Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

SIMULATING HEAVY TAILED PROCESSES USING DELAYED HAZARD RATE TWISTING

Sandeep Juneja
Indian Institute of Technology
New Delhi 110016, INDIA

Perwez Shahabuddin
Columbia University
New York, NY 10027, U.S.A.

Anurag Chandra
Massachusetts Institute of Technology,
Cambridge, MA 02139, U.S.A.

ABSTRACT

Consider the problem of estimating the small probability that the maximum of a random walk exceeds a large threshold, when the process has a negative drift and the underlying random variables may have heavy tailed distributions. We consider one class of such problems that has applications in estimating the ruin probability associated with insurance claim processes with subexponentially distributed claim sizes, and in estimating the probability of large delays in single server $M/G/1$ queues with subexponentially distributed service times. Significant work has been done on analogous problems for the light tailed case (when the moment generating function exists in a neighborhood around zero, so that the tail decreases at an exponential rate or faster) involving importance sampling methods that use exponential twisting. However, for the subexponential case, moment generating functions do not exist in the pertinent regions making exponential twisting infeasible. In this paper we introduce importance sampling techniques where the new probability measure is obtained by twisting the hazard rate of the original distribution. For subexponential distributions this amounts to twisting at a subexponential rate. We also introduce the technique of "delaying" the change of measure and show that the combination of the two techniques produces asymptotically optimal estimates of the small probabilities mentioned above for a large class of subexponential distributions.

1 INTRODUCTION

Consider a random walk $H_n = \sum_{i=0}^{n} \xi_i$, where each ξ_i can be expressed as a difference of the random variables (rv) B_i and A_i. The sequence of rv $(B_i : i \geq 0)$ is i.i.d. and has a heavy tailed distribution, i.e., a distribution whose tail decays at a subexponential rate. The sequence of rv $(A_i : i \geq 0)$ is also i.i.d., having an exponential distribution with rate λ, and is independent of $(B_i : i \geq 0)$. Assume that $E(\xi_i) < 0$. In this paper we develop techniques for efficient estimation of the small probability $P(\max_n H_n > u)$ as $u \to \infty$.

One application of such probabilities arises in the insurance industry (see, e.g., Embrechts, Kluppelberg and Mikosch (1997)). For many insurance companies it is reasonable to assume that the premium accumulates deterministically at a rate r. The claim sizes $(B_i : i \geq 0)$ are i.i.d. and subexponentially distributed and their inter-arrival times $(A_i : i \geq 0)$ are exponentially distributed. Let u denote the initial reserve of such a company. Then its wealth at time t equals

$$U(t) = u + rt - \sum_{i=1}^{N(t)} B_i,$$

where $N(t) = \sup(n : \sum_{j=1}^{n} A_j \leq t)$, is the Poisson counting process. An important parameter for selecting the premium rate, determining the appropriate level of initial reserve and for managing the risk of the portfolio is the probability of eventual ruin, i.e., $P(U(t) < 0 \text{ for some } t)$. Noting that ruin can occur only at instants of claim arrivals, we can re-express the ruin probability as $P(H_n > u \text{ for some } n) = P(\max_n H_n > u)$, where $\xi_i = B_i - rA_i$.

Another application arises in estimating steady state probability of large delays in a $M/G/1$ queue observing first-come-first-serve rule (see, e.g., Section VI.9 of Feller (1966)). If W_n denotes the waiting time of customer n, B_n denotes its service time and A_n denotes the inter-arrival time between customer n and $n+1$, then it is well known that $W_{n+1} = (W_n + B_n - A_n)^+ = (W_n + \xi_n)^+$. Progressing recursively,

$$W_{n+1} = \max(\xi_n + \cdots + \xi_1, \xi_n + \cdots + \xi_2, \ldots, \xi_n).$$

This implies that W_{n+1} has the same distribution as $\max_{i \leq n} H_n$. Thus, the steady state probability of waiting time exceeding u equals $P(\max_n H_n > u)$. Such probabilities are useful in performance analysis of certain communications networks (see, e.g., Heidelberger (1995)).

The probability $P(\max_n H_n > u)$, typically, cannot be determined analytically. Simulation has proved an effective tool for accurate estimation of such probabilities. However, as is well known, naive simulation takes a prohibitive amount of computational effort to estimate small probabilities. Importance sampling (IS) has been successfully used to efficiently estimate small probabilities when the underlying distributions are light tailed, i.e., when the tail cumulative distribution function (cdf) decreases at an exponential or faster rate and hence the moment generating function is finite in a neighborhood of zero. We refer the reader to Glynn and Iglehart (1989) and Heidelberger (1995) for an introduction to the use of importance sampling and a survey, respectively. It involves simulating the system under a new probability P^* and appropriately modifying the resulting output to get an un-biased estimate of the small probability. The key problem is to select a P^* which significantly reduces the variance of the resulting estimator.

We refer the reader to Siegmund (1976), Asmussen(1985), Lehtonen and Nyrhinen (1992) for IS techniques for the efficient estimation of $P(\max_n H_n > u)$ when the $(B_i : i \geq 0)$ have light tailed distributions. These techniques involve using appropriate "exponentially twisted" distributions for each A_i and B_i (A new distribution F_θ is said to be obtained by exponentially twisting distribution F by an amount θ if $dF_\theta(x) = e^{\theta x} dF(x)/M(\theta)$, where $M(\cdot)$ denotes the moment generating function of distribution F). In particular, B_i needs to be positively twisted (i.e., $\theta > 0$). However, when $(B_i : i \geq 0)$ has a tail that decays at a subexponential rate, then such a exponential twisting is no longer feasible and alternative techniques need to be developed.

Initial research suggests that the efficient simulation for estimating the probability of a random walk exceeding a large threshold, when the underlying variables are non-negative and subexponentially distributed, is feasible. In particular, in Asmussen and Binswanger (1997) the problem of estimating $P(\max_n H_n > u)$ is tackled by using the Pollaczeck-Khinchine formula to represent this probability as $P(\sum_{i=1}^{N} X_i > u)$, where $(X_i : i \geq 1)$ is a sequence of non-negative i.i.d. random variables corresponding to the ladder heights of the random walk $(H_n : n \geq 0)$ and N is a geometrically distributed random variable with parameter $\rho = \lambda E(A_i)$. It is easy to check that $E(\xi_i) < 0$ is equivalent to the condition $\rho < 1$. It is well known that X_i has an integrated tail distribution with cdf $\frac{1}{E(B_i)} \int_0^x \overline{F}_B(y)dy$, where $\overline{F}_B(\cdot)$ denotes the tail cdf of B_i. Typically, X_i is subexponentially distributed if B_i is (see Embrechts, Kluppelberg and Mikosch (1997)). These papers also focus on the easier problem of efficiently estimating $P(\sum_{i=1}^{n} X_i > u)$, where n is fixed, to gain insight towards developing efficient simulation techniques for $P(\sum_{i=1}^{N} X_i > u)$. Let $S_n = \sum_{i=1}^{n} X_i$.

Asmussen and Binswanger (1997) propose a conditioning approach for estimating $P(S_n > u)$ and $P(S_N > u)$ and show that it is "asymptotically optimal" (i.e., as $u \to \infty$) when the tail distribution of X_i decreases at a polynomial rate (i.e., X_i has a Pareto like, regularly varying distribution; see, Feller (1966) for a discussion on regularly varying distributions). Asymptotic optimality (a.o.) is the standard criteria used in the literature to judge the efficiency of rare event simulation and it is reviewed in Section 2. However when the tail distribution of X_i decreases at a subexponential, but a higher rate, as when it has a Weibull distribution with shape parameter less than 1, this approach is not asymptotically optimal (also abbreviated as "a.o." in the rest of the paper). Asmussen, Binswanger and Hojgaard (1998) propose an importance sampling approach which works for such distributions for estimating $P(S_n > u)$. The basic idea in that approach is to use a distribution that has a tail that is (in an informal sense) an order of magnitude heavier than that of X_i. However, for $P(S_N > u)$, it suffers from the drawback that the IS estimator has infinite variance when ρ is sufficiently close to 1.

In this paper we also focus on the problem of developing efficient simulation techniques using IS, to estimate $P(S_n > u)$ and $P(S_N > u)$, albeit our hazard rate approach is different from the approach in Asmussen, Binswanger and Hojgaard (1998). Our major contributions are:

1. We arrive at the IS distribution to efficiently estimate $P(S_n > u)$, by *twisting the hazard rate* of the original distribution. This hazard rate twisted distribution is determined as follows: Suppose that the original probability density function (pdf) of the random variable X_i exists and is represented as $\lambda(x)e^{-\int_0^x \lambda(y)dy}$ where $\lambda(x)$ is its hazard rate, i.e., its the ratio of the pdf and the tail distribution function. Then after twisting the hazard rate by an amount $0 \leq \theta < 1$ the resulting twisted pdf equals $\lambda(x)(1-\theta)e^{-\int_0^x (1-\theta)\lambda(y)dy}$. For example, for a Pareto distribution with pdf

$$\begin{aligned} f(x) &= \frac{\alpha - 1}{x^\alpha}, \quad x \geq 1 \quad (\alpha > 1) \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

   the hazard rate twisting amounts to polynomial twisting leading to the twisted pdf

$$\begin{aligned} f_\theta(x) &= \frac{(\alpha-1)(1-\theta)}{x^{1+(\alpha-1)(1-\theta)}}, \quad x \geq 1, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

   Through appropriate selection of θ as a function of u, with the property that θ increases to 1 as $u \to \infty$, we show that for $(X_i : i \geq 0)$ subexponentially distributed and under mild regularity conditions, $P(S_n > u)$ can be estimated a.o.
2. The above approach, suffers from a drawback similar to that suffered by the IS approach

suggested in Asmussen, Binswanger and Hojgaard (1998); it performs poorly for large values of ρ. To remedy this, we introduce the concept of *delayed hazard rate twisting* to a.o. estimate $P(S_N > u)$ for all values of $\rho < 1$. Again, if $f(\cdot)$ denotes the pdf of the original distribution, then under *delayed hazard rate twisting*, the new pdf

$$\begin{aligned} f_{\theta,x^*}(x) &= f(x) \qquad x < x^*, \\ &= \frac{\overline{F}(x^*)}{\overline{F}_\theta(x^*)} f_\theta(x) \qquad x \geq x^*, \end{aligned} \quad (1)$$

where $\overline{F}_\theta(\cdot)$ denotes the tail cdf corresponding to $f_\theta(\cdot)$ and (θ, x^*) are appropriately selected increasing functions of u. Thus, the tails become heavier from the twisting, but the probability that the random variable takes small values remains unchanged.

3. We generalize the concept of delayed hazard rate twisting to *weighted delayed hazard rate twisting*. Here, the IS distribution has slightly less probability of taking small values compared to the original distribution. The difference is used to make the tails even heavier as compared to the tails under delayed hazard rate twisting. We show experimentally that this gives much larger variance reduction compared to delayed hazard rate twisting.

Section 2 defines subexponential distributons and describes importance sampling and related work in more detail. Sections 3 elaborates on each of the items listed above. Experimental results are presented in Section 4.

2 PRELIMINARIES AND RELATED WORK

2.1 Subexponential Distributions

Let $(X_i : i \leq n)$ be i.i.d. rv's with cdf $F(\cdot)$ and taking values on $(0, \infty)$. Let $\overline{F}(x) = 1 - F(x)$ denote the tail cdf and $F^{*n}(\cdot)$ denote the n-fold convolution of $F(\cdot)$. For any two functions $g_1(u)$ and $g_2(u)$ (that take non-zero values for all sufficiently large u), we use the notation $g_1(u) \sim g_2(u)$ to denote that $\lim_{u\to\infty} g_1(u)/g_2(u) = 1$. Finally, let $X_{(i)}$, $1 \leq i \leq n$ denote the ith order statistics of $(X_1, \ldots X_n)$ with $X_{(n)}$ being the maximum.

A cdf $F(\cdot)$ is said to be subexponentially distributed if

$$\frac{\overline{F^{*n}}(u)}{n\overline{F}(u)} \sim \frac{P(S_n > u)}{nP(X_1 > u)} \sim 1 \qquad (2)$$

for all n. In fact, it suffices to show that (2) holds for $n = 2$. Thus subexponentiality implies that $P(S_n > u) \sim n\overline{F}(u)$. It can also be shown that for case where N is a geometric random variable with parameter $1 - \rho$ (i.e. $P(N = i) = (1 - \rho)\rho^i$ for $i = 0, 1, 2 \ldots$),

$$P(S_N > u) \sim \frac{\rho}{1-\rho}\overline{F}(u)$$

(see, e.g., Embrechts, Kluppelberg and Mikosch (1997)). These asymptotic results are used in proving a.o. in the estimation of these quantities. Also note that $P(X_{(n)} > u) \sim nP(X_1 > u)$, indicating that on the set $\{S_n > u\}$, we have $S_n \approx X_{(n)}$.

To develop our hazard rate twisting framework, we assume that X is a continuous random variable with a pdf given by $f(x)$. As mentioned in the Introduction, $\lambda(x) = f(x)/\overline{F}(x)$ denotes the hazard rate. Let $\Lambda(x) = \int_0^x \lambda(y)dy$ denote the hazard function. Then, $f(x) = \lambda(x)e^{-\Lambda(x)}$ and $\overline{F}(x) = e^{-\Lambda(x)}$. In Pitman(1980), it is shown that amongst the distributions whose hazard rate $\lambda(x)$ eventually decreases to 0, a necessary and sufficient condition for a distribution to belong to the sub-exponential class is that:

$$\lim_{u\to\infty} \int_0^u \lambda(x)e^{x\lambda(u)-\Lambda(x)}dx = 1. \qquad (3)$$

This is useful in checking membership to the subexponential class. For a detailed exposition on subexponential distributions the reader is referred to Embrechts, Kluppelberg and Mikosch (1997).

Let $S(\cdot)$ be a slowly varying function, i.e., $\lim_{x\to\infty} S(tx)/S(x) = 1$ for all $t > 0$. In our analysis, we need additional mild regularity conditions on the distribution of the rv X_i; the complete set of assumptions that we make on the distribution of X_i are as follows:

Assumption 1 *The hazard rate is eventually decreasing and eventually everywhere differentiable. Furthermore, it has the form,*

$$\lambda(x) = S(x)/x^\beta,$$

for $0 < \beta \leq 1$.

Assumption 2 *For* $\beta = 1$*, there exists a* $p > 0$ *such that*

$$\Lambda(x) \geq [\log(x)]^p$$

for all x *large enough.*

It can be checked that the commonly used families of subexponential distributions: Pareto, Weibull with decreasing failure rate (i.e., $\overline{F}(x) = e^{-\lambda x^\alpha}$, for $\alpha < 1$) and log-normal distributions, satisfy these assumptions. Also, Assumption (1) is sufficient for (3) to hold and hence it guarantees subexponentiality. The limitations imposed by these assumptions are minor from practical viewpoint and are discussed in Juneja, Shahabuddin and Chandra (1999).

As mentioned in Section 1, we are interested in estimating $\mu_n(u) \equiv P(S_n > u) = E(1_{\{S_n>u\}})$ for n fixed, and $\mu_N(u) \equiv P(S_N > u) = E(1_{\{S_N>u\}})$ where $1_{\{\cdot\}}$ is an indicator random variable. The basic ideas in this paper can be easily extended to the case where N has a general distribution that is bounded by a geometrically decreasing function.

Typically, probabilities such as $\mu_n(u)$ and $\mu_N(u)$ are efficiently estimated by developing an alternate representation of these probabilities as $E_{P^*}(Z)$, where Z is a random variable whose expectation under the probability measure P^* (that may be same as P) equals the value of the desired probability but whose variance is much smaller than the variance of the naive estimator, i.e., the random variable $1_{\{S_n>u\}}$ under probability P (for estimating $\mu_n(u)$). Then, many independent samples of Z under P^* are generated, and the average of the samples provides a point estimator of the desired probability. The variance is also estimated from the given sample and the normal approximation provided by the central limit theorem is used to construct confidence intervals.

We are interested in cases when u is large, i.e., when the probabilities are very small. A standard criteria to judge whether any rare event probability estimator is efficient or not for large values of u, is a.o.. Let $\sigma_{P^*}{}^2(Z)$ denote the variance of rv Z, when generated using the probability P^*.

Definition 1 *The estimator corresponding to Z is said to be a.o., iff*

$$\liminf_{u\to\infty} \frac{\log(\sigma_{P^*}(Z))}{\log(E_{P^*}(Z))} \geq 1. \tag{4}$$

One can easily show that the naive simulation estimator is not a.o. as in that case, the ratio in (4) tends to 1/2.

2.2 Importance Sampling and Related Work

Let $G(\cdot)$ denote a distribution function with pdf $g(\cdot)$ such that $g(x) > 0$ if $f(x) > 0$. Let P^* denote the resultant probability measure on $(X_1, \ldots X_n)$ when each X_i has the distribution $G(\cdot)$. The probability $\mu_n(u)$ may be re-expressed as follows:

$$\begin{aligned}\mu_n(u) &= E_P(1_{\{S_n>u\}}) \\ &= E_{P^*}(1_{\{S_n>u\}} L(X_1, \ldots, X_n))\end{aligned}$$

where $L(X_1, \ldots X_n) \equiv \prod_{i=1}^n f(X_i)/g(X_i)$ is called the likelihood ratio. Hence, in this case, $Z = 1_{\{S_n>u\}} L(X_1, \ldots, X_n)$, generated using P^*, is the new estimator for $\mu_n(u)$. Similarly, the estimator for the probability $\mu_N(u)$ is $Z = 1_{\{S_N>u\}} L(X_1, \ldots, X_N)$, where N is generated from a geometric distribution with parameter $1-\rho$ and $(X_i : i \geq 0)$ are generated using P^*. The key issue is to select a P^* (or equivalently, the $G(\cdot)$) so that the variance of Z is reduced, and whenever possible, a.o. is achieved.

Asmussen, Binswanger and Hojgaard (1998) propose that the IS pdf $g(\cdot)$ have a very heavy tail so that the following condition is satisfied:

Condition 1 *The distribution K with pdf $k(x) = \frac{f^2(x)}{cg(x)}$, where $c = \int_0^\infty \frac{f^2(x)}{g(x)} dx$, is subexponential and its tail cdf $\overline{K}(x)$ is of order $\overline{F}(x)^2$ in the sense that $\log \overline{K}(x) \sim \log \overline{F}(x)^2$.*

Once this condition holds then

$$E_{P^*}(Z^2) = c^n P_K(S_n > u) \sim nc^n P_K(X > u)$$

and a.o. follows.

In particular, they propose the following IS pdf to be used to simulate $(X_i : i \leq n)$:

$$g(x) = \frac{1}{(x+e)\log(x+e)^2} \quad \text{for} \quad x \geq 0.$$

Note that this distribution has a very heavy tail as its tail distribution function $\overline{G}(x) = 1/\log(x+e)$ decreases at a logarithmic rate, and the first moment does not exist. They show that for many commonly used distributions the above density satisfies condition 1 and hence a.o. estimates $\mu_n(u)$. Intuitively, one may see this by noting that on the set $\{S_n > u\}$, typically $X_{(n)} > u$, and $X_{(i)}$ for $(i \leq n-1)$ are small and insensitive to u. The rv Z, on this set equals the product of the ratio $f(X_{(i)})/g(X_{(i)})$ for $(i \leq n)$. Now, the ratio $f(X_{(n)})/g(X_{(n)})$ is small and decreases as u increases. The other ratios, $f(X_{(i)})/g(X_{(i)})$ for $(i \leq n-1)$, typically, are each greater than 1, but are relatively unaffected as u increases to infinity. Hence, for u large enough, with high probability, Z is small, and with little variation, on the set $\{S_n > u\}$.

For the case of $\mu_N(u)$, it is easily seen that $E(Nc^N)$ needs to be finite for a.o.. In particular, $c < 1/\rho$. This, creates problems for large values of ρ. Intuitively, this can be seen by the fact that the value of the product $\Pi_{i=1}^{n-1}[f(X_{(i)})/g(X_{(i)})]^2$ increases geometrically with n, while its relative frequency of occurrence $\approx \rho^n(1-\rho)$ decreases at a lower geometric rate, leading to infinite second moment. In the next section we develop the hazard rate twisting framework and show that "delayed" hazard rate twisting overcomes this problem of infinite variance for large values of ρ.

3 HAZARD RATE TWISTING AND DELAYED CHANGES OF MEASURE

3.1 Hazard Rate Twisting

Consider again the pdf $f(x) = \lambda(x)e^{-\Lambda(x)}$. As, mentioned in the Introduction, we cannot select a new pdf by positive exponential twisting, i.e., such that its value at x is proportional to $e^{\theta x}\lambda(x)e^{-\Lambda(x)}$ for some $\theta > 0$, as the normalisation constant, $\int_0^\infty e^{\theta x}\lambda(x)e^{-\Lambda(x)}$ is infinite. However, subexponential twisting such that the new pdf is proportional to $e^{\theta\Lambda(x)}\lambda(x)e^{-\Lambda(x)}$ can be performed as then the normalization constant

$$\int_0^\infty e^{\theta\Lambda(x)}\lambda(x)e^{-\Lambda(x)} = 1/(1-\theta),$$

is finite for $\theta < 1$. This results in the new hazard rate twisted pdf

$$f_\theta(x) = (1-\theta)\lambda(x)e^{-(1-\theta)\Lambda(x)}, \qquad x \geq 0.$$

Consider the problem of estimating $\mu_n(u)$ using IS density $f_\theta(\cdot)$ to generate the X_i's. In this case

$$\begin{aligned} Z &= \frac{f(X_1)}{f_\theta(X_1)}\cdots\frac{f(X_n)}{f_\theta(X_n)}1_{\{S_n>u\}} \\ &= \frac{1}{(1-\theta)^n}e^{-\theta\sum_{i=1}^n\Lambda(X_i)}1_{\{S_n>u\}}. \end{aligned} \tag{5}$$

The following lemma gives a lower bound for $\sum_{i=1}^n \Lambda(X_i)$ in (5). This is useful in upper bounding Z.

Lemma 1 *Consider any set of non-negative numbers $(x_1, \ldots x_n)$. Under Assumption (1), for every $\epsilon > 0$ and for $\sum_{i=1}^n x_i$ large enough, the hazard function satisfies the following "asymptotic concavity" property:*

$$\sum_{i=1}^n \Lambda(x_i) \geq \Lambda(\sum_{i=1}^n x_i) - \epsilon. \tag{6}$$

Using (6) we get the following bound on Z, on the set $\{S_n > u\}$, for u sufficiently large and $0 \leq \theta < 1$:

$$Z \leq \frac{1}{(1-\theta)^n}e^{-\theta(\Lambda(u)-\epsilon)} \leq \frac{1}{(1-\theta)^n}e^{-\theta\Lambda(u)+\epsilon}.$$

Differentiating the RHS, we find that $\theta = 1 - n/\Lambda(u)$ achieves the minimum upper bound. Using this bound we see that,

$$E_{P^*}(Z^2 1_{\{S_n>u\}}) \leq e^{2\epsilon}(\frac{e\Lambda(u)}{n})^{2n}e^{-2\Lambda(u)}.$$

From the fact that $\mu_n(u) \sim n\overline{F}(u) = ne^{-\Lambda(u)}$, it follows that the hazard rate twisting with $\theta = 1 - n/\Lambda(u)$ is a.o. (One can also show that the hazard rate twisting is also a.o. for $\theta = 1 - b/\Lambda(u)$ for all constants $b > 0$.) However, this approach does not work quite so well, for estimating $\mu_N(u)$. In that case, since the probability that $N = n$ for $n \geq 1$, equals $\rho^n(1-\rho)$, the unconditioned bound on $E_{P^*}(Z^2 1_{\{S_N>u\}})$ equals

$$e^{2\epsilon}e^{-2\Lambda(u)}(1-\rho)\sum_{n=1}^\infty(\frac{e\Lambda(u)}{n})^{2n}\rho^n.$$

Using Stirling's formula, we see that $(2n/e)^{2n} \sim (2n-1)!/2\sqrt{\pi n}$ and then

$$\sum_{n=1}^\infty(\frac{e\Lambda(u)}{n})^{2n}\rho^n$$

is close to

$$\sum_{n=1}^\infty 2\sqrt{\pi n}\frac{(2\Lambda(u)\sqrt{\rho})^{2n}}{(2n-1)!}.$$

This term, in turn is bounded by an appropriately selected constant, K_0, times $\Lambda^2(u)\sum_{n=0}^\infty \frac{(2\Lambda(u)\sqrt{\rho})^{2n}}{(2n)!}$, and therefore by $K_0\Lambda^2(u)e^{2\Lambda(u)\sqrt{\rho}}$. This implies that the variance is finite even as $u \to \infty$, for all $\rho < 1$. Also,

$$\lim_{u\to\infty}\frac{\log(\sigma_{P^*}(Z))}{\log(\mu(u))} \geq 1 - \sqrt{\rho}.$$

Thus a significant variance reduction over naive simulation may be expected for small values of ρ. However for $\rho > 0.25$ it is likely to (asymptotically) performs worse than standard simulation. In the next section we tackle this problem using delayed hazard rate twisting.

3.2 Delayed Hazard Rate Twisting

The delayed hazard rate twisted pdf $f_{\theta,x^*}(\cdot)$ has the form shown in (1). We now show that delayed hazard rate twisting for an appropriately selected (x^*, θ), leads to a.o. estimation of $\mu_N(u)$ when $X_i's$ satisfy Assumptions (1) and (2). Set $\theta = \theta_u = 1 - b/\Lambda(u)$ where b is a positive constant. Let $x^* = x_u^*$ where x_u^*, for sufficiently large u, is the unique solution to the equation

$$\Lambda(x) = 2\log\Lambda(u) - \log a, \tag{7}$$

where $a > 0$ is any constant so that $(1+a)\rho < 1$. It follows that $x_u^* \to \infty$ as $u \to \infty$. Let P^* denote the probability measure under which N has Geometric$(1-\rho)$ distribution and $(X_i : i \geq 0)$ have pdf $f_{\theta_u, x_u^*}(\cdot)$. As before, let Z be the product of the likelihood ratio and the indicator $1_{\{S_N > u\}}$.

Theorem 1 *For $(X_i : i \geq 0)$ with hazard rates that satisfy Assumptions (1) and (2), the IS technique using P^* asymptotically optimally estimates $\mu_N(u)$, i.e.,*

$$\lim_{u \to \infty} \frac{\log(\sigma_{P^*}(Z))}{\log(\mu_N(u))} = 1.$$

The strategy for proving the above theorem is to partition the set $\{S_N > u\}$ into subsets and then bound the likelihood ratio over sets of significant probability and show that sets where likelihood ratios are large have small probability. In particular, for $n \geq 1$ and $k \leq n$ let

$$A_{n,k} = \{N = n, S_N > u, X_{(k)} < x^*, X_{(k+1)} \geq x^*\}.$$

where for convenience we define $X_{(0)} = 0$ and $X_{(N+1)} = \infty$. Also let $A_n = \cup_{k=0}^{k=n} A_{n,k}$. Then

$$\{S_N > u\} = \cup_{n=1}^{\infty} A_n.$$

Another version of the inequality in Lemma 1 is used for the bounding of likelihood ratios. The detailed proof is given in Juneja, Shahabuddin and Chandra(1999).

In most cases we want to build confidence intervals for the estimators, i.e., we need to estimate the variance of the importance sampling estimator. Hence in addition to ensuring that the estimator has a small variance, we also need to ensure that the variance of the variance estimator is small. The problem with x_u^* selected using (7) is that, while $E_{P^*}(Z^2)$ is small, the fourth moment $E_{P^*}(Z^4)$, which governs the variance of the variance can be seen to be infinite. Hence, in practice we recommend that x_u^* be selected so that $\Lambda(x_u^*) = 4 \log \Lambda(u) - \log a$. Repeating the above analysis with a view to upper bounding $E_{P^*}(Z^4)$, it can be seen that our new choice of x_u^* gives a.o. both in the estimation of $\mu_N(u)$ and $E_{P^*}(Z^2)$.

However, there is a cost involved in selecting a higher x_u^*. This is discussed in greater detail in Juneja, Shahabuddin and Chandra(1999). The next section discusses an improved version of delayed hazard rate twisting which reduces the required x_u^*.

3.3 Weighted Delayed Hazard Rate Twisting

We now consider a more general method of delayed hazard rate twisting where the new probability of taking values less than x_u^* is less than the original probability of taking these values. The difference is used to make the tails heavier than in the case of delayed twisting. Let w be a constant. Then the new measure is given by

$$\begin{aligned} f_{\theta_u, x_u^*}(x) &= \frac{f(x)}{1+w} \quad \text{for} \quad x \leq x_u^* \\ &= \left(1 - \frac{F(x_u^*)}{1+w}\right) \frac{f_\theta(x)}{\overline{F}_\theta(x_u^*)} \quad \text{for} \quad x > x_u^*. \end{aligned}$$

Note that $w = 0$ in the above expression gives ordinary delayed hazard rate twisting. We now consider the case where

$$0 < w < \frac{1}{\rho} - 1. \tag{8}$$

We use the same θ_u as before. Let $0 < a < 1/[\rho(1+w)] - 1$ and let x_u^* be defined by

$$\Lambda(x_u^*) = \log \Lambda(u) - \frac{1}{2}\log(aw).$$

Note that x_u^* has similar asymptotic properties as under delayed twisting. It is also easy to check that $P_{\theta, x_u^*}(X > x_u^*) = 1 - F(x_u^*)/(1+w) \sim w/(1+w)$. This is contrast to ordinary delayed hazard rate twisting where this probability equaled $e^{-\Lambda(x_u^*)}$ and decreased to 0 as $u \to \infty$. Also note that $P_{\theta, x_u^*}(X > u) \sim w/(e(1+w))$. Hence in general we are giving more weight to the tail of the distribution than before and thus making the rare event happen more frequently. The penalty we pay is that now we incur a likelihood ratio $(1+w) > 1$ when $X < x_u^*$, which we did not do before. However, due to the upperbound in (8), the geometric increase in this penalty with n is absorbed by the larger geometric decrease in the frequency of observing large N.

The proof of the following theorem is given in Juneja, Shahabuddin and Chandra (1999):

Theorem 2 *For $(X_i : i \geq 0)$ with hazard rates that satisfy Assumptions (1) and (2), the IS technique using the weighted delayed hazard twisting distribution a.o. estimates $\mu_N(u)$.*

Again for variance estimation considerations, we select

$$0 < w < \frac{1}{\rho^{1/3}} - 1$$

determine x_u^* using

$$\Lambda(x_u^*) = \log \Lambda(u) - \frac{1}{4}\log(aw^3)$$

where now a satisfies $0 < a < 1/[\rho(1+w)^3] - 1$. Even though both delayed hazard rate twisting and weighted delayed hazard rate twisting give a.o. estimates of $\mu_N(u)$ and $E_{P^*}(Z^2)$, the variance $\sigma^2_{P^*}(Z)$ in weighted delayed hazard rate twisting seems to be orders of magnitude less. This is discussed further in Juneja, Shahabuddin and Chandra (1999). It is also evident from the experiments in the next section.

4 EXPERIMENTAL RESULTS

We consider the case where $f(x)$ is the Weibull density with rate parameter 1 and scale parameter 0.5. We consider several values of u and ρ. For each case we estimate the probability and the relative error (99% confidence interval half width divided by the quantity that is estimated) using delayed and weighted delayed hazard rate twisting. We used 10,000,000 replications for each case. We also compute the variance reduction obtained over standard simulation in each case. For this we use the accurate estimates of $\mu_N(u)$, obtained using weighted delayed hazard rate twisting, to estimate the variance of the standard simulation estimator that uses the same run-length; the latter is given by $\mu_N(u)(1-\mu_N(u))/10,000,000$.

For the ordinary delayed case we use $b = 1$ and $a = (1/2\rho) - 0.5$. For the weighted delayed case we choose $b = 1$, $a = (1/2\rho^{1/4}) - 0.5$, and $w = a$. These values of a and w satisfy the conditions given previously for the case when we also desire good estimates for the variance. Table 1 gives the a and x_u^* computed for various values of ρ and u when we use the two methods. Note that x_u^* increases as u increases but at a much slow rate for the weighted delayed case as compared to the plain delayed case.

Table 2 gives the results of the simulation using the two methods with the parameter values in Table 1. Notice how ordinary delayed hazard rate twisting gives substantial improvement over standard simuation. However, the RE grows significantly as u becomes larger. But weighted delayed hazard rate twisting does much better and the RE using this method remains almost bounded.

Table 1: Values of a and x^* corresponding to Delayed Hazard Rate Twisting (D) and Weighted Delayed Hazard Rate Twisting (W) for different ρ and u. The b was 1 in each case and the w in W was the same as a.

		Values of (a, x^*)		
u	Me.	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$
100	D	(1.5, 77.5)	(0.5, 98.1)	(.166, 121)
	W	(.207, 15.0)	(.0946, 21.7)	(.0373, 31.3)
200	D	(1.5, 104)	(0.5, 127)	(.166, 153)
	W	(.207, 17.8)	(.0946, 25.1)	(.0373, 35.3)
400	D	(1.5, 134)	(0.5, 161)	(.166, 190)
	W	(.207, 20.9)	(.0946, 28.7)	(.0373, 39.5)
800	D	(1.5, 168)	(0.5, 198)	(.166, 230)
	W	(.207, 24.2)	(.0946, 32.5)	(.0373, 44.0)

Table 2: Estimates of probabilities and relative errors using Delayed Hazard Rate Twisting (D) and Weighted Delayed Hazard Rate Twisting (W) for different ρ and u. For each case 10,000,000 replications were simulated. The quantities in brackets are estimated variance reduction factors over standard simulation. The estimates of the probabilities, relative errors and variance reduction factors for method D with $u = 400, 800$ are not reliable, as they did not sufficiently converge in the given number of replications.

		Estimates of Probabilities	
u	ρ	D	W
100	0.25	$1.84E{-5} \pm 27\%$	$1.68E{-5} \pm 1.1\%$
		(0.51)	(320)
	0.5	$6.76E{-5} \pm 20\%$	$6.40E{-5} \pm 1.3\%$
		(.24)	(67)
	0.75	$4.39E{-4} \pm 4.7\%$	$4.59E{-4} \pm 2.3\%$
		(.69)	(2.7)
200	0.25	$1.48E{-7} \pm 84\%$	$2.55E{-7} \pm 1.4\%$
		(6.3)	(1.4E+4)
	0.50	$1.06E{-6} \pm 59\%$	$8.94{-7} \pm 1.2\%$
		(1.8)	(5272)
	0.75	$4.68E{-6} \pm 55\%$	$4.55E{-6} \pm 2.3\%$
		(.48)	(288)
400	0.25	$2.40E{-10} \pm 153\%$	$7.04E{-10} \pm 1.7\%$
		(1179)	(3.3E+6)
	0.5	$2.09E{-9} \pm 182\%$	$2.33E{-9} \pm 1.4\%$
		(96)	(1.4E+6)
	0.75	$6.71E{-9} \pm 175\%$	$9.49E{-9} \pm 1.6\%$
		(32)	(2.8E+5)
800	0.25	$1.86E{-14} \pm 171\%$	$1.77E{-13} \pm 2.1\%$
		(1.2E+7)	(8.7E+9)
	0.50	$4.69E{-14} \pm 258\%$	$5.62E{-13} \pm 1.7\%$
		(2.1E+6)	(3.9E+9)
	0.75	$4.83E{-13} \pm 209\%$	$2.02E{-12} \pm 1.7\%$
		(3.1E+5)	(1.2E+9)

5 ACKNOWLEDGEMENT

This work was partially supported by NSF Career Award Grant DMI-96-25297. The work of Anurag Chandra was performed while he was at the Indian Institute of Technology, Delhi.

REFERENCES

Asmussen, S. 1985. Conjugate processes and the simulation of ruin problems. *Stochastic Processes and Applications* 20, 213-229.

Asmussen, S., and Binswanger, K. 1997. Simulation of ruin probabilities for subexponential claims. *ASTIN BULLETIN* 27, 2, 297-318.

Asmussen, S., Binswanger, K., and Hojgaard, B. 1998. Rare events simulation for heavy tailed distributions.

Research Report, Dept. of Mathematical Statistics, Lund University, Box 118, SE-22100 Lund, Sweden.

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

Feller, W. 1966. *An Introduction to Probability Theory and its Applications Volume II*. John Wiley & Sons, Inc..

Glynn, P.W., and Iglehart, D.L. 1989. Importance sampling for stochastic simulations. *Management Science* 35, 11, 1367-1393.

Heidelberger, P. 1995. Fast Simulation of Rare events in queueing and reliability models. *ACM Transactions on Modeling and Computer Simulation* 5, 1, 43-85.

Juneja, S., Shahabuddin, P., and Chandra, A. 1999. Rare event simulation of heavy tailed risk processes using hazard rate twisting and delayed change of measure. Research Report, Department of IEOR, Columbia University, New York, NY 10027, USA.

Lehtonen, T., and Nyrhinen, H. 1992. Simulating level-crossing probabilities by importance sampling. *Advances in Applied Probability* 24, 858-874.

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

Siegmund, D. 1976. Importance sampling in the Monte Carlo study of sequential tests. *The Annals of Statistics* 4, 673-684.

AUTHOR BIOGRAPHIES

SANDEEP JUNEJA is an Assistant Professor in the Mechanical Engineering Department at the Indian Institute of Technology (IIT) Delhi. Prior to joining the faculty at IIT he was a Senior Consultant at Andersen Consulting, India (1995-96), and Director Quantitative Analysis at American Credit Indemnity, U.S.A.(1993-95). He received a B.Tech. in Mechanical Engineering from IIT Delhi (1989) and a Masters in Statistics and Ph.D. in Operations Research from Stanford University (1993).

PERWEZ SHAHABUDDIN is an Associate Professor in the Industrial Engineering and Operations Research Department at Columbia University. From 1990 to 1995, he was a Research Staff Member at the IBM T.J. Watson Research Center. He received a B.Tech. in Mechanical Eng. from the Indian Institute of Technology, Delhi (1984), and a M.S. in Statistics and a Ph.D. in Operations Research from Stanford University (1990).

ANURAG CHANDRA is in the Ph.D. program in Operations Research at the Sloan School of Management, Massachusetts Institute of Technology. He obtained his B.Tech. in Computer Science from the Indian Institute of Technology, Delhi, in 1999.