-based importance sampling techniques for $P(T > u)$ when all activities are light tailed. In Section 4, we introduce asymptotic hazard rate twisting and we again consider the problem of efficient estimation of $P(T > u)$ when all activities are light-tailed, albeit this time we use asymptotic hazard rate twisting. In Section 5, we develop exact asymptotics and asymptotic hazard rate twisting based importance sampling techniques for $P(T > u)$ when some of the activities are heavy tailed. Some practical implementation considerations related to the proposed importance sampling techniques are discussed in Section 6. Numerical experiments, verifying the efficacy of proposed fast simulation techniques are given in Section 7. Finally, in Section 8 we provide a brief conclusion and discuss some areas for further research. All proofs are relegated to the appendix.

2. MATHEMATICAL FRAMEWORK AND BACKGROUND

In this section, we first discuss the importance sampling approach in the context of estimating $P(T > u)$ and $P(S_n > u)$. We then classify the distributions as light and heavy tailed and illustrate the definitions through examples.

Let $(\Omega, \mathcal{F}, P)$ denote the underlying probability space on which the mutually independent rvs $(X_1, \ldots, X_n)$ are defined. It suffices to let $\mathcal{F}$ correspond to the σ algebra generated by $(X_1, \ldots, X_n)$.

2.1 Importance Sampling

Recall that naive estimation of $P(T > u)$ involves generating many independent samples of $(X_i : i \leq n)$ and hence of $I(T > u)$ via simulation and then taking their average (here $I(\cdot)$ denotes the indicator function). As is well known, the number of samples needed to get a fixed degree of relative accuracy (ratio of standard deviation and the mean of the estimate) is inversely proportional to $P(T > u)$ and hence becomes prohibitively large for large values of u. This motivates the need for importance sampling which we now briefly review: Let F_i denote the distribution function of X_i. To keep the notation simple, we assume that each X_i has a probability density function (pdf) given by f_i. Let P^* denote another probability measure on the space $(\Omega, \mathcal{F})$ under which each X_i has a distribution function F_i^* and a pdf f_i^*, such that P is absolutely continuous with respect to P^* (i.e., $f_i^*(x) > 0$ if $f_i(x) > 0$ almost everywhere), and the X_i's are mutually independent under P^* as well. Importance sampling under the measure P^* involves generating samples of $(X_1, X_2, \ldots, X_n)$ using P^*. Then, an unbiased estimator of $P(T > u)$ equals $I(T > u)L$ where

$$L = \frac{f_1(X_1)}{f_1^*(X_1)} \frac{f_2(X_2)}{f_2^*(X_2)} \cdots \frac{f_n(X_n)}{f_n^*(X_n)} \ a.s.,$$

and is referred to as the likelihood ratio. The average of many independent samples of $I(T > u)L$ provides an unbiased estimator of $P(T > u)$. For any probability $\tilde{P}$, let $E_{\tilde{P}}$ denote the associated expectation operator. Note that $E_{P^*}[L^2 I(T > u)] \geq (E_{P^*}[LI(T > u)])^2 = P(T > u)^2$ and hence

$$\limsup_{u \to \infty} \frac{\log E_{P^*}[L^2 I(T > u)]}{\log P(T > u)} \leq 2.$$

Such a P^* is defined to be an asymptotically optimal estimator of $P(T > u)$ iff:

$$\lim_{u \to \infty} \frac{\log E_{P^*}[L^2 I(T > u)]}{\log P(T > u)} = 2. \tag{1}$$

The discussion above holds for $P(S_n > u)$ as well with T replaced by S_n.

2.2 Classification of Distributions Based on Tail Behavior

2.2.1 *Light-Tailed and Superexponential Distributions.* Consider a rv X with distribution function F and pdf f. Let $\bar{F}$ denote its tail distribution function so that $\bar{F}(x) = 1 - F(x)$. Let Λ denote its hazard function, that is, $\Lambda(x) = -\log \bar{F}(x)$. Since $\bar{F}$ is nonincreasing and $\bar{F}(x) \to 0$ as $x \to \infty$, it follows

that Λ is non-decreasing and $\Lambda(x) \to \infty$ as $x \to \infty$. We say that X has a *light-tailed distribution* if there exists a positive extended real number $\lambda \in (0, \infty]$ and an $\alpha \geq 1$ such that

$$\frac{\Lambda(x)}{x^\alpha} \to \lambda.$$

We say that X is *superexponentially distributed* if above holds with $\alpha > 1$. In this case, we say that its tail distribution decays at a superexponential rate. To keep the exposition simple, we do not consider distributions where the above limit does not exist.

Two functions f and g are said to be asymptotically similar if $\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1$. This is denoted by $f(x) \sim g(x)$. The following examples illustrate some light-tailed distributions:

Example 2.1.　Suppose that X is Gamma(n, λ) distributed. Its pdf

$$f(x) = \frac{\lambda^n x^{n-1} \exp[-\lambda x]}{(n-1)!},$$

for $x > 0$ and $f(x) = 0$ otherwise, where n is a positive integer and $\lambda > 0$. Its tail df equals

$$\exp[-\lambda x] \sum_{m=0}^{n-1} \frac{(\lambda x)^m}{m!}.$$

Its hazard function is asymptotically similar to λx. Hence, it is light tailed but not superexponentially distributed.

Example 2.2.　Suppose that X has a Weibull(α, λ) distribution, that is, $P(X > x) = \exp(-(\lambda x)^\alpha)$, $x > 0$, where λ is called the scale parameter and α is called the shape parameter. Then, for $\alpha \geq 1$, it is light tailed and for $\alpha > 1$ it is superexponentially distributed. Here the hazard function is $\Lambda(x) = (\lambda x)^\alpha$.

Example 2.3.　Suppose that X has a Normal distribution with mean μ and variance σ^2 (let $N(\mu, \sigma^2)$ denote a rv with this distribution). Its tail df $\bar{F}$ evaluated at x equals the tail df of a $N(0, 1)$ rv evaluated at $(x - \mu)/\sigma$. Thus, the following inequalities follow (see, for example, Feller [1970], VII.1, Lemma 2, pg. 175):

$$\frac{1}{\sqrt{2\pi}} \frac{\sigma}{x - \mu} \left(1 - \left[\frac{\sigma}{x - \mu}\right]^2\right) \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] < \bar{F}(x)$$

$$< \frac{1}{\sqrt{2\pi}} \frac{\sigma}{x - \mu} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right],$$

for $x \geq \mu$. Although the hazard function Λ of $N(\mu, \sigma^2)$ does not have a simple explicit representation, it follows that for any $\epsilon > 0$ and x sufficiently large:

$$\frac{(x - \mu)^2}{2\sigma^2} + \log\left(\frac{x - \mu}{\sigma}\right) + \log(\sqrt{2\pi}) \leq \Lambda(x) \leq \frac{(x - \mu)^2}{2\sigma^2} + \log\left(\frac{x - \mu}{\sigma}\right) + \log(\sqrt{2\pi}) + \epsilon.$$

In particular, $\Lambda(x) \sim \frac{(x-\mu)^2}{2\sigma^2} \sim \frac{x^2}{2\sigma^2}$ and X is superexponentially distributed.

2.2.2 *Heavy-Tailed or Subexponential Distributions.* A rv X is said to be *subexponentially* distributed if it is nonnegative and if

$$\frac{P(X_1 + X_2 > u)}{2P(X_1 > u)} \to 1$$

as $u \to \infty$, where X_1 and X_2 are independent and have the same distribution as X (see, for example, Embrechts et al. [1997] and Pakes [2004] generalizes this definition to remove the non-negativity restriction. However, we consider only non-negative subexponential random variables in this paper as this considerably simplifies some of our analysis).

It can be shown that if X is subexponentially distributed then its moment generating function evaluated at any positive value equals ∞. Furthermore, if the hazard function Λ of X is sublinear, that is, $\Lambda(x)/x \to 0$ as $x \to \infty$ then under certain regularity conditions X is subexponentially distributed (see Pitman [1980]). For instance, if for all x sufficiently large, $\Lambda(x) = x/\log(x)$, then X is subexponentially distributed.

Consider the following well-known family of subexponential distributions (see, e.g., Embrechts et al. [1997]):

Example 2.4. Suppose that X has a Weibull(α, λ) distribution, that is, $P(X > x) = \exp(-(\lambda x)^\alpha)$, $x > 0$ with shape parameter $\alpha < 1$. Here the hazard function $\Lambda(x) = (\lambda x)^\alpha$ is sublinear.

Example 2.5. Suppose that $X = \exp[N(\mu, \sigma^2)]$, that is, it has a Lognormal distribution. Then, its pdf

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\frac{(-(\log x - \mu)^2}{2\sigma^2)} \quad \text{for} \quad x > 0.$$

Note that its tail df evaluated at x equals the tail df of $N(\mu, \sigma^2)$ evaluated at $\log(x)$. The same therefore is true for the hazard function Λ. Then, for any $\epsilon > 0$ and x sufficiently large:

$$\frac{(\log(x) - \mu)^2}{2\sigma^2} + \log\left(\frac{\log(x) - \mu}{\sigma}\right) + \log(\sqrt{2\pi}) \le \Lambda(x),$$

and

$$\Lambda(x) \le \frac{(\log(x) - \mu)^2}{2\sigma^2} + \log\left(\frac{\log(x) - \mu}{\sigma}\right) + \log(\sqrt{2\pi}) + \epsilon.$$

In particular, $\Lambda(x) \sim \frac{(\log(x)-\mu)^2}{2\sigma^2}$. It is well known that X is subexponentially distributed (see, e.g., Embrechts et al. [1997]).

Example 2.6. Suppose that X has a Pareto(α, λ) distribution, that is,

$$P(X > x) = \frac{1}{(1 + \lambda x)^\alpha} \quad \text{for} \quad x \ge 0.$$

Then,

$$\Lambda(x) = \alpha \log(1 + \lambda x).$$

In the literature there are different notions of heavy-tailed random variables. Most general classification considers random variable X to be heavy tailed if

$$\exp[\epsilon x] P(X \geq x) \to \infty$$

as $x \to \infty$ for all $\epsilon > 0$. Under this definition, subexponentially distributed random variables are a subclass of heavy-tailed random variables. Certain classifications include random variables with Pareto-type polynomially-decaying tails as heavy, while excluding the random variables with lighter Weibull-type tails. Our analysis in Section 5 focuses on subexponential distributions that includes Pareto as well as Weibull distributions (with shape parameter less than 1). In this article, we refer to subexponential distributions as heavy tailed.

In Sections 3 and 4, we allow the light-tailed random variables to take values in $\Re$ (to accommodate Normally distributed random variables commonly used in PERT analysis). In Section 5, our focus is on subexponential random variables. As indicated earlier, there we restrict all random variables to be nonnegative for ease of analysis.

3. LIGHT-TAILED ANALYSIS USING EXPONENTIAL TWISTING

In this section, we consider the case where $(X_i : i \leq n)$ are light tailed. We first review the importance sampling technique of exponential twisting. We then consider $P(S_n > u)$ as this provides insights for analyzing the more general probability $P(T > u)$. Specifically, we develop the logarithmic tail asymptotics for $P(S_n > u)$ and show that appropriate exponential twisting based importance sampling distribution asymptotically optimally estimates $P(S_n > u)$. Through an example we illustrate that simple exponential twisting may not work to asymptotically optimally estimate $P(T > u)$. We then propose a convex combination of exponentially twisted distributions to asymptotically optimally estimate $P(T > u)$. (See, e.g., Sadowsky and Bucklew [1990] for an application of a similar idea in the large deviations settings involving multi-dimensional light-tailed random variables.)

3.1 Logarithmic Tail Asymptotics and Simulation for $P(S_n > u)$

Let $\mathcal{H}_i(\theta) = \log \int_{-\infty}^{\infty} \exp(\theta x) f_i(x) dx$ denote the log-moment generating function of X_i. For θ such that $\mathcal{H}_i(\theta) < \infty$, let f_i^{θ} denote the pdf obtained by exponentially twisting f_i by θ, that is,

$$f_i^{\theta}(x) = \exp[\theta x - \mathcal{H}_i(\theta)] f_i(x).$$

Let P_θ denote the probability measure under which each X_i has a pdf f_i^{θ} and $(X_i : i \leq n)$ are mutually independent. Thus, if P_θ is used to estimate $P(S_n > u)$, the corresponding likelihood ratio L_θ equals

$$L_\theta = \exp\left[-\theta S_n + \sum_{i \leq n} \mathcal{H}_i(\theta) \right] \ a.s. \tag{2}$$

Let $\lambda_i(\cdot)$ and $\Lambda_i(\cdot)$ denote the hazard rate and hazard function of X_i, respectively. Then, $f_i(x) = \lambda_i(x) \exp[-\Lambda_i(x)]$. As is well known, $\lambda_i(x) = \Lambda'_i(x) = f_i(x)/\bar{F}_i(x)$, where $\Lambda'_i(x)$ denotes the derivative of $\Lambda_i(x)$.

Note that if $\frac{\Lambda_i(x)}{\Lambda_j(x)} \to \infty$ as $x \to \infty$, then X_i has a lighter tail compared to X_j, as

$$\frac{P(X_i > x)}{P(X_j > x)} = \exp(-\Lambda_i(x) + \Lambda_j(x)) \to 0.$$

The following assumption is needed to prove Theorem 3.2:

ASSUMPTION 3.1. *There exists an $\alpha \geq 1$, $k : 1 \leq k \leq n$ and constants $\tilde{\lambda}_i$, $0 < \tilde{\lambda}_i < \infty$, $1 \leq i \leq k$ such that for all $i \leq k$:*

$$\lambda_i(x) \sim \tilde{\lambda}_i \alpha x^{\alpha-1}. \tag{3}$$

(*Note that this implies that*

$$\Lambda_i(x) \sim \tilde{\lambda}_i x^\alpha, \tag{4}$$

see [Feller 1971, VIII.9, Theorem 1(b), pg 281]. Furthermore, if $k < n$, then $\Lambda_i(x)/x^\alpha \to \infty$ as $x \to \infty$, for $k < i \leq n$.

Thus, each of the rvs $(X_i : i > k)$ have a lighter tail distribution compared to the rvs $(X_i : i \leq k)$.

Let

$$\lambda^* = \frac{1}{\left(\sum_{i \leq k} 1/\tilde{\lambda}_i^{\frac{1}{\alpha-1}} \right)^{\alpha-1}}, \tag{5}$$

for $\alpha > 1$ and

$$\lambda^* = \min_{i \leq k} \tilde{\lambda}_i$$

for $\alpha = 1$. Recall that $E_{\tilde{P}}$ denotes the expectation operator associated with probability $\tilde{P}$.

THEOREM 3.2. *Under Assumption 3.1,*

$$\lim_{u \to \infty} \frac{\log P(S_n > u)}{u^\alpha} = -\lambda^*, \tag{6}$$

and

$$\lim_{u \to \infty} \frac{\log E_{P_{\theta_u}} L^2 I(S_n > u)}{u^\alpha} = -2\lambda^*, \tag{7}$$

where, for $\alpha > 1$, $\theta_u \sim \alpha \lambda^* u^{\alpha-1}$, and for $\alpha = 1$, $\theta_u \sim \min_{i \leq k} \tilde{\lambda}_i$ such that $\mathcal{H}_i(\theta_u) = o(u)$ for each $i \leq n$.

From Theorem 3.2, it follows that P_{θ_u} asymptotically optimally estimates $P(S_n > u)$.

Remark 3.3. For $\alpha > 1$, the choice $\theta_u \sim \alpha \lambda^* u^{\alpha-1}$ is motivated by the fact that the value that minimizes an upper bound $\exp[-\theta u + \sum_{i \leq n} \mathcal{H}_i(\theta)]$ (for $\theta > 0$) on the right-hand side in (2) is asymptotically similar to $\alpha \lambda^* u^{\alpha-1}$; the latter is given by the solution to the equation

$$\sum_{i \leq n} \mathcal{H}_i'(\theta) = u, \quad \theta > 0. \tag{8}$$

The analysis supporting this is omitted for brevity. When $\alpha = 1$, typically, θ_u can be easily selected so that the conditions in Theorem 3.2 hold. For instance, suppose that X_1 is exponentially distributed with rate η, X_2 has a Gamma distribution with rate η and shape parameter α, while all other $(X_i : i \geq 3)$ have lighter tails, for example, they are exponentially distributed with rate $\mu > \eta$. Then, we may set $\theta_u = \eta(1 - \frac{1}{u^\beta})$, for any $\beta > 0$. For this θ_u, $\mathcal{H}_1(\theta_u) = \beta \log u$, $\mathcal{H}_2(\theta_u) = \beta\alpha \log u$, while $\mathcal{H}_i(\theta_u) = \log \frac{\mu}{\mu-\eta-1/u^\beta}$ for $i \geq 3$. Hence, all $\mathcal{H}_i(\theta_u)$ are $o(u)$.

Remark 3.4.   The proof of Theorem 3.2 does not require the existence of pdf f_i, or equivalently, the hazard rate λ_i, for each i. It holds even if relation (4) is assumed to hold in lieu of (3) in Assumption 3.1. However, since (3) is useful in simplifying later analysis, we retain Assumption 3.1 in its present form.

3.2 Logarithmic Tail Asymptotics for $P(T > u)$

The logarithmic tail asymptotics for $P(T > u)$ follows easily from Theorem 3.2. Some notation is needed for this purpose.

Recall from Assumption 3.1 that, roughly speaking, $(X_i : i \leq k)$ have tails of a similar magnitude that are much heavier than the tails of $(X_i : i > k)$. Let $\tilde{\mathcal{P}}_j = \mathcal{P}_j \cap \{1, 2, \ldots, k\}$. For each $j \leq \tau$ such that $\tilde{\mathcal{P}}_j$ is nonempty, when $\alpha > 1$, let

$$\lambda_j^* = \frac{1}{\left(\sum_{i \in \tilde{\mathcal{P}}_j} 1/\tilde{\lambda}_i^{\frac{1}{\alpha-1}} \right)^{\alpha-1}}. \tag{9}$$

Let $\lambda_j^* = \min_{i \in \tilde{\mathcal{P}}_j} \tilde{\lambda}_i$ when $\alpha = 1$. Set λ_j^* to ∞ if $\tilde{\mathcal{P}}_j$ is empty. Let $\mathcal{L}_j = \sum_{i \in \mathcal{P}_j} X_i$.

THEOREM 3.5.   *Under Assumption* 3.1,

$$\lim_{u \to \infty} \frac{\log P(T > u)}{u^\alpha} = -\min_{j \leq \tau} \lambda_j^*. \tag{10}$$

Remark 3.6.   Thus, in the PERT network terminology, the tail decay rate of T is governed by the tail decay rate of the most likely 'path' $\mathcal{P}_j$, $j \leq \tau$ in the network. Note that when $\alpha = 1$, $\min_{j \leq \tau} \lambda_j^*$ equals $\min_{j \leq k} \tilde{\lambda}_j$. Thus, when the tail distribution function of a rv in the network with the heaviest tail decays at an exponential rate, the decay rate of the network delay T is also exponential with the same rate.

3.3 Asymptotically Optimal Simulation for $P(T > u)$

In Example 3.7, we illustrate that under exponential twisting the change of measure that asymptotically optimally estimates $P(S_n > u)$ need not asymptotically optimally estimate $P(T > u)$.

Example 3.7.   Suppose that $n = 2$ and that X_1 and X_2 are iid with $N(0, \sigma^2)$ distribution. Then,

$$\lim_{u \to \infty} \frac{\log P(X_i > u)}{u^2} = -\frac{1}{2\sigma^2}$$

for $i = 1, 2$. In particular, Assumption 3.1 holds with $k = 2$, $\alpha = 2$, $\tilde{\lambda}_1 = \tilde{\lambda}_2 = \frac{1}{2\sigma^2}$. Also, $\lambda^* = \frac{1}{4\sigma^2}$. Further suppose that $T = \max(X_1, X_2)$. It follows from Theorem 3.2 that

$$\lim_{u \to \infty} \frac{\log P(S_2 > u)}{u^2} = -\frac{1}{4\sigma^2},$$

and

$$\lim_{u \to \infty} \frac{\log E_{P_{\theta_u}} L_{\theta_u}^2 I(S_2 > u)}{u^\alpha} = -\frac{1}{2\sigma^2},$$

where $\theta_u = \frac{u}{2\sigma^2}$. (Recall the definition of P_θ from Section 3.1). Again, since X_1 and X_2 are iid,

$$P(T > u) = 1 - P(X_1 \le u)P(X_2 \le u) = 2P(X_1 > u) - P(X_1 > u)^2 \sim 2P(X_1 > u),$$

and it follows that

$$\lim_{u \to \infty} \frac{\log P(T > u)}{u^2} = -\frac{1}{2\sigma^2}.$$

We now argue that

$$\liminf_{u \to \infty} \frac{\log E_{P_{\theta_u}} L_{\theta_u}^2 I(T > u)}{u^2} > -\frac{1}{\sigma^2}. \tag{11}$$

Hence, while P_{θ_u} asymptotically optimally estimates $P(S_2 > u)$, it does not do so for $P(T > u)$. To see (11), note that $\mathcal{H}_i(\theta) = \theta^2 \sigma^2/2$ for $i = 1, 2$. Hence, $\mathcal{H}_i(\theta_u) = \frac{u^2}{8\sigma^2}$ for $i = 1, 2$. Specializing (2) to our example, we get:

$$L_{\theta_u} = \exp\left[-\frac{u}{2\sigma^2}(X_1 + X_2) + \frac{u^2}{4\sigma^2} \right]$$

almost surely. In particular, it follows that $E_{P_{\theta_u}} L_{\theta_u}^2 I(T > u) = E_P L_{\theta_u} I(T > u)$ is bounded from below by

$$\exp\left[-\frac{1}{2\sigma^2}u(u + \sqrt{u} + c) + \frac{u^2}{4\sigma^2} \right] P(X_1 \in (u, u + \sqrt{u})) P(0 \le X_2 \le c) \tag{12}$$

for a positive constant c. Since X_1 has a $N(0, \sigma^2)$ distribution, it is easily seen that $\frac{P(X_1 \in (u, u+\sqrt{u}))}{P(X_1 > u)} \to 1$ as $u \to \infty$. Hence,

$$\lim_{u \to \infty} \frac{\log P(X_1 \in (u, u + \sqrt{u}))}{u^2} = -\frac{1}{2\sigma^2}. \tag{13}$$

Taking logarithm of (12), dividing by u^2 and taking $\liminf$ as $u \to \infty$, noting (13), the result can be seen to equal $-\frac{3}{4\sigma^2}u^2$ and hence (11) follows.

Similarly, it is easily argued that just by exponentially twisting X_1 by any parameter and not X_2, or vice-versa, asymptotic optimality cannot be achieved in this example. This is easily seen by considering the contribution to the second moment by the set where the untwisted component taking values greater than u.

In Example 3.7, we illustrated that the exponential-twisted-based change of measure that asymptotically optimally estimates $P(S_n > u)$ may not be as

effective in estimating $P(T > u)$. A brief intuitive explanation is as follows: It is well known that a change of measure that sufficiently emphasizes the most likely paths to the rare event provides an effective importance sampling estimator for the rare event probability (see, e.g., Juneja and Shahabuddin [2006]). However, since the most likely way the event $\{S_n > u\}$ occurs may differ from the way the event $\{T > u\}$ occurs, an importance sampling estimator that is effective in estimating $P(S_n > u)$ may not be as effective in estimating $P(T > u)$. For instance, in Example 3.7, for large u, a likely way for $\{S_2 > u\}$ to occur can be roughly seen to correspond to both X_1 and X_2 taking values close to $u/2$, while a likely way for $\{T > u\}$ to occur involves X_1 exceeding u and X_2 taking a usual small value (relative to u), or vice-versa. It can be seen in Example 3.7 that under P_{θ_u} each X_i has a $N(u/2, \sigma^2)$ distribution. This explains its lack of effectiveness in estimating $P(T > u)$ in that example.

We now propose an asymptotically optimal change of measure to estimate $P(T > u)$ that involves taking a convex combination of τ (τ is defined in the Introduction) exponentially twisted changes of measure. It can be seen that this change of measure assigns significant probability to events $\{\mathcal{L}_j > u\}$ for $j \leq \tau$ and $\tilde{\mathcal{P}}_j$ nonempty (recall that $\mathcal{L}_j = \sum_{i \in \mathcal{P}_j} X_i$), and hence to the event

$$\{T > u\} = \bigcup_{j \leq \tau} \{\mathcal{L}_j > u\}.$$

When $\tilde{\mathcal{P}}_j$ is nonempty, let $P_{\theta_{u_j}}$ for $j \leq \tau$, denote a change of measure under which the distribution of each X_i for $i \in \mathcal{P}_j$ is obtained by exponentially twisting the original distribution by θ_{u_j}, where $\theta_{u_j} \sim \alpha \lambda_j^* u^{\alpha - 1}$, if $\alpha > 1$. When $\alpha = 1$, select θ_{u_j} such that $\theta_{u_j} \sim \lambda_j^*$ and $\mathcal{H}_i(\theta_{u_j}) = o(u)$ for each $i \in \mathcal{P}_j$. Set $\theta_{u_j} = 0$ if $\tilde{\mathcal{P}}_j$ is empty. From Theorem 3.2, it follows that $P_{\theta_{u_j}}$ is an asymptotically optimal importance sampling distribution to estimate $P(\mathcal{L}_j > u)$ when $\tilde{\mathcal{P}}_j$ is nonempty.

Now consider a probability measure P^* for $(X_i : i \leq n)$ such that for any $A \in \mathcal{F}$:

$$P^*(A) = \sum_{j \leq \tau} p_j P_{\theta_{u_j}}(A),$$

where $(p_j > 0 : j \leq \tau)$ and $\sum_{j \leq \tau} p_j = 1$. Note that the likelihood ratio L of P with respect to P^* equals

$$L = \frac{1}{\sum_{j \leq \tau} p_j \exp\left(\theta_{u_j} \sum_{i \in \mathcal{P}_j} X_i - \sum_{i \in \mathcal{P}_j} \mathcal{H}_i(\theta_{u_j})\right)} \quad a.s. \tag{14}$$

THEOREM 3.8.   *Under Assumption* 3.1,

$$\lim_{u \to \infty} \frac{\log E_{P^*}[L^2 I(T > u)]}{u^\alpha} = -2 \min_{j \leq \tau} \lambda_j^*. \tag{15}$$

From this theorem and (10), it follows that P^* asymptotically optimally estimates $P(T > u)$ for any arbitrary $(p_j > 0 : j \leq \tau)$ such that $\sum_{j \leq \tau} p_j = 1$. Later in Section 6, we discuss a heuristic for arriving at a good choice for $(p_j : j \leq \tau)$.

4. LIGHT-TAILED ANALYSIS USING HAZARD RATE TWISTING

As mentioned in the Introduction, for a subexponential distribution, exponential twisting is no longer feasible as the moment generating function evaluated at any positive value equals ∞. To remedy this, Juneja and Shahabuddin [2002] introduced hazard rate twisting to efficiently estimate $P(S_n > u)$ (recall that $S_n = \sum_{i \leq n} X_i$) where $(X_i : 1 \leq i \leq n)$ are mutually independent, and are identically and subexponentially distributed. In this and the next section we show that hazard rate twisting can be implemented to efficiently estimate $P(T > u)$ (and hence also $P(S_n > u)$) even when all the underlying rvs are allowed to have different distributions and may even be light tailed. A potential drawback of this approach is that both for light and heavy-tailed rvs, this twisting may be difficult to implement as it may be difficult to sample from the hazard rate twisted distribution. Fortunately, in many such cases it may be feasible to twist with a function that is asymptotically similar to the hazard function, such that it is easy to sample from the 'asymptotically hazard rate twisted' distribution and the resultant distribution is asymptotically optimal in estimating the rare event probabilities that we consider.

In this section, we first review the hazard rate twisting technique and introduce the more general asymptotic hazard rate twisting technique when all $(X_i : i \leq n)$ are light tailed. We also prove that this technique asymptotically optimally estimates $P(T > u)$ in the light-tailed settings. In Section 5, we develop the idea of asymptotic hazard rate twisting when one or more $(X_i : i \leq n)$ are subexponentially distributed.

4.1 Hazard Rate Twisting

As mentioned in Section 3.1, the pdf $f_i(x)$ may be re-expressed as $\lambda_i(x) \exp[-\Lambda_i(x)]$, where $\lambda_i(\cdot)$ and $\Lambda_i(\cdot)$ denote the hazard rate and the hazard function of X_i, respectively. Then the pdf obtained by hazard rate twisting of the original distribution by an amount $\theta < 1$ is given by h_i^θ, where

$$h_i^\theta(x) = \lambda_i(x)(1 - \theta) \exp[-(1 - \theta)\Lambda_i(x)].$$

Note that h_i^θ has corresponding hazard rate $\lambda_i(x)(1 - \theta)$ and hazard function $\Lambda_i(x)(1 - \theta)$ at x. For instance, the pdf obtained by hazard rate twisting of the Weibull(α, λ) distribution in Examples 2.2 and 2.4 by an amount θ is Weibull($\alpha, \lambda(1 - \theta)^{1/\alpha}$). Similarly, hazard rate twisting of Pareto(α, λ) distribution in Example 2.6 by θ results in Pareto($\alpha(1 - \theta), \lambda$) distribution. The pdf obtained by hazard rate twisting of $N(\mu, \sigma^2)$ has a more complex form. Note that $\lambda_i(x) = f_i(x)/\bar{F}_i(x)$. Hence, $h_i^\theta(x)$ may alternatively be expressed as $f_i(x)(1 - \theta)/\bar{F}_i(x)^\theta$. In particular, the pdf obtained by hazard rate twisting $N(\mu, \sigma^2)$ by amount θ has the form

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{1-\theta} \frac{(1 - \theta) \exp\left[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\right]}{\left(\int_x^\infty \exp\left[-\frac{1}{2}(\frac{y-\mu}{\sigma})^2\right]dy\right)^\theta}. \tag{16}$$

To see that hazard rate twisting results in a significantly different distribution compared to exponential twisting note that exponentially twisting a

Weibull(α, λ) distribution is feasible only when $\alpha \geq 1$, and for $\alpha > 1$ the resultant distribution is no longer Weibull. As mentioned earlier, exponential twisting a Pareto distribution is not feasible. Furthermore, it is easy to check that $N(\mu + \theta\sigma^2, \sigma^2)$ is the distribution obtained when $N(\mu, \sigma^2)$ is exponentially twisted by an amount θ.

Juneja and Shahabuddin [2002] show that hazard rate twisting with $\theta_u = 1 - \frac{n}{\Lambda_i(u)}$, asymptotically optimally estimates $P(S_n > u)$ when X_i's are iid and subexponentially distributed (asymptotically as $u \to \infty$, for a fixed n). This can be easily seen to be true more generally for $\theta_u = 1 - \frac{1}{\Theta(\Lambda_i(u))}$. Here, we refer to any non-negative function $a(u)$ as $\Theta(b(u))$, where $b(u)$ is another non-negative function, if there exist constants $0 < K_1 < K_2$, and $K_1 b(u) \leq a(u) \leq K_2 b(u)$ for all sufficiently large u.

4.2 Conditions on Asymptotic Hazard Functions

Whenever hazard rate twisting results in a distribution from which it is easy to sample (e.g., Weibull distribution and Pareto distribution discussed at the end of Section 4.1), it may be practical to conduct hazard rate twisting to estimate $P(T > u)$. However, when this is not the case, more general methods may be needed. For instance, it may be seen from (16) that hazard rate twisting a Normal distribution does not result in another Normal distribution. In fact it results in a distribution from which generating random variates may be difficult. The same can be seen to be true for the Lognormal distribution. Fortunately, as noted in Example 2.3, the hazard function $\Lambda(x)$ of a Normal distribution is asymptotically similar to $\frac{(x-\mu)^2}{2\sigma^2}$. If this function is used to twist the Normal distribution, then we get another Normal distribution. Similarly, from Example 2.5, it can be seen that if we twist a Lognormal distribution with $\frac{(\log x - \mu)^2}{2\sigma^2}$ (asymptotically similar to its hazard function), we again get a Lognormal distribution. We now describe some conditions imposed on the asymptotic hazard functions that, with a suitably selected twisting parameter, ensure that the associated importance sampling asymptotically optimally estimates $P(T > u)$.

Consider non-negative 'asymptotic hazard' functions $\tilde{\Lambda}_i(\cdot) : \Re \to \Re^+$, $i \leq k$ such that

$$\tilde{\Lambda}_i(x) \sim \Lambda_i(x). \tag{17}$$

Further suppose that $\tilde{\Lambda}_i(x)$ satisfy the condition

$$\tilde{\Lambda}_i(x) \leq \Lambda_i(x) + a_i \max(\log \Lambda_i(x), 1) \tag{18}$$

for some $a_i \geq 0$, and for all x and $i \leq k$.

Clearly, the conditions (17) and (18) are satisfied by $\tilde{\Lambda}_i(x) = \Lambda_i(x)$.

Let $\Gamma_i(\theta) = \log \int_{-\infty}^{\infty} \exp[\theta \tilde{\Lambda}_i(x)] f_i(x) dx$. Note that if $\tilde{\Lambda}_i$ is used as a proxy for Λ_i in hazard rate twisting of X_i by amount θ, the resultant pdf equals

$$\exp[\theta \tilde{\Lambda}_i(x) - \Gamma_i(\theta)] f_i(x).$$

Also note that $\Gamma_i(\theta) = \frac{1}{1-\theta}$ if $\tilde{\Lambda}_i = \Lambda_i$.

4.3 Asymptotically Optimal Simulation

Consider an importance sampling distribution P^* under which all X_i's remain mutually independent. Each X_i for $i \leq k$, has a pdf given by

$$f_i^*(x) = \exp[\theta_u \tilde{\Lambda}_i(x) - \Gamma_i(\theta_u)] f_i(x)$$

and $\theta_u = 1 - 1/\Theta(u^\alpha)$. Each X_i for $i > k$ has the original distribution.

Note that the likelihood ratio of P with respect to P^* equals

$$L = \exp\left[-\theta_u \sum_{i=1}^{k} \tilde{\Lambda}_i(X_i) + \sum_{i=1}^{k} \Gamma_i(\theta_u) \right] \quad a.s. \tag{19}$$

THEOREM 4.1.   *Under Assumption* 3.1, (17) *and* (18),

$$\lim_{u \to \infty} \frac{\log E_{P^*} L^2 I(T > u)}{u^\alpha} = -2 \min_{j \leq \tau} \lambda_j^*. \tag{20}$$

In view of Theorem 3.5, this establishes that P^* estimates $P(T > u)$ asymptotically optimally. Also note that Theorem 4.1 implies that,

$$\lim_{u \to \infty} \frac{\log E_{P^*} L^2 I(S_n > u)}{u^\alpha} = -2\lambda^*. \tag{21}$$

Remark 4.2.   In light of Example 3.7, this result may appear surprising as that example illustrates that the exponential twisting based importance sampling distribution that asymptotically optimally estimates $P(S_n > u)$ may not be as efficient in estimating $P(T > u)$. The intuitive explanation for effectiveness of asymptotic hazard rate twisting for both the probabilities is that under this twisting, the probability of a rv taking large values is much larger than that under naive simulation and is typically more spread out compared to exponential twisting. For example, it can be seen that if an asymptotic hazard function $\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$ is used to twist $N(\mu, \sigma^2)$ rv and θ_u is set to $1 - \frac{c}{u^2}$ for a constant c, then, the resulting distribution is $N(\mu, u^2\sigma^2/c)$. It is easy to see that $P(N(\mu, u^2\sigma^2/c) > u/2)$ and $P(N(\mu, u^2\sigma^2/c) > u)$ both converge to positive constants as $u \to \infty$.

Remark 4.3.   Empirically we observed that the above hazard rate twisting distribution gives large variance reduction compared to naive simulation only when the probability of interest is extremely small. A significant improvement over it is achieved if in a manner analogous to that described in Section 3.3, we use a convex combination of τ probability measures each using asymptotic hazard rate twisting geared to a particular path as described above (again, the original measure may be used for paths that do not contain any activity X_i, $i \leq k$). The proof of asymptotic optimality then follows exactly as the proof of Theorem 3.8 (with Theorem 4.1 replacing Theorem 3.2 in the proof) and is therefore omitted. In our experiments we report the results using this mixture of probability measures.

5. HEAVY-TAILED ANALYSIS

We now develop an exact tail-asymptotic and asymptotic hazard rate twisting based simulation methods to estimate $P(T > u)$ for the case

where the heaviest-tailed rv in the network has a subexponential distribution.

5.1 Exact Asymptotic

Theorem 5.2 develops an exact asymptotic for $P(T > u)$. The following assumption is needed to prove it.

ASSUMPTION 5.1. *The random variables $(X_i : i \leq n)$ are non-negative. Furthermore, X_1 is subexponentially distributed and there exists a $(k : 1 \leq k \leq n)$ and positive constants $(\gamma_i \leq 1 : i \leq k)$ such that for $i \leq k$,*

$$\lim_{u \to \infty} \frac{P(X_i > u)}{P(X_1 > u)} = \gamma_i$$

and

$$\lim_{u \to \infty} \frac{P(X_i > u)}{P(X_1 > u)} = 0$$

for $i > k$ (if $k < n$).

Under Assumption 5.1, it is well known that

$$P\left(\sum_{i \leq n} X_i > u\right) \sim P(\max_{i \leq n} X_i > u) \sim \sum_{i \leq n} P(X_i > u), \tag{22}$$

(see, e.g., Sigman [1999]).

THEOREM 5.2. *Under Assumption 5.1,*

$$P(T > u) \sim \left(\sum_{i=1}^{k} \gamma_i\right) P(X_1 > u). \tag{23}$$

Example 5.3.   Consider the case where all X_i's have a Weibull distribution, that is, $P(X_i > x) = \exp(-(\eta_i x)^{\alpha_i})$, $x > 0$. Suppose that $\alpha_1 < 1$ so that X_1 is subexponentially distributed (see, e.g., Embrechts et al. [1997]). Since it has the heaviest tail, it follows that if $\alpha_i = \alpha_1$ then, $\eta_i \geq \eta_1$. In this setting, note that

$$\lim_{u \to \infty} \frac{P(X_i > u)}{P(X_1 > u)}$$

equals either one or zero. Hence, k corresponds the number of rvs with the distribution identical to X_1, and $P(T > u) \sim k P(X_1 > u) = k \exp(-(\eta_1 u)^{\alpha_1})$.

Similarly, if for each i, X_i has a Lognormal(μ_i, σ_i^2) distribution, that is, its pdf

$$f_i(x) = \frac{1}{x\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(\log x - \mu_i)^2}{2\sigma_i^2}\right) \quad \text{for} \quad x > 0,$$

then again $\sigma_i^2 \leq \sigma_1^2$ and if $\sigma_i^2 = \sigma_1^2$, then $\mu_i \leq \mu_1$. Here again, k denotes the number of rvs with the distribution identical to X_1, and

$$P(T > u) \sim k P(X_1 > u) \sim \frac{k\sigma_1}{\sqrt{2\pi}\log u} \exp\left(\frac{-(\log u - \mu_1)^2}{(2\sigma_1^2)}\right)$$

(an exact asymptotic of $P(X_1 > u)$ may be inferred from Example 2.3).

Finally, suppose that all X_i have a Pareto(α_i, η_i) distribution, that is,

$$P(X_i > x) = \frac{1}{(1 + \eta_i x)^{\alpha_i}} \quad \text{for} \quad x > 0.$$

Now X_i, for $i \leq k$ are rvs with $\alpha_i = \alpha_1$, while $\alpha_i > \alpha_1$ for $i > k$. Hence,

$$P(T > u) \sim \sum_{i=1}^{k} \frac{1}{(1 + \eta_i u)^{\alpha_1}} \sim \frac{1}{u^{\alpha_1}} \sum_{i=1}^{k} \frac{1}{\eta_i^{\alpha_1}}.$$

5.2 Conditions on Asymptotic Hazard Functions

We now describe the conditions imposed on the non-negative asymptotic hazard functions $\tilde{\Lambda}_i(\cdot) : \Re \to \Re^+$, $i \leq n$, that can be used to asymptotically optimally estimate $P(T > u)$ via appropriate importance sampling. As in the light-tailed settings, we assume that

$$\tilde{\Lambda}_1(x) \sim \Lambda_1(x), \tag{24}$$

and for all i, there exists $a_i \geq 0$,

$$\tilde{\Lambda}_i(x) \leq \Lambda_i(x) + a_i \max(\log \Lambda_i(x), 1). \tag{25}$$

Rather than selecting $\tilde{\Lambda}_i(u) \sim \Lambda_i(u)$ for $i \geq 2$, as in the light-tailed settings, we allow greater flexibility in the choice of $\tilde{\Lambda}_i$ by considering those that satisfy

$$\liminf_{u \to \infty} \frac{\tilde{\Lambda}_i(u)}{\tilde{\Lambda}_1(u)} \geq 1 \tag{26}$$

for $i \geq 2$. It is easy to see that the Assumption 5.1 ensures that the hazard functions satisfy this condition. To see this, note that under Assumption 5.1,

$$\lim_{u \to \infty} \Lambda_1(u) \left(\frac{\Lambda_i(u)}{\Lambda_1(u)} - 1 \right) \geq 0.$$

Therefore,

$$\liminf_{u \to \infty} \frac{\Lambda_i(u)}{\Lambda_1(u)} \geq 1.$$

To ease the analysis, we further restrict $\tilde{\Lambda}_1$ to be eventually a nondecreasing function and to satisfy the following property: For every $\epsilon > 0$, there exists a u_ϵ such that

$$\sum_{i=1}^{n} \tilde{\Lambda}_1(x_i) \geq \tilde{\Lambda}_1 \left(\sum_{i=1}^{n} x_i \right) - \epsilon. \tag{27}$$

for all $(x_1, \ldots, x_n) \geq 0$ with $\sum_{i=1}^{n} x_i \geq u_\epsilon$. This property is shown to be true in Juneja and Shahabuddin [2002] for any non-negative function $\Lambda(\cdot)$ that is eventually everywhere differentiable and whose derivative is eventually decreasing to zero. It is satisfied by hazard rates of commonly encountered subexponential distributions such as Pareto, Lognormal and Weibull distributions with shape parameter less than 1.

As mentioned in Section 4, for Lognormal(μ_i, σ_i^2) rv, the hazard-rate-twisted distribution is not Lognormal and may be difficult to sample from. However, $\Lambda_i(x) \sim (\log x - \mu_i)^2/(2\sigma_i^2)$. Hence, we can set $\tilde{\Lambda}_i(x) = (\log x - \mu_i)^2/(2\sigma_i^2)$. This function satisfies the restrictions imposed on $\tilde{\Lambda}_i(\cdot)$ in (25) and the resultant twisted distribution can be seen to be Lognormal.

5.3 Asymptotically Optimal Simulation

Let P^* denote the probability measure under which the X_i's remain mutually independent and each X_i has the distribution

$$d F_i^*(x) = \exp[\theta_u \tilde{\Lambda}_i(x) - \Gamma_i(\theta_u)]d F_i(x),$$

where $\theta_u = 1 - 1/\Theta(\tilde{\Lambda}_1(u))$. In Juneja and Shahabuddin [2002], this value of θ_u is shown to work well for estimating $P(S_n > u)$ when all rvs are iid and subexponentially distributed and one uses plain hazard rate twisting. The fact that $P(T > u)$ has the same exact asymptotic as $P(S_n > u)$ suggests that a similar θ_u may work in this setting as well.

Theorem 5.2 implies that

$$\lim_{u \to \infty} \frac{\log P(T > u)}{\Lambda_1(u)} = -1. \tag{28}$$

Theorem 5.4 states the main result of this section. Together with (28), it implies asymptotic optimality of P^*.

THEOREM 5.4. *Under Assumption* 5.1, (24), (25) *and* (27),

$$\lim_{u \to \infty} \frac{\log E_{P^*} L^2 I(T > u)}{\Lambda_1(u)} = -2. \tag{29}$$

Remark 5.5. Theorem 5.4 can be seen to hold even if the light-tailed $X_i's$ retain their original distribution under P^*. The proof then follows along the lines of Theorem 3.2 where all rvs with $\alpha_i > \alpha$ retain their original distribution under the asymptotically optimal change of measure. However, we do not explicitly prove this to avoid repetition.

Remark 5.6. As we discussed in Remark 4.3 in the light-tailed settings, significant performance improvement over P^* may be achieved, if in a manner analogous to that described in Section 3.3, we use a convex combination of τ probability measures each using asymptotic hazard rate twisting geared to a particular path. The proof of asymptotic optimality then follows exactly as the proof of Theorem 3.8. As in the light-tailed settings, in our experiments we report the results using this mixture of probability measures.

6. PARAMETER SELECTION FOR IMPORTANCE SAMPLING

In the previous sections we gave order of magnitude specifications of optimal θ_u for asymptotically optimal estimation of $P(S_n > u)$ and $P(T > u)$. Similarly, we proved that any choice of $(p_j : j \leq \tau)$ where each $p_j > 0$ and $\sum_{j \leq \tau} p_j = 1$ can be used to asymptotically optimally estimate $P(T > u)$ with appropriate

probability measures corresponding to each set $\mathcal{P}_j$. Empirically, we observe that when superexponential rvs are involved and exponential twisting is conducted in estimating $P(S_n > u)$, setting θ_u as a solution to (8) performs very well.

Empirically we also observe that when asymptotic hazard rate twisting or hazard rate twisting is applied, setting θ_u to any $1 - 1/\Theta(u^\alpha)$ function in light-tailed settings (Theorem 4.1) and $1 - 1/\Theta(\Lambda_1(u))$ function in the heavy-tailed settings (Theorem 5.4) may achieve variance reduction only for extremely large values of u. To achieve simulation efficiency increase for practical ranges of u, say when the rare event probabilities are of the order of 10^{-1} to 10^{-6}, we need good heuristics to select θ_u. Similarly, though setting each $p_j = 1/\tau$ gives reliable results, the results are improved substantially (about 2-3 times in our experiments) by selecting them cleverly, as outlined in this section.

6.1 Selecting θ_u

We specialize a heuristic from Huang and Shahabuddin [2004] that considers a more general set-up. We focus on arriving at a good θ_u for estimating $P(S_n > u)$. This then extends to finding good $(\theta_{u_j}, j \leq \tau)$ for estimating $P(T > u)$ by setting each θ_{u_j} to the value that is good for estimating the corresponding $P(\mathcal{L}_j > u)$.

For estimating $P(S_n > u)$, the second moment obtained by asymptotic hazard rate twisting by amount θ is given by (see (19))

$$E_{P^*}(L^2 I(S_n > u)) = \exp\left(2\sum_{i \leq n} \Gamma_i(\theta)\right) E_{P^*}\left[\exp\left(-2\theta \sum_{i \leq n} \tilde{\Lambda}_i(X_i)\right) I(S_n > u)\right],$$
(30)

both under light and heavy-tailed settings. Here, for light-tailed settings, we have assumed that k defined in Assumption 3.1 equals n. When, $k < n$, the impact of $(X_i : i > k)$ on the rare event probability $P(S_n > u)$ is negligible, so heuristically we can proceed with finding best twisting parameter by ignoring these rvs.

We now follow two steps. In the first step, for a given $\theta \geq 0$, we find a deterministic upper bound on $E_{P^*}(L^2 I(S_n > u))$. This is done by minimizing $\sum_{i \leq n} \tilde{\Lambda}_i(x_i)$ over all $(x_i : i \leq n)$ satisfying $\sum_{i \leq n} x_i \geq u$. In the case where any random variable X_i is non-negative we add an additional constraint $x_i \geq 0$. Let $g(u)$ denote such a solution. This yields the bound

$$E_{P^*}(L^2 I(S_n > u)) \leq \exp\left(2\sum_{i \leq n} \Gamma_i(\theta) - 2\theta g(u)\right).$$
(31)

Often, $g(u)$ may not be available in closed form, however a simple approximate solution to $g(u)$ may be found that is close to it for all sufficiently large values of u (this is illustrated through an example in Case 1 in Section 7). Then, we replace $g(u)$ above by its approximation. In the second step we find the $\theta = \theta_u \geq 0$ that minimizes the above bound. In all the examples reported in Section 7, the θ_{u_j} so obtained (by focussing on $P(\mathcal{L}_j > u)$) have a form that is consistent with the requirements in the corresponding theorem statements, thus assuring asymptotic optimality. For small u, the θ_{u_j} so obtained may be negative. Then, we set $\theta_{u_j} = 0$, corresponding to naive simulation.

6.2 Selection of p_j's

Here, we use the ideas reported in Juneja and Shahabuddin [2006]. We illustrate them in the framework in Section 3.3. Again, first an upper bound to $E_{P^*}(L^2 I(T > u))$ is developed where L is given by (14). Then, we find the vector $(p_j : j \leq \tau)$ that minimizes this bound. Note that for $\theta_{u_j} > 0$,

$$LI(\mathcal{L}_j > u) \leq \frac{K_j}{p_j} I(\mathcal{L}_j > u),$$

where $K_j = \exp(-\theta_{u_j} u + \sum_{i \in \mathcal{P}_j} \mathcal{H}_i(\theta_{u_j}))$. Therefore,

$$LI(T > u) \leq \left(\max_{j \leq \tau} \frac{K_j}{p_j} \right) I(T > u),$$

and

$$E_{P^*}(L^2 I(T > u)) \leq \left(\max_{j \leq \tau} \frac{K_j}{p_j} \right)^2.$$

Minimizing the above upper bound subject to $\sum_{j \leq \tau} p_j = 1$, and $p_j \geq 0$ for $j \leq \tau$, yields the weights

$$p_j^* = \frac{K_j}{\sum_{i=1}^m K_i}.$$

The asymptotic hazard rate twisting case, both in light and heavy-tailed settings, is similar; in this case we get the same expression for p_j^* except that now $K_j = \exp(-\theta_{u_j} g(u) + \sum_{i \in \mathcal{P}_j} \Gamma_i(\theta_{u_j}))$. Here, we make use of the bound in (31) evaluated at $\theta = \theta_{u_j}$. This methodology is used to determine p_j's in the experiments in Section 7.

7. NUMERICAL RESULTS

We consider an example with $n = 15$ and $\tau = 10$. The $(\mathcal{L}_j : j \leq 10)$, respectively, equal

$$X_1 + X_4 + X_{11} + X_{15},$$
$$X_1 + X_4 + X_{12},$$
$$X_2 + X_5 + X_{11} + X_{15},$$
$$X_2 + X_5 + X_{12},$$
$$X_2 + X_6 + X_{13},$$
$$X_2 + X_7 + X_{14},$$
$$X_3 + X_8 + X_{11} + X_{15},$$
$$X_3 + X_8 + X_{12},$$
$$X_3 + X_9 + X_{15},$$
$$\text{and } X_3 + X_{10} + X_{14}.$$

We test the performance of the proposed techniques for five sets of distributions of $(X_i : i \leq 15)$ described below. For $j \leq 10$, let n_j denote the number of terms in $\mathcal{P}_j$.

Case 1. *Subexponential-tailed with Hazard Rate Twisting.* In this case $X_1, \ldots, X_8$ have the Pareto$(2, 1)$ distribution and $X_9, \ldots, X_{15}$ have the Weibull$(0.5, 2)$ distribution. Note that in this case all X_i's are heavy tailed. We set $\tilde{\Lambda}_i(x) = \Lambda_1(x) = 2\ln(1 + x)$ for $i = 1, \ldots, 8$. Also, $\tilde{\Lambda}_i(x) = \Lambda_i(x) = (2x)^{0.5}$ for $i = 9, \ldots, 15$. It follows that $\Gamma_i(\theta) = -\ln(1 - \theta)$ for all $i = 1, \ldots, 15$. We apply the procedure outlined in Section 6 to determine good $(\theta_{u_j} : j \leq 10)$. As an illustration, consider the procedure to determine θ_{u_1}. Here, we first need to minimize

$$2\ln(1 + x_1) + 2\ln(1 + x_4) + \sqrt{2x_{11}} + \sqrt{2x_{15}},$$

subject to $x_1 + x_4 + x_{11} + x_{15} \geq u$ and $x_i \geq 0$, $1 \leq i \leq 4$. Note that $2\ln(1 + x)$ is dominated by $\sqrt{2x}$ for all sufficiently large x. Due to the concavity of $\ln(1 + x)$ it is easy to see that the minimum occurs around $(x_1, x_4, x_{11}, x_{15}) = (u, 0, 0, 0)$ or $(0, u, 0, 0)$ (note that the solution also gives an indication of the most likely way $\{\mathcal{L}_1 > u\}$ occurs: X_1 exceeds u *or* X_4 exceeds u). The minimum value is $g(u) = 2\ln(1 + u)$.

In the second step, we need to minimize

$$\frac{1}{(1 - \theta)^8} e^{-2\theta g(u)}$$

for $\theta \geq 0$. For u sufficiently large, this yields

$$\theta \equiv \theta_{u_1} = 1 - \frac{4}{g(u)}.$$

Since $g(u) = \Lambda_1(u)$, θ_{u_1} is of the form consistent with the requirement in Theorem 6. In general, since all $\mathcal{P}_j$ contain at least one Pareto distributed rv, it can be seen that

$$\theta_{u_j} = 1 - \frac{n_j}{\tilde{\Lambda}_1(u)}.$$

Case 2. *Exponential-Tailed with Hazard Rate Twisting.* In this case $X_1, \ldots, X_8$ have an exponential distribution with rate 0.5 and $X_9, \ldots, X_{15}$ have an exponential distribution with rate 1. In this case $\alpha = 1$, $\tilde{\lambda}_i = 0.5$ for $i = 1, \ldots, 8$, and $\tilde{\lambda}_i = 1$ otherwise. Also, $\Gamma_i(\theta) = -\ln(1 - \theta)$ for all $i = 1, \ldots, 15$. The procedure in Section 6 yields

$$\theta_{u_j} = 1 - \frac{n_j}{\lambda_j^* u}$$

(recall that $\lambda_j^* = \min_{i \in \mathcal{P}_j} \tilde{\lambda}_i$).

Case 3. *Superexponential-Tailed with Hazard Rate Twisting.* In this case, $X_1, \ldots, X_8$ have the Weibull$(2, 1)$ distribution and $X_9, \ldots, X_{15}$ have the Weibull$(2, 2)$ distribution. Here $\alpha = 2$, $\tilde{\lambda}_i = 1$ for $i = 1, \ldots, 8$, and $\tilde{\lambda}_i = 4$ otherwise. Also, $\Gamma_i(\theta) = -\ln(1 - \theta)$ for all $i = 1, \ldots, 15$. The procedure in Section 6 yields

$$\theta_{u_j} = 1 - \frac{n_j}{\lambda_j^* u^\alpha},$$

where, recall that λ_j^* is given by (9).

Case 4. *Superexponential-Tailed with Asymptotic Hazard Rate Twisting.* In this case $X_1, \ldots, X_8$ have the $N(1, 0.02)$ distribution and $X_9, \ldots, X_{15}$ have the $N(1, 0.01)$ distribution. In this case $\alpha = 2$, and $\tilde{\lambda}_i = 1/(2\sigma_i^2)$ for $i = 1, \ldots, 15$ (where σ_i^2 denotes the variance of X_i). We twist each component i using the asymptotic hazard function $\tilde{\Lambda}_i(x) = (x - \mu_i)^2/(2\sigma_i^2)$. Hence $\Gamma_i(\theta) = -0.5 \ln(1 - \theta)$ for all $i = 1, \ldots, 15$. The procedure in Section 6 yields

$$\theta_{u_j} = 1 - \frac{n_j}{2\lambda_j^*(u - a_j)^\alpha}, \tag{32}$$

where a_j is the mean of $\mathcal{L}_j$. As an illustration, consider $j = 2$. Here, $n_2 = 3$, $a_2 = \mu_1 + \mu_4 + \mu_{12}$ and $\lambda_2^* = \frac{1}{2(\sigma_1^2 + \sigma_4^2 + \sigma_{12}^2)}$. Then,

$$\theta_{u_2} = 1 - \frac{3(\sigma_1^2 + \sigma_4^2 + \sigma_{12}^2)}{(u - (\mu_1 + \mu_4 + \mu_{12}))^2}.$$

Under asymptotic hazard rate twisting the distribution of each X_i for $i = 1, 4$ and 12, becomes $N(\mu_i, \sigma_i^2/(1 - \theta_{u_2}))$. In particular,

$$(X_1 + X_4 + X_{12}) \sim N\left(\mu_1 + \mu_4 + \mu_{12}, \frac{(u - (\mu_1 + \mu_4 + \mu_{12}))^2}{3}\right)$$

under the new measure. Thus, the mean is unchanged but the variance is increased (so that the probability of exceeding u is no longer rare).

Case 5. *Superexponential-Tailed with Exponential Twisting.* Same as Case 4, but now we apply the exponential-twisting based algorithm of Section 3.3 to each path. For each of the ten paths, we find the θ_{u_j} as the solution to (8). Note that $\mathcal{H}_i(\theta) = \mu_i \theta + \sigma_i^2 \theta^2/2$ for all $i = 1, \ldots, 15$ (where μ_i and σ_i^2 denote the mean and variance of X_i). As an illustration, consider $j = 2$. Here, the solution to (8) yields

$$\theta_{u_2} = (u - (\mu_1 + \mu_4 + \mu_{12}))/(\sigma_1^2 + \sigma_4^2 + \sigma_{12}^2).$$

Applying exponential twisting by amount θ_{u_2} to each X_i in $\mathcal{P}_2$ makes $(X_1 + X_4 + X_{12}) \sim N(u, \sigma_1^2 + \sigma_4^2 + \sigma_{12}^2)$, that is, under the new measure, the mean duration of that path is changed to u but the variance is unchanged.

For each case, we simulate for different values of u, and estimate the probability, the 99% relative error (i.e., the 99% confidence interval half-width upon the quantity to be estimated), the variance reduction factor (VRF) and the efficiency increase factor (EIF). The VRF is the ratio of the variance of the naive simulation estimator to that of the importance sampling estimator; the variance of the naive simulation estimator is estimated by substituting the accurate importance sampling estimator of $P(T > u)$ in $P(T > u)(1 - P(T > u))$. The EIF is the ratio of the expected run-length of naive simulation to that of expected run-length of the importance sampling simulation, for achieving the same relative error. We estimate it by running naive simulation for the same number of replications as the importance sampling simulation, computing the ratio of the two CPU times, and then multiplying it with the VRF. We use MATLAB for the computation. Since MATLAB makes extensive use of matrix methods,

Table I. Estimates of Probabilities, 99% Relative Errors, Variance Reduction Factors (VRF), and Efficiency Increase Factors (EIF) for the Different Cases and for Various u's. The Paired Quantities in Brackets are (VRF, EIF)

Case 1 (Subexp)		Case 2 (Exp)		Case 3 (Superexp)	
u	Estimates (100,000 rep)	u	Estimates (100,000 rep)	u	Estimates (100,000 rep)
10	$2.41 * 10^{-1} \pm 1.1\%$ (1.7, 0.7)	10	$3.62 * 10^{-1} \pm .85\%$ (1.6, .55)	4	$1.36 * 10^{-1} \pm 1.2\%$ (2.7, 1.2)
20	$4.87 * 10^{-2} \pm 2.1\%$ (3.0, 1.3)	15	$5.66 * 10^{-2} \pm 1.7\%$ (3.9, 1.2)	5	$7.69 * 10^{-2} \pm 2.3\%$ (17, 7.7)
50	$4.18 * 10^{-3} \pm 4.6\%$ (7.4, 3.1)	20	$6.80 * 10^{-3} \pm 2.5\%$ (15, 5.2)	5.5	$1.21 * 10^{-3} \pm 2.9\%$ (65, 32)
100	$8.39 * 10^{-4} \pm 7.1\%$ (16, 6.5)	25	$7.14 * 10^{-4} \pm 3.4\%$ (80, 28)	6	$1.57 * 10^{-4} \pm 3.4\%$ (356, 155)
400	$4.70 * 10^{-5} \pm 11\%$ (109, 45)	30	$7.39 * 10^{-5} \pm 4.2\%$ (498, 142)	6.5	$1.61 * 10^{-5} \pm 4.0\%$ (2510, 990)
1600	$3.07 * 10^{-6} \pm 20\%$ (519, 216)	35	$7.00 * 10^{-6} \pm 5.0\%$ (3708, 1117)	7.0	$1.44 * 10^{-6} \pm 4.6\%$ (21444, 9760)

the CPU time is usually not proportional to the number of replications of the simulation. Hence, the EIF numbers (but not the VRF numbers) may change substantially if we change the number of replications.

Table I gives the results for the first three cases. Note that different sets of u's have been selected for the three cases, so that the estimated probabilities lie in similar ranges. The results are as expected; both the VRF and EIF become larger as u increases.

Table II gives the results for Case 4 and Case 5, that apply different simulation algorithms to the same problem. Note that as u increases, exponential twisting far exceeds asymptotic hazard rate twisting in terms of efficiency. Hence for the case of superexponential X_i's, whenever exponential twisting is easily implementable as in the Normal distribution case, we recommend using the exponential-twisting based algorithm of Section 3.3. In cases where exponential twisting is hard to implement (e.g., Case 3), asymptotic hazard rate twisting may be a viable alternative.

7.1 Effect of Problem Size

In this section, through simple examples we test the efficacy of proposed importance sampling techniques as n increases. We consider two extreme cases for different values of n:

(1) $\tau = 1$. In PERT network terminology, this corresponds to all the activities in series.

(2) $\tau = n$. In PERT network terminology, this corresponds to all the activities in parallel.

We refer to the first as the *series* case and to the second as the *parallel* case. We observe that the variance reduction obtained by asymptotic hazard rate twisting deteriorates significantly with n in the series case. The deterioration is insignificant in the parallel case. In contrast, exponential twisting, whenever applicable, is robust in both these dimensions. Recall that in the parallel case,

Table II. Estimates of Probabilities, 99% Relative Errors,
Variance Reduction Factors (VRF), and Efficiency
Increase Factors (EIF) for Normally Distributed Activity
Durations, Using Two Different Methods. The Paired
Quantities in Brackets are (VRF, EIF)

u	Case 4 (Asymp. HRT) Estimates (100,000 rep)	Case 5 (Exp. Twist.) Estimates (100,000 rep)
4.3	$2.57 * 10^{-1} \pm 1.1\%$ (1.5, 1.3)	$2.61 * 10^{-1} \pm .71\%$ (3.8, 1.7)
4.5	$5.61 * 10^{-2} \pm 2.7\%$ (1.6, 1.4)	$5.61 * 10^{-2} \pm 0.97\%$ (12, 5.6)
4.7	$6.31 * 10^{-3} \pm 5.0\%$ (4.2, 3.3)	$6.20 * 10^{-3} \pm 1.17\%$ (78, 35)
4.9	$3.60 * 10^{-4} \pm 9.3\%$ (21, 18)	$3.57 * 10^{-4} \pm 1.3\%$ (1030, 471)
5.0	$7.24 * 10^{-5} \pm 11\%$ (67, 60)	$6.65 * 10^{-5} \pm 1.4\%$ (4960, 2350)
5.2	$1.60 * 10^{-6} \pm 19\%$ (1137, 890)	$1.43 * 10^{-6} \pm 1.6\%$ (187610, 89550)

Table III. Estimates of Variance Reduction Factors (VRF) Using
Exponential Twisting and Asymptotic Hazard Rate Twisting, for
Series and Parallel Cases. The Activity Durations are $N(0, 1)$, and
the u is Selected so that $P(T > u) = 0.0001$. (The VRF in the series
column were analytically calculated.)

n	Series VRF-exp	Series VRF-hazard	Parallel VRF-exp (100,000 rep)	Parallel VRF-hazard (100,000 rep)
1	2388	317	2388	317
5	2388	28	2085	296
10	2388	5	2108	228
50	2388	1	1980	190
100	2388	1	1801	187

both for exponential twisting and asymptotic hazard rate twisting, we use a mixture of appropriate probability distributions.

Specifically, we first consider the case where all the X_i have the $N(0, 1)$ distribution in the series and the parallel settings. We vary n, but at the same time vary u so that $P(T > u)$ remains fixed at 10^{-4}. Results with both aymptotic hazard rate twisting and exponential twisting are presented in Table III. In the series case, it is easy to compute the second moment in each setting analytically, so no simulations were conducted.

We also do the same for the case where all the X_i's have Pareto(2,1) distribution. Results are presented in Table IV. Here, in each setting u was determined through trial-and-error so that $P(T > u)$ is approximately 10^{-4}.

8. CONCLUSION AND FUTURE WORK

In this article, we developed tail asymptotics and fast simulation techniques for estimating the probability that the maximum of sums of a few random variables

Table IV. Estimates of Variance
Reduction Factors (VRF) Using
Asymptotic Hazard Rate Twisting, for
Series and Parallel Cases. The Activity
Durations are Pareto(2,1), and the u is
Selected so that $P(T > u) = 0.0001$

	Series	Parallel
n	VRF-Hazard (100,000 rep)	VRF-Hazard (100,000 rep)
1	875	875
5	68	673
10	14	633
50	1	575
100	1	540

exceeds a large threshold. We considered both light and heavy-tailed random variables and introduced asymptotic hazard rate twisting based importance sampling in these settings. This may have implementation advantages over hazard rate twisting or exponential twisting, although the latter whenever easy to implement, is empirically seen to perform much better. We also noted that straightforward application of exponential twisting and asymptotic hazard rate twisting may not be very effective in estimating $P(T > u)$ when the associated PERT network has many paths. This may be remedied by using a probability measure that is a mixture of probability measures so that for each path there is a probability measure in this mixture that is tailored to it.

Our work focussed on small number of random variables. As mentioned in the introduction, one useful generalization corresponds to extending these results to large networks when the random variables involved are light tailed. Then the tail asymptotics may be developed with the help of large deviations theory. In addition, it may be straightforward to extend the exponential twisting based importance sampling methods discussed here to large networks.

Consider a resource allocation problem where the distribution time of each activity is a function of the resources allocated to it. Further suppose that the objective function corresponds to the probability of large delays. In this setting, the tail asymptotics developed here may serve as a surrogate objective function and facilitate determining the asymptotic solution to this resource allocation problem. The authors explore this in future research.

APPENDIX: PROOFS

PROOFS FOR SECTION 3: LIGHT TAILS, EXPONENTIAL TWISTING

We first prove some lemmas useful in proving Theorem 3.2.

LEMMA 8.1. *Under Assumption* 3.1,

$$\liminf_{u \to \infty} \frac{\log P(S_n > u)}{u^\alpha} \geq -\lambda^*. \tag{33}$$

PROOF.   First consider the case where $\alpha = 1$. Since all X_i's are real valued random variables, there exist constants $(b_i : 2 \leq i \leq n)$ such that for each such i, $P(X_i \geq b_i) > 0$. Thus,

$$P(S_n > u) \geq P\left(X_1 \geq u - \sum_{i=2}^{n} b_i\right) \prod_{i=2}^{n} P(X_i \geq b_i).$$

From this, it easily follows that

$$\liminf_{u \to \infty} \frac{\log P(S_n > u)}{u} \geq -\tilde{\lambda}_1.$$

Similarly, the above inequality holds for right-hand side equal to $-\tilde{\lambda}_i$ for $2 \leq i \leq k$ and hence for $-\lambda^* = -\min_{i \leq k} \tilde{\lambda}_i$.

Now consider $\alpha > 1$. Suppose that $(\beta_i \geq 0 : i \leq k)$ are such that $\sum_{i \leq k} \beta_i = 1$. Again,

$$P(S_n > u) \geq \left[\prod_{i \leq k} P\left(X_i \geq \beta_i\left(u - \sum_{j>k}^{n} b_j\right)\right)\right]\left[\prod_{j>k}^{n} P(X_j \geq b_j)\right],$$

(with the usual convention that if $k = n$, then $\sum_{j>k}^{n} b_j = 0$ and $\Pi_{j>k}^{n} P(X_j \geq b_j) = 1$). Let $\bar{b} = \sum_{j>k}^{n} b_j$. Thus

$$\frac{\log P(S_n > u)}{u^\alpha} \geq \sum_{i \leq k} \frac{\log P(X_i \geq \beta_i(u - \bar{b}))}{u^\alpha} + \sum_{j>k} \frac{\log P(X_j \geq b_j)}{u^\alpha}.$$

Taking $\liminf$ as $u \to \infty$, from Assumption 3.1, it follows that

$$\liminf_{u \to \infty} \frac{\log P(S_n > u)}{u^\alpha} \geq -\sum_{i \leq k} \beta_i^\alpha \lambda_i.$$

Now by setting $\beta_i = 1/\tilde{\lambda}_i^{1/(\alpha-1)} \big/ \sum_{j \leq k} 1/\tilde{\lambda}_j^{1/(\alpha-1)}$ the required relation (33) follows.   □

To prove Theorem 3.2 for $\alpha > 1$ case, we need to develop an asymptotic upper bound for log moment generating function in this setting. To this end, we begin with a simple observation regarding the log-moment generating function. Let $\mathcal{H}$ be the log-moment generating function corresponding to a distribution function F. Using $\exp[\theta s] = 1 + \theta \int_0^s \exp[\theta x] dx$, we get

$$\int_0^\infty \exp[\theta s] dF(s) = \bar{F}(0) + \theta \int_0^\infty \int_0^s \exp[\theta x] dx \, dF(s)$$

$$= \bar{F}(0) + \theta \int_0^\infty \exp[\theta x] \bar{F}(x) dx.$$

Since, for $\theta > 0$, $\int_{-\infty}^0 \exp[\theta s] dF(s) \leq F(0)$, it follows that

$$\exp[\mathcal{H}(\theta)] \leq 1 + \theta \int_0^\infty \exp[\theta x] \bar{F}(x) dx \quad \forall \theta > 0. \tag{34}$$

Let $\log^+(\theta)$ denote $\max(\log \theta, 1)$.

LEMMA 8.2. *Suppose that the tail distribution function $\bar{F}(\cdot)$ of a random variable satisfies the asymptotic relation*

$$\limsup_{u \to \infty} \frac{\log \bar{F}(u)}{u^\alpha} \le -\lambda, \quad 0 < \lambda < \infty \tag{35}$$

for $\alpha > 1$ and $\lambda > 0$. Then for every sufficiently small ϵ there exists a constant C_ϵ and constants C_1 and C_2 that are independent of θ such that

$$\mathcal{H}(\theta) \le \left(\frac{(\theta/\alpha)^{\alpha/(\alpha-1)}}{(\lambda - \epsilon)^{1/(\alpha-1)}} \right) (\alpha - 1) + C_\epsilon + C_1 \log^+(\theta) + C_2 \log^+(\lambda - \epsilon). \tag{36}$$

PROOF. Given $\epsilon > 0$, it follows from (35) that there exists K_ϵ such that

$$\bar{F}(x) \le K_\epsilon \exp[-(\lambda - \epsilon)x^\alpha],$$

for all $x \ge 0$. Using (34) and Lemma 8.3 given below, it follows that

$$\begin{aligned} \mathcal{H}(\theta) &\le 1 + \log^+(\theta) + \log^+ \left(\int_0^\infty \exp[\theta x] \bar{F}(x) dx \right), \\ &\le 1 + \log^+(\theta) + \log^+ \left(K_\epsilon \int_0^\infty \exp[\theta x - (\lambda - \epsilon)x^\alpha] dx \right). \end{aligned}$$

This in turn is upper bounded by

$$1 + \log^+(\theta) + \log^+ K_\epsilon + \left(\frac{(\theta/\alpha)^{\alpha/(\alpha-1)}}{(\lambda - \epsilon)^{1/(\alpha-1)}} \right) (\alpha - 1) + K_1 \log^+(\theta) + K_2 \log^+(\lambda - \epsilon).$$

The result follows from this. □

LEMMA 8.3. *For $\alpha > 1$ and $\gamma > 0$, there exist constants K_1 and K_2 independent of $\theta > 0$ such that*

$$\log \left(\int_0^\infty \exp[\theta x - \gamma x^\alpha] dx \right) \le \left(\frac{(\theta/\alpha)^{\alpha/(\alpha-1)}}{\gamma^{1/(\alpha-1)}} \right) (\alpha - 1) + K_1 \log^+(\theta) + K_2 \log^+(\gamma).$$

PROOF. Consider the integral

$$\int_0^\infty \exp[\theta x - \gamma x^\alpha] dx. \tag{37}$$

Let $\tilde{x} = (\frac{\theta}{\gamma\alpha})^{1/(\alpha-1)}$ so that $\theta = \gamma\alpha\tilde{x}^{\alpha-1}$. From Taylor's expansion, it follows that

$$x^\alpha = \tilde{x}^\alpha + (x - \tilde{x})\alpha\tilde{x}^{\alpha-1} + \frac{(x - \tilde{x})^2}{2}\alpha(\alpha - 1)\xi_x^{\alpha-2},$$

where ξ_x lies between $\tilde{x}$ and x. Thus, (37) may be re-expressed as

$$\exp[\theta\tilde{x} - \gamma\tilde{x}^\alpha] \int_0^\infty \exp\left[-\frac{(x - \tilde{x})^2}{2}\gamma\alpha(\alpha - 1)\xi_x^{\alpha-2} \right] dx. \tag{38}$$

Note that

$$\theta\tilde{x} - \gamma\tilde{x}^\alpha = \left(\frac{(\theta/\alpha)^{\alpha/(\alpha-1)}}{\gamma^{1/(\alpha-1)}} \right) (\alpha - 1).$$

Thus, to complete the proof, it remains to prove that

$$\log\left(\int_0^\infty \exp\left[-\frac{(x-\tilde{x})^2}{2}\gamma\alpha(\alpha-1)\xi_x^{\alpha-2}\right]dx\right) \leq K_1 \log^+(\theta) + K_2 \log^+(\gamma) \quad (39)$$

for constants K_1 and K_2 independent of θ. Now,

$$\int_0^\infty \exp\left[-\frac{(x-\tilde{x})^2}{2}\gamma\alpha(\alpha-1)\xi_x^{\alpha-2}\right]dx \leq 2\tilde{x}$$

$$+ \int_{2\tilde{x}}^\infty \exp\left[-\frac{(x-\tilde{x})^2}{2}\gamma\alpha(\alpha-1)\xi_x^{\alpha-2}\right]dx. \quad (40)$$

Consider the case $\alpha \geq 2$, so that $\xi_x^{\alpha-2} \geq \tilde{x}^{\alpha-2}$ for $x \geq \tilde{x}$. Hence, the integral in the right-hand side of (40) is bounded by

$$\int_{-\infty}^\infty \exp\left[-\frac{(x-\tilde{x})^2}{2}\gamma\alpha(\alpha-1)\tilde{x}^{\alpha-2}\right]dx.$$

This integral equals $\frac{\sqrt{2\pi}}{\sqrt{\gamma\alpha(\alpha-1)\tilde{x}^{\alpha-2}}}$ and hence (39) is true.

Now consider the case where $\alpha < 2$. Here, $\xi_x^{\alpha-2} \geq x^{\alpha-2}$ for $x \geq \tilde{x}$ and $(x-\tilde{x})^2 \geq (x/2)^2$ for $x \geq 2\tilde{x}$ and hence the integral in the right-hand side of (40) is bounded by

$$\int_0^\infty \exp\left[-\frac{x^\alpha}{8}\gamma\alpha(\alpha-1)\right]dx = \gamma^{\frac{-1}{\alpha}}\int_0^\infty \exp\left[-\frac{y^\alpha}{8}\alpha(\alpha-1)\right]dy.$$

Thus, (39) is true. □

PROOF OF THEOREM 3.2. First consider $\alpha > 1$. Fix $\epsilon > 0$. Recall that $\log^+ x$ denotes $\max(\log x, 1)$. From Lemma 8.2 and the fact that $\theta_u \sim \alpha\lambda^* u^{\alpha-1}$, it follows that for $i \leq k$ and u sufficiently large, $\mathcal{H}_i(\theta_u)$ is $\leq$

$$\lambda^{*\frac{\alpha}{\alpha-1}}u^\alpha \frac{1}{(\tilde{\lambda}_i - \epsilon)^{1/(\alpha-1)}}(\alpha-1) + o(u^\alpha) + C_\epsilon + C_1 \log^+(\alpha\lambda^* u^{\alpha-1}) + C_2 \log^+(\tilde{\lambda}_i - \epsilon)$$
$$(41)$$

and for $i > k$, for any $M < \infty$, $\mathcal{H}_i(\theta_u)$ is $\leq$

$$\lambda^{*\frac{\alpha}{\alpha-1}}u^\alpha \frac{1}{(M - \epsilon)^{1/(\alpha-1)}}(\alpha-1) + o(u^\alpha) + C_\epsilon + C_1 \log^+(\alpha\lambda^* u^{\alpha-1}) + C_2 \log^+(M - \epsilon).$$
$$(42)$$

It is easy to select C_ϵ, C_1 and C_2 so that they are independent of i and u. Recall that for u sufficiently large so that $\theta_u > 0$,

$$L_{\theta_u} = \exp\left[-\theta_u S_n + \sum_{i \leq n}\mathcal{H}_i(\theta_u)\right] \leq \exp\left[-\theta_u u + \sum_{i \leq n}\mathcal{H}_i(\theta_u)\right] \quad a.s.$$

It then follows that

$$P(S_n > u) = E_{P_{\theta_u}}L_{\theta_u}I(S_n > u) \leq \exp\left[-\theta_u u + \sum_{i \leq n}\mathcal{H}_i(\theta_u)\right]. \quad (43)$$

Using the estimates in (41) and (42), recalling that $\theta_u \sim \alpha\lambda^* u^{\alpha-1}$, taking the logarithms, dividing by u^α and taking lim sup, we get $\limsup_{u\to\infty}\frac{\log P(S_n > u)}{u^\alpha}$ is

less than or equal to

$$-\alpha\lambda^* + \sum_{i\leq k}\lambda^{*\frac{\alpha}{\alpha-1}}\frac{1}{(\tilde{\lambda}_i-\epsilon)^{1/(\alpha-1)}}(\alpha-1) + \sum_{i>k}\lambda^{*\frac{\alpha}{\alpha-1}}\frac{1}{(M-\epsilon)^{1/(\alpha-1)}}(\alpha-1).$$

This holds for all sufficiently small $\epsilon > 0$ and $M < \infty$. Thus, taking limit as $M \to \infty$ and then limit as $\epsilon \to 0$, it follows that

$$\limsup_{u\to\infty}\frac{\log P(S_n > u)}{u^\alpha} \leq -\alpha\lambda^* + \sum_{i\leq k}\lambda^{*\frac{\alpha}{\alpha-1}}\frac{1}{\tilde{\lambda}_i^{1/(\alpha-1)}}(\alpha-1).$$

The right-hand side above equals $-\lambda^*$ and (6) follows from this and Lemma 8.1. Similarly, by essentially repeating the analysis from (43) onwards, $E_{P_{\theta_u}}[L_{\theta_u}^2 I(S_n > u)]$ may be bounded from above to establish (7).

Now consider $\alpha = 1$. Without loss of generality, let $\lambda_1 = \min_{i\leq k}\tilde{\lambda}_i$. Again,

$$P(S_n > u) = E_{P_{\theta_u}}[L_{\theta_u} I(S_n > u)] \leq \exp\left[-\theta_u u + \sum_{i\leq n}\mathcal{H}_i(\theta_u)\right],$$

and

$$E_{P_\theta}\left[L_{\theta_u}^2 I(S_n > u)\right] \leq \exp\left[-2\theta_u u + 2\sum_{i\leq n}\mathcal{H}_i(\theta_u)\right].$$

Since $\theta_u \sim \lambda^*$ and each $\mathcal{H}_i(\theta_u) = o(u)$, the result follows. □

PROOF OF THEOREM 3.5. From Theorem 3.2, it follows that when $\tilde{\mathcal{P}}_j$ is nonempty, then

$$\lim_{u\to\infty}\frac{\log P(\mathcal{L}_j > u)}{u^\alpha} = -\lambda_j^*.$$

Otherwise, if $\tilde{\mathcal{P}}_j$ is empty, then

$$P(\mathcal{L}_j > u) \leq P\left(\bigcup_{i\in\mathcal{P}_j}\{X_i > u/|\mathcal{L}_j|\}\right) \leq \sum_{i\in\mathcal{P}_j}P(X_i > u/|\mathcal{L}_j|).$$

Since, due to Assumption 3.1, $\frac{\Lambda_i(u/|\mathcal{L}_j|)}{u^\alpha} \to \infty$, it follows that

$$\lim_{u\to\infty}\frac{1}{u^\alpha}\log P(\mathcal{L}_j > u) = -\infty. \tag{44}$$

The proof now follows from the inequalities

$$\max_{j\leq\tau}P(\mathcal{L}_j > u) \leq P(T > u) = P\left(\bigcup_{j\leq\tau}\mathcal{L}_j > u\right)$$

$$\leq \sum_{j\leq\tau}P(\mathcal{L}_j > u)$$

$$\leq \tau\max_{j\leq\tau}P(\mathcal{L}_j > u).$$

Taking logarithms in the above inequalities, dividing by u^α, and taking limits as $u \to \infty$, the result follows. □

PROOF OF THEOREM 3.8. Note again that

$$E_{P^*}[L^2 I(T > u)] \leq \sum_{j \in K} E_{P^*}[L^2 I(\mathcal{L}_j > u)].$$

Thus, it suffices to show that

$$\limsup_{u \to \infty} \frac{\log E_{P^*}[L^2 I(\mathcal{L}_j > u)]}{u^\alpha} \leq -2\lambda_j^*, \tag{45}$$

for all $j \leq \tau$ (recall that $\lambda_j^* = \infty$ if $\tilde{\mathcal{P}}_j$ is empty). To see this, first consider j such that $\tilde{\mathcal{P}}_j$ is nonempty. Again consider the expression for L given in (14). This may be upper bounded by

$$\frac{1}{p_j} \exp\left[-\theta_{u_j} \sum_{i \in \mathcal{P}_j} X_i + \sum_{i \in \mathcal{P}_j} \mathcal{H}_i(\theta_{u_j}) \right]. \tag{46}$$

Thus, (45) follows from the arguments used in proof of Theorem 3.2 specialized to activities in $\mathcal{P}_j$.

Now consider the case j such that $\tilde{\mathcal{P}}_j$ is empty. Note that $E_{P^*}[L^2 I(\mathcal{L}_j > u)] = E_P[LI(\mathcal{L}_j > u)]$ and that L is uniformly upper bounded by $\frac{1}{p_j}$ along the sample paths in the set $\{\mathcal{L}_j > u\}$. Thus,

$$E_P[LI(\mathcal{L}_j > u)] \leq \frac{1}{p_j} P(\mathcal{L}_j > u).$$

The result then follows from (44). □

PROOFS FOR SECTION 4: LIGHT TAILS, HAZARD RATE TWISTING

The following lemma is useful in proving Theorem 4.1.

LEMMA 8.4.   *If $\tilde{\Lambda}_i$ satisfies (18), then*

$$\exp[\Gamma_i(\theta)] \leq p\left(\frac{1}{1-\theta} \right)$$

for $0 < \theta < 1$, where $p(x)$ is a polynomial in x.

PROOF.   We need to show that

$$\int_{-\infty}^{\infty} \exp[\theta \tilde{\Lambda}_i(x)] \lambda_i(x) \exp(-\Lambda_i(x)) dx \tag{47}$$

is upper bounded by a polynomial term in $\frac{1}{1-\theta}$. From (18), we can upper bound this term by

$$\int_{-\infty}^{\infty} \lambda_i(x) \exp[a_i \max(\log \Lambda_i(x), 1) - (1-\theta)\Lambda_i(x)] dx.$$

Now simply set $y = (1-\theta)\Lambda_i(x)$, then $dy = (1-\theta)\lambda_i(x)dx$. Note that $\Lambda_i(-\infty) = 0$ and $\Lambda_i(\infty) = \infty$. Thus, the above integral is upper bounded by

$$\frac{1}{(1-\theta)^{a_i+1}} \int_0^{\infty} \max(\exp[1], y)^{a_i} \exp[-y] dy,$$

establishing the result. □

PROOF OF THEOREM 4.1.   We first establish (21). In view of Lemma 8.1, it suffices to show that

$$\limsup_{u\to\infty} \frac{\log E_{P^*}[L^2 I(S_n > u)]}{u^\alpha} \leq -2\lambda^*.$$

Some notation is useful in displaying intermediate results used to prove (21). For $M > 0$, let

$$A_M = \{S_n > u\} \cap \bigcap_{i>k}\{X_i \leq u/M\},$$

and

$$B_M = \bigcup_{i>k}\{X_i > u/M\}.$$

Note that $\{S_n > u\} \subset A_M \cup B_M$. Equation (21) follows easily from Lemmas 8.5 and 8.6 stated below.

To complete the proof of Theorem 4.1 note that

$$E_{P^*}[L^2 I(T > u)] \leq \sum_{j \leq \tau} E_{P^*}\big[L^2 I(\mathcal{L}_j > u)\big]. \tag{48}$$

Thus, it suffices to show that

$$\limsup_{u\to\infty} \frac{\log E_{P^*}\big[L^2 I(\mathcal{L}_j > u)\big]}{u^\alpha} \leq -2\lambda_j^*.$$

To see this, first consider j such that $\tilde{\mathcal{P}}_j$ is nonempty. Consider again the expression for L given in (19). This may be bounded from above by

$$\exp\left[-\theta_u \sum_{i\in\tilde{\mathcal{P}}_j} \tilde{\Lambda}_i(X_i) + \sum_{i\in\tilde{\mathcal{P}}_j} \Gamma_i(\theta_u) + o(u^\alpha)\right] \ a.s., \tag{49}$$

where we use the fact that $\sum_{i\leq k, i\notin\tilde{\mathcal{P}}_j} \Gamma_i(\theta_u)$ is $\exp[o(u^\alpha)]$ (see Lemma 8.4). The proof follows as in the proof of (21), with the caveat that the additional $\exp[o(u^\alpha)]$ in the likelihood ratio in no way affects that proof.

Now consider the case j such that $\tilde{\mathcal{P}}_j$ is empty. Note that $E_{P^*}[L^2 I(\mathcal{L}_j > u)] = E_P[LI(\mathcal{L}_j > u)]$. Again, note that L is uniformly upper bounded by an $\exp[o(u^\alpha)]$ term on the set $\{\mathcal{L}_j > u\}$. Thus,

$$E_P[LI(\mathcal{L}_j > u)] \leq \exp[o(u^\alpha)]P(\mathcal{L}_j > u).$$

The result then follows from (44).   □

LEMMA 8.5.   *Under Assumption 3.1, (17) and (18),*

$$\lim_{M\to\infty} \limsup_{u\to\infty} \frac{\log E_{P^*}[L^2 I(A_M)]}{u^\alpha} \leq -2\lambda^*$$

LEMMA 8.6.   *Under Assumption 3.1, (17) and (18),*

$$\lim_{M\to\infty} \limsup_{u\to\infty} \frac{\log E_{P^*}[L^2 I(B_M)]}{u^\alpha} \leq -2\lambda^*$$

PROOF OF LEMMA 8.5. Due to Lemma 8.4, $(1-\theta)\Gamma_i(\theta) \to 0$ as $\theta \uparrow 1$, and hence,

$$\frac{\Gamma_i(\theta_u)}{\Theta(u^\alpha)} \to 0 \qquad (50)$$

as $u \to \infty$. In the following discussion we assume that u is sufficiently large so that $0 < \theta_u < 1$. Recall display (19) and note that $A_M \subset \{\sum_{i=1}^k X_i > u(1 - \frac{n-k}{M})\}$. In view of these, it suffices to show that for any positive $\epsilon < 1$, there exists a constant $\tilde{C}_\epsilon$ such that,

$$\exp\left[-\theta_u \sum_{i=1}^k \tilde{\Lambda}_i(X_i) \right] I\left(\sum_{i=1}^k X_i > u\left(1 - \frac{n-k}{M}\right) \right)$$

$$\leq \exp\left[-\lambda^*(1-\epsilon)\theta_u u^\alpha \left(1 - \frac{n-k}{M}\right)^\alpha + \tilde{C}_\epsilon \right] \quad (51)$$

as such. The result then follows by squaring the above equation, taking expectation on both sides with respect to P^* (note that the right-hand side is deterministic), dividing the logarithm of the resultant value of by u^α, taking the limit as $u \to \infty$, and letting $\epsilon \to 0$ and $M \to \infty$ in the right-hand side.

To see (51), note that since $\tilde{\Lambda}_i$ is non-negative and $\tilde{\Lambda}_i(x) \sim \tilde{\lambda}_i x^\alpha$, it follows that for $x \geq 0$, for any positive $\epsilon < 1$, there exists a positive constant C_ϵ such that

$$\tilde{\Lambda}_i(x) \geq \tilde{\lambda}_i x^\alpha(1 - \epsilon) - C_\epsilon.$$

Using this, and letting $\mathcal{I}$ denote the random set of rvs amongst $(X_1, \ldots, X_k)$ that are positive, we can upper bound the left-hand side of (51) with

$$\exp\left[-\theta_u(1-\epsilon) \sum_{i \in \mathcal{I}} \tilde{\lambda}_i X_i^\alpha + kC_\epsilon \right] I\left(\sum_{i=1}^k X_i > u\left(1 - \frac{n-k}{M}\right) \right).$$

Using Jensen's inequality, it is easy to infer that

$$\sum_{i \in \mathcal{I}} \tilde{\lambda}_i x_i^\alpha \geq \frac{1}{\left(\sum_{i \in \mathcal{I}} 1/\tilde{\lambda}_i^{\frac{1}{\alpha-1}} \right)^{\alpha-1}} \left(\sum_{i \in \mathcal{I}} x_i \right)^\alpha.$$

To see this, consider a rv that takes value $\lambda_j^{\frac{1}{\alpha-1}} x_j$ with probability $1/\lambda_j^{\frac{1}{\alpha-1}} / \sum_{i \in \mathcal{I}} 1/\tilde{\lambda}_i^{\frac{1}{\alpha-1}}$ for $j \in \mathcal{I}$, and use the fact that α moment of a non-negative rv is greater than or equal to the first moment raised to power α for $\alpha \geq 1$. Note that

$$\frac{1}{\left(\sum_{i \in \mathcal{I}} 1/\tilde{\lambda}_i^{\frac{1}{\alpha-1}} \right)^{\alpha-1}} \geq \lambda^*.$$

Thus, the left-hand side of (51) may be upper bounded by

$$\exp\left[-\theta_u(1-\epsilon)\lambda^* \left(\sum_{i=1}^k X_i \right)^\alpha + kC_\epsilon \right] I\left(\sum_{i=1}^k X_i > u\left(1 - \frac{n-k}{M}\right) \right),$$

and (51) follows. □

PROOF OF LEMMA 8.6. Note that $E_{P^*}[L^2 I(B_M)] = E_P[LI(B_M)]$. From (50), and the fact that $\tilde{\Lambda}(x) \geq 0$ for all x in (19), it follows that $L \leq \exp[o(u^\alpha)]$ *as such*. Thus, $E_P[LI(B_M)] \leq \exp[o(u^\alpha)]P(B_M)$. It thus suffices to show that

$$\lim_{M \to \infty} \limsup_{u \to \infty} \frac{\log P(B_M)}{u^\alpha} \leq -2\lambda^*.$$

Note that

$$P\left(\bigcup_{i>k} X_i > u/M\right) \leq \sum_{i>k} P(X_i > u/M)$$

and $P(X_i > u/M) = \exp(-\Lambda_i(u/M))$. Hence, it suffices to show that for $i > k$

$$\lim_{M \to \infty} \liminf_{u \to \infty} \frac{\Lambda_i(u/M)}{u^\alpha} \geq 2\lambda^*.$$

From Assumption 3.1, for any M, $\frac{\Lambda_i(u/M)}{(\frac{u}{M})^\alpha}$ increases to infinity. Hence, $\liminf_{u \to \infty} \frac{\Lambda_i(u/M)}{u^\alpha} \geq 2\lambda^*$ and the result follows. $\square$

PROOFS FOR SECTION 5: HEAVY TAILS

PROOF OF THEOREM 5.2. Note that,

$$P(T > u) \geq P\left(\bigcup_{i \leq n}\{X_i > u\}\right)$$

$$\geq \sum_{i \leq n} P(X_i > u) - \sum_{i,j,i \neq j} P(X_i > u)P(X_j > u).$$

Hence, dividing both sides by $P(X_1 > u)$, it is easily seen that

$$\liminf_{u \to \infty} \frac{P(T > u)}{P(X_1 > u)} \geq \sum_{i=1}^{k} \gamma_i.$$

Also note that,

$$P(T > u) \leq P(S_n > u),$$

so from (22), it follows that

$$\limsup_{u \to \infty} \frac{P(T > u)}{P(X_1 > u)} \leq \sum_{i=1}^{k} \gamma_i. \quad \square$$

PROOF OF THEOREM 5.4. Note that the likelihood ratio of P with respect to P^* equals

$$L = \exp\left[-\theta_u \sum_{i=1}^{n} \tilde{\Lambda}_i(X_i) + \sum_{i=1}^{n} \Gamma_i(\theta_u)\right]. \tag{52}$$

In view of (26) and the fact that each $\tilde{\Lambda}_i$ is non-negative, it follows that there exists a constant $D_\epsilon > 0$ such that

$$(1 - \epsilon)\tilde{\Lambda}_1(x) - D_\epsilon \leq \tilde{\Lambda}_i(x),$$

for all i and x. Using this, for u sufficiently large, L may be upper bounded by

$$\exp\left[nD_\epsilon - \theta_u(1-\epsilon)\sum_{i=1}^{n}\tilde{\Lambda}_1(X_i) + \sum_{i=1}^{n}\Gamma_i(\theta_u)\right].$$

In particular, for $u \geq u_\epsilon$, using (27), noting that $\tilde{\Lambda}_1(S_n) \geq \tilde{\Lambda}_1(u)$ when $S_n > u$ for u sufficiently large, it follows that

$$L^2 I(T > u) \leq L^2 I(S_n > u) \leq \exp\left[2nD_\epsilon - 2\theta_u(1-\epsilon)\tilde{\Lambda}_1(u) \;+\; 2\theta_u\epsilon(1-\epsilon) \;+\; 2\sum_{i=1}^{n}\Gamma_i(\theta_u)\right].$$

In particular, the logarithm of the right-hand side upper bounds $\log E_{P^*}[L^2 I(T > u)]$. The result now follows by noting that due to (25) and Lemma 8.4, $\Gamma_i(\theta_u)$ is bounded from above by $\log p[\Theta(\tilde{\Lambda}_1(u))]$ (since $\tilde{\Lambda}_1(u) = 1/(1-\theta_u)$) and so due to (24)

$$\frac{\Gamma_i(\theta_u)}{\Lambda_1(u)} \to 0$$

for each i, $\theta_u < 1$, and that ϵ is arbitrary.   □

REFERENCES

ADALAKHA, V. G. AND KULKARNI, V. G. 1989. A classified bibliography of research on stochastic PERT networks. *INFOR 27,* 3, 272–296.

CROVELLA, M., TAQQU, M. S., AND BESTAVROS, A. 1998. Heavy tails in the world wide web. In *Practical Guide to Heavy Tails*, R. Adler, R. Feldman, and M. S. Taqqu, Eds. Birkhauser, Boston, MA 24–31.

DEMBO, A. AND ZEITOUNI, O. 1998. *Large Deviations Techniques and Applications*, Second ed. Springer, New York, NY.

ELMAGHRABY, S. E. 1977. *Activity Networks: Project Planning and Control by Network Models*. Wiley, New York, NY.

EMBRECHTS, P., KLUPPELBERG, C., AND MIKOSCH, T. 1997. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin, Heidelberg, Germany.

FELLER, W. 1970. *An Introduction to Probability Theory and Its Applications*, Third ed. Vol. 1. Wiley, New York, NY.

HEIDELBERGER, P. 1995. Fast simulation of rare events in queueing and reliability models. *ACM Trans. Model. Comput. Simul. 5,* 1, 43–85.

HUANG, Z. AND SHAHABUDDIN, P. 2004. A unified approach for finite dimensional, rare-event Monte Carlo simulation. In *Proceedings of the 2004 Winter Simulation Conference*, R. Ignalls, M. Rossetti, J. Smith, and B. Peters, Eds. (Piscataway, NJ). 1616–1624.

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

JUNEJA, S. AND SHAHABUDDIN, P. 2006. Rare event simulation techniques. In *Handbook on Simulation*, S. Henderson and B. Nelson, Eds. Elsevier, Amsterdam, The Netherlands, 291–350.

JURECKOVA, J. 1981. Tail behavior of location estimators. *Ann. Stat. 9,* 3, 578–585.

LELAND, W. E., TAQQU, M. S., WILLINGER, W., AND WILSON, D. V. 1994. On the self-similar nature of ethernet traffic. *IEEE/ACM Trans. Netw. 2*, 1–15.

PAKES, A. G. 2004. Convolution equivalence and infinite divisibility. *J. Appl. Prob. 41*, 407–424.

PETROV, V. V. 1975. *Sums of Independent Random Variables*. Springer-Verlag, New York, NY.

PITMAN, E. J. G. 1980. Subexponential distribution functions. *J. Austral. Math. Soc. Ser. A 29*, 337–347.

SADOWSKY, J. S. AND BUCKLEW, J. 1990. On large deviation theory and asymptotically efficient Monte Carlo estimation. *IEEE Trans. Inf. Theory 36,* 3, 579–588.

SIGMAN, K. 1999. A primer on heavy-tailed distributions. *Queu. Syst. 33*, 261–275.