Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.

SIMULATING RUIN PROBABILITIES IN INSURANCE RISK PROCESSES WITH SUBEXPONENTIAL CLAIMS

Nam Kyoo Boots

Department of Economics
& Business Administration
Vrije Universiteit
De Boelelaan 1105
Amsterdam, 1081 HV, THE NETHERLANDS

Perwez Shahabuddin

Department of Industrial Engineering
& Operations Research
Columbia University
116th street & Broadway
New York, NY 10027, U.S.A.

ABSTRACT

We describe a fast simulation framework for simulating small ruin probabilities in insurance risk processes with subexponential claims. Naive simulation is inefficient since the event of interest is rare, and special simulation techniques like importance sampling need to be used. An importance sampling change of measure known as sub-exponential twisting has been found useful for some rare event simulations in the subexponential context. We describe conditions that are sufficient to ensure that the infinite horizon probability can be estimated in a (work-normalized) *large set asymptotically optimal* manner, using this change of measure. These conditions are satisfied for some large classes of insurance risk processes – e.g., processes with Markov-modulated claim arrivals and claim sizes – where the heavy tails are of the 'Weibull type'. We also give much weaker conditions for the estimation of the finite horizon ruin probability. Finally, we present experiments supporting our results.

1 INTRODUCTION

Consider an insurance risk process. Let X_n denote the size of claim n which arrives at epoch η_n, $n \geq 1$, with $0 \equiv \eta_0 < \eta_1 < \eta_2 < \cdots$. Define ξ_n as the interarrival time between claim n and claim $n-1$, i.e., $\xi_n := \eta_n - \eta_{n-1}$, $n \geq 1$. The premium rate is c, so in the absence of claims the insurance reserve builds up at that rate. Let $N(t)$ be the random number of arrivals in $(0, T]$. If we let $U(t)$ be the reserve at time t, then

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

where u is the reserve at time 0. Finally, let $S(t)$ be the claim-surplus process defined by $u - U(t)$. We are interested in the finite horizon ruin probability given by

$$\begin{aligned}\psi(u,T) &:= \mathbb{P}\,(U(t) < 0 \text{ for some } t < T)\\ &= \mathbb{P}\left(\sup_{t<T} S(t) > u\right), \end{aligned} \qquad (1)$$

and the infinite horizon ruin probability defined by $\psi(u) = \psi(u, \infty)$. Unless we specifically state the $T = \infty$ in the notation $\psi(u, T)$, take $T < \infty$ in that notation.

We assume very general conditions on the random drivers of this process, i.e., the sequence $\{(\xi_i, X_i) : i \geq 1\}$. The claim sizes X_i's can be one of m different types, corresponding to m different distributions, denoted by F_1, F_2, …, F_m. We assume that at least one of them is subexponentially distributed. Let the (random) type of the claim that arrives at time ξ_i be denoted by W_i. The $\{(\xi_i, W_i) : 1 \geq 1\}$ can be any general discrete time stochastic process. However, given $\{(\xi_i, W_i) : i \geq 1\}$, we assume that the X_i has distribution F_{W_i}, and that the X_i's are independent of one another. This includes the renewal arrrivals model, which is the most basic sub-case. In this model the interarrival times are i.i.d., the claim sizes are i.i.d. and the interarrival times are independent of the claim-size sequence. It also includes cases where the claim sizes and the interarrival times are environment dependent and the state of the environment fluctuates randomly. One example of the latter to which our techniques apply, is when the state of the environment is given by a finite state, continuous time Markov chain, arrivals are Poisson with the rate dependent on the state, and the distribution of the claim sizes are also dependent on the state (given the states, they are independent of one another and the arrival stream).

Let $\lim_{t\to\infty} N(t)/t = \omega$ w.p. 1 and let $\lim_{t\to\infty} \sum_{i=1}^{N(t)} X_i/t = \omega\beta$ w.p. 1. For most practical situations, the $S(t)$ process is assumed to have a downward drift in steady state, (otherwise, the infinite horizon ruin probability will be 1), i.e., $c > \omega\beta$ or $\rho := \omega\beta/c < 1$.

For even the simplest of these problems, no closed form solution exists for either $\psi(u)$ or $\psi(u, T)$. Asymptotic (i.e., as $u \to \infty$) approximations abound in the literature, but there are very few bounds. Hence simulation presents a viable alternative. However note that as $u \to \infty$, $\psi(u) \to 0$ and $\psi(u, T) \to 0$. Hence when u is large, the probability of ruin is small, and then naive simulation is not efficient, as then many sample paths need to be generated in order to get a sample path where the ruin actually happens. The case of $\psi(u)$ suffers from the additional problem of not knowing when to stop the simulation.

Importance sampling is a widely used simulation technique to speed up the occurrence of rare events (see, e.g., Heidelberger (1995), Asmussen (2000)). The basic idea is to change the probability dynamics of the system so that the event of interest happens faster. One then uses likelihood ratios to unbias the estimator. However, any arbitrary probability dynamics that speeds up the occurrence of rare events may give estimators that have higher variances than those using naive simulation. Hence there exists a vast amount of literature on coming up with efficient changes of measure for rare event simulation problems in different settings (e.g., Heidelberger (1995), Lehtohnen and Nyrhinen (1992)).

It is well-known that for the new probability dynamics to be effective, it should closely approximate the probability dynamics of the original system *conditioned* on the rare event happening. In the case when the claim distribution is *light-tailed*, large deviations theory is used to determine the 'most likely path' under some scaling along which the rare event happens (see, e.g., Cottrell, Fort and Malgouyres (1983), Bucklew (1990), and Sadowsky (1991)). Then *exponential twisting* (see, e.g., Siegmund (1976), and Asmussen (1985)) is used to direct the process along the most likely path. Such an approach is known to give large reductions in variance over naive simulation for large classes of systems, including i.i.d. interarrival times of claims, and Markov-modulated arrival times of claims (see, e.g., Lehtohnen and Nyrhinen (1992)). In particular, many of the importance sampling techniques based on this approach are known to be *asymptotically optimal*, i.e., the asymptotic (exponential) rate of decay of the second moment is twice that of the first moment, which is the slowest possible rate (since the second moment is always greater than or equal to the square of the first moment). Asymptotic optimality is the standard criterion used in the literature to classify a rare event simulation technique as being efficient (see, e.g., Heidelberger (1995)).

However, when the claim size is subexponential, then one cannot use traditional large deviations theory as then the most likely path is much more complex, and moment generating functions that are a key component of traditional large deviations theory, do not exist. Hence simple extensions of approaches used in the light-tailed case cannot be used here.

To the best of our knowledge, the first fast simulation approach in the subexponential setting was presented by Asmussen and Binswanger (1997). They gave an approach for simulating for the probability that the geometric sum of i.i.d. subexponential random variables exceeds a given high value. One property of subexponentially distributed random variables is that given that the sum of i.i.d. subexponentially distributed random variables is 'large', the sum behaves very much like the maximum of that group. Hence computing the probability of the maximum exceeding u, after simulating for the other lower order statistics, eliminates much of the variance. This insight forms the basis of their technique and they showed that it was asymptotically optimal for subexponential distributions with regularly varying tails (e.g. Pareto).

Juneja and Shahabuddin (1999) introduced the notion of *subexponential twisting* or more specifically *hazard rate twisting*, were one twists at a sub-exponential rate. They proved that a *delayed* version of hazard rate twisting was asymptotically optimal for most subexponential distributions for the problem described in the preceding paragraph. Asmussen, Binswanger, and Hojgaard (2000) present an alternative importance sampling approach that is also asymptotically optimal for most subexponential distributions.

However all the afore-mentioned techniques are for estimating the probability that the geometric sum of i.i.d. random variables exceeds a given level. This has applications in simulating for the ruin probability in insurance risk processes with Poisson claim arrivals and i.i.d. claim sizes (or by duality, for the probability that the steady-state waiting time exceeds a given value in a M/GI/1 queue). This is accomplished by using the Pollaczek-Khintchine transformation to transform the actual problem into the problem of geometric sums mentioned above. However, even for one of the simplest cases, where one has i.i.d. renewal interarrival times instead of Poisson arrivals, even though such a representation exists, it is not useful (as then the parameters of the geometric distribution, and the distribution of the i.i.d. random variables that appear in the geometric sum, are not easily calculable). Hence the approach breaks down.

Boots and Shahabuddin (2000b; see Boots and Shahabuddin 2000a for a preliminary version) considered the problem of estimating $\psi(u)$ for the case with i.i.d. interarrival times and i.i.d. claim sizes, where the sequence of claim sizes is independent of the sequence of interarrival times. Instead of using the Pollaczek-Khintchine transformation, they simulated the claim-surplus process directly. In particular they simulated the random walk associated with the claim-surplus process, i.e., $S_i = S_{i-1} + (X_i - c\xi_i)$, which is the claim-surplus process viewed at the moments of claim arrivals (or more precisely, an infinitesimal time after these moments). Then they used the fact that $\psi(u)$ can also be expressed as $\mathbb{P}(\sup_{n\in\mathbb{N}} S_n > u)$. The change of measure that they used consists of subexponentially twisting the claim sizes X_i. They showed that for claim-size distributions that have 'Weibull type tails', the new estimator is *large set asymptotically optimal*. The "large set" means that instead of estimating the probability of the rare event, one estimates the probability of a large subset of the rare event in an asymptotically optimal manner, i.e., the subset on which the variance is well behaved. To satisfy this criterion, the large set needs to be such that asymptotic relative bias is bounded by a prespecified value of one's choice. Hence in practical terms, it is very close to asymptotic optimality. The proof for claim-size distributions that have 'Pareto type tails' or 'lognormal type tails' is still an open problem.

The main contribution of this paper is to generalize the approach of Boots and Shahabuddin (2000b) to:

- more general arrival processes and claim-size processes like those described at the beginning of this section.
- estimating $\psi(u, T)$.

The approach we take in each case is to state very general sufficient conditions on the claim-surplus process and the claim-size distributions, for the change of measure that we propose to be large set asymptotically optimal. Then we give specific examples where these conditions are satisfied. In particular we show that the case where the arrival process and the claim-size process are Markov-modulated satisfy these conditions. Some 'perturbed' versions of commonly used insurance risk processes also satisfy these conditions; see Boots and Shahabuddin (2001). All the proofs for the theoretical results in this paper are also given in Boots and Shahabuddin (2001).

2 MODELS AND NOTATION

We start with some commonly used notation. For a distribution F with tail $\overline{F} \equiv 1 - F$ and density f, let $\lambda(x) \equiv f(x)/\overline{F}(x)$ be the hazard-rate function, and $\Lambda(x) = \int_{s=0}^{x} \lambda(s)\mathrm{d}s$ be the hazard function (also called the cumulant function). The tail of any distribution, $\overline{F}(x)$, may be written as $\overline{F}(x) = \exp(-\Lambda(x))$. The integrated tail of F is defined by $F_I(x) := \int_0^x \overline{F}(y)\mathrm{d}y / \int_0^\infty \overline{F}(y)\mathrm{d}y$ when $\int_0^\infty \overline{F}(y)\mathrm{d}y < \infty$. Let $\lambda_I(x)$ be the hazard-rate function and $\Lambda_I(x)$ be the hazard function corresponding to F_I. For any functions $z_1(x)$ and $z_2(x)$, we use the notation $z_1(x) \sim z_2(x)$ as $x \to \infty$, to mean that the ratio of $z_1(x)$ to $z_2(x)$ converges to 1 as x goes to infinity. We define $F^{\leftarrow}(y) = \inf\{x : F(x) = y\}$. If the inverse function of F is well-defined, then $F^{\leftarrow} \equiv F^{-1}$. Finally, the minimum of x_1 and x_2 is denoted by $x_1 \wedge x_2$ and the indicator random variable corresponding to the event A is denoted by $I(A)$.

To model the possibility of large claims we allow claims to be subexponentially distributed. The definition of subexponentiality is due to Chistyakov (1964) (see Embrechts, Klüppelbeg and Mikosch (1997) for details about subexponential distributions):

Definition 2.1 *Let (X_n) be a sequence of i.i.d. non-negative random variables with distribution F. The F is subexponential (denoted by $F \in \mathcal{S}$) iff for all $n \geq 2$*

$$\frac{\mathbb{P}(X_1 + \cdots + X_n > u)}{n\mathbb{P}(X_1 > u)} \to 1 \ (u \to \infty). \tag{2}$$

We call a distribution light-tailed if it decays at an exponential or faster rate. Subexponential distributions are *not* light-tailed, but decay at a sub-exponential rate.

Section 2.1 presents the insurance risk model in its most general form, as well as some examples.

2.1 The Insurance Risk Model

Consider the insurance risk model described in Section 1. We assume that F_i has a finite mean β_i. Define $\mathcal{M} := \{F_1, \ldots, F_m\}$. We assume that we can partition $\mathcal{M}$ into $\mathcal{M}_1$ and $\mathcal{M}_2$, such that there exists a subexponential distribution $F \equiv F_{i^\star} \in \mathcal{M}$ that satisfies $\overline{F}_j(x) \sim b_j \overline{F}(x)$, with $b_j = 0$ for $j \in \mathcal{M}_1$, $b_j > 0$ for $j \in \mathcal{M}_2$. Without loss of generality, we assume that $\max_{i \in \{1,\ldots,m\}} b_i = 1$, i.e., the claim-size distribution $F \equiv F_{i^\star}$ has the heaviest tail. Let Λ_i denote the hazard function corresponding to F_i and let Λ denote the hazard function corresponding to F. We assume that for x large enough, $\Lambda(x) \leq \Lambda_i(x)$ for all $F_i \in \mathcal{M}$. The following assumption is satisfied by most subexponential distributions (see Juneja and Shahabuddin (1999) for a discussion).

Assumption 1 *The hazard-rate function $\lambda(x)$ is eventually decreasing.*

The time to ruin $\sigma(u)$ is defined by $\inf\{t : S(t) > u)\}$. Define $\tau(u) := N(\sigma(u))$. If ruin occurs, $\tau(u)$ can be interpreted as the number of claims that arrive before the ruin occurs. For $T < \infty$, define $\tau(u, T) := N(\sigma(u) \wedge T)$. One can interpret $\tau(u, T)$ as the number of claims that arrive either before ruin occurs or the horizon of length T is exceeded. Note that $\psi(u, T) = \mathbb{P}(\sigma(u) < T)$ and that $\psi(u) = \mathbb{P}(\sigma(u) < \infty) = \mathbb{P}(\tau(u) < \infty)$. Two specific examples of insurance risk processes that fall in the above framework are the renewal arrivals model and the Markov-modulated compound Poisson model.

2.1.1 Markov-Modulated Compound Poisson Model

In this model the claim-interarrival times follow a Poisson process, and both the arrival intensity and the claim-size distribution are governed by some continuous time Markov chain $\{J(t), t \geq 0\}$ with a finite state space $\{1, \ldots, m\}$ and an unique stationary distribution $\{\pi_1, \ldots, \pi_m\}$, $\pi_i > 0$ for $i = 1, \ldots, m$. When $J(t)$ is in state i, the arrival intensity equals ω_i, the premium rate equals $c \equiv 1$ and the claim-size distribution equals F_i, i.e., $F_i = \mathbb{P}(X_j \le x \mid J(\eta_j) = i)$. Let $(J_n)_{n\in\mathbb{N}}$ be the discrete time Markov chain which is embedded in the process $\{J(t), t \ge 0\}$ at claim arrival epochs, i.e. $J_n = J(\eta_n)$. Define $\mathcal{M} := \{F_1, \ldots, F_m\}$. In this model $\omega = \sum_{i=1}^m \pi_i \omega_i$, $\beta = \sum_{i=1}^m \omega_i \pi_i \beta_i / \sum_{i=1}^m \omega_i \pi_i$, and thus $\rho = \sum_{i=1}^m \pi_i \omega_i \beta_i$. The probability of ruin before time T with initial environment i is denoted by $\psi_i(u, T)$. Define $\mathbb{P}_i(\cdot)$ by $\mathbb{P}(\cdot \mid J(0) = i)$. In most cases we omit the subscript $_i$ and we use the concise notation $\mathbb{P}$ instead of $\mathbb{P}_i$, $\psi(u)$ instead of $\psi_i(u)$ and $\psi(u, T)$ instead of $\psi_i(u, T)$.

3 RARE EVENTS SIMULATION

As mentioned in Section 1, for u large, $\psi(u, T)$, $0 < T \le \infty$, is small, and naive simulation, i.e., estimating $\psi(u, T)$ by the sample mean of many independent replications of $I(\sup_{0\le t<T} S(t) > u)$ (note that $\mathbb{E}[I(\sup_{0\le t<T} S(t) > u)] = \mathbb{P}(\sup_{0\le t<T} S(t) > u)$), is too slow to give a reliable estimate of $\psi(u, T)$ in a reasonable amount of time. We now briefly describe the concept of importance sampling, that is used to speed up the simulation. Suppose the stochastic process that we wish to simulate is defined on some probability space with measure $\mathbb{P}$. Let $A(u)$ be some event with the property that $\alpha(u) := \mathbb{P}(A(u)) \to 0$ as $u \to \infty$. Let $\mathbb{Q}$ be some other measure on the same probability space such that $\mathbb{P}$ is absolutely continuous relative to $\mathbb{Q}$. One can then express $\alpha(u) = \mathbb{E}_{\mathbb{Q}}[I(A(u))\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}]$, where $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$ is called the likelihood-ratio and subscript $\mathbb{Q}$ indicates that the expectation is with respect to the new measure $\mathbb{Q}$. In importance sampling one generates the sample paths under the $\mathbb{Q}$ measure, computes the likelihood-ratio in each case, and estimates $\alpha(u)$ by the sample mean of the $I(A(u))(\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q})$'s, which is an unbiased estimator of $\alpha(u)$. The problem is to determine the new measure $\mathbb{Q}$.

As mentioned in Section 1 in the light-tailed setting, sample path large deviations arguments are used to derive changes of measure that are asympotically optimal. However, such large deviations arguments do not seem to work in the subexponential case (see Asmussen (2000), Pg. 287, and Asmussen, Binswanger and Hojgaard (2000) for counter examples). In Boots and Shahabuddin (2000b) it is shown that in the renewal model with subexponentially distributed claim sizes, if one tolerates a small bias, then in some cases it is possible to come up with an efficient importance sampling change of measure. Following Boots and Shahabuddin (2000b), we use the following criterion to classify estimators as efficient (think of δ as the maximum asymptotic relative bias that one is willing to tolerate in the simulation).

Definition 3.1 Work-normalized large set asymptotically optimal.
Let $\delta \in (0, 1)$ be a fixed constant. If there exists a family of decompositions (parameterized by β) of $\alpha(u)$ into two positive quantities $\alpha(u) = \gamma_\beta(u) + \epsilon_\beta(u)$ s.t.

1. *for any given β, $\limsup_{u\to\infty} \epsilon_\beta(u)/\alpha(u) \le \delta$,*
2. *for any fixed u, $\limsup_{\beta\to\infty} \epsilon_\beta(u)/\alpha(u) = 0$,*
3. *for any given β, there exists an unbiased estimator $\widehat{\gamma_\beta}(u)$ of $\gamma_\beta(u)$ that is work-normalized asymptotically optimal, i.e.,*

$$\liminf_{u\to\infty} \frac{\log\left(\mathrm{work}(u) \times \mathbb{V}\mathrm{ar}\left[\widehat{\gamma_\beta}(u)\right]\right)}{\log(\gamma_\beta^2(u))} \ge 1, \qquad (3)$$

then $\widehat{\gamma_\beta}(u)$ is said to be a work-normalized large set asymptotically optimal (a.o.) estimator of $\alpha(u)$.

The parameter β in the above definition is used to make the *relative bias* of the estimator small enough (say smaller than δ) for any fixed value of u (in contrast to the asymptotic relative bias which we are assured is less than δ). For conciseness, we write $\gamma_\beta(u) \equiv \gamma(u)$ and $\epsilon_\beta(u) \equiv \epsilon(u)$, and we use $\beta \equiv 1$, except in the section with the experimental results.

The new measure $\mathbb{Q}$ we use consists of applying hazard rate twisting (HRT) on each of the claim-size distributions; no other distributions in the insurance risk process are changed. Hazard rate twisting (Juneja and Shahabuddin (1999)) is implemented by replacing the distribution F_j by the new distribution $F_{j,\theta_u}(x) := 1 - e^{-\Lambda_j(x)(1-\theta_u)}$, where $0 \le \theta_u < 1$ is an appropriately selected function of u. The density corresponding to F_{j,θ_u} is given by $f_{j,\theta_u}(x) := (1-\theta_u)\lambda_j(x)e^{-(1-\theta_u)\Lambda_j(x)}$. Weighted delayed hazard rate twisting (WDHRT) extends HRT by introducing a weighting parameter w_u and a delaying parameter $x_u^\star$. Both w_u and $x_u^\star$ are possibly functions of u. The WDHRT density is defined by

$$f_{j,\theta_u,x_u^\star}(x) := \begin{cases} \dfrac{f_j(x)}{1+w_u}, & \text{for } x \le x_u^\star, \\ \left(1 - \dfrac{F_j(x_u^\star)}{1+w_u}\right) \dfrac{f_{j,\theta_u}(x)}{\overline{F}_{j,\theta_u}(x_u^\star)}, & \text{for } x > x_u^\star. \end{cases} \qquad (4)$$

Denote the corresponding cdf by $F_{j,\theta_u,x_u^\star}(x)$.

The parameter θ_u is chosen such that under $\mathbb{Q}$ the occurrence of a large claim is far more likely than under $\mathbb{P}$. The parameter w_u is used to control the frequency of the occurrence of large claims under $\mathbb{Q}$. The parameter $x_u^\star$ is used to guarantee that the distribution of a claim size, conditioned on the claim being small, looks similar under both $\mathbb{P}$ and $\mathbb{Q}$. All these try to capture the basic principle that the importance sampling measure should be very close to the original measure conditioned on the rare event happening. The exact selections of these parameters will be given later.

4 INFINITE HORIZON RUIN PROBABILITIES

As mentioned before, it is very difficult to find an a.o. estimator for $\psi(u) = \mathbb{E}[I(\tau(u) < \infty)]$. Let us say we use hazard rate twisting represented by $\mathbb{Q}_u \equiv \mathbb{Q}$ and in that case $\mathbb{E}[I(\tau(u) < \infty)] = \mathbb{E}_{\mathbb{Q}}[I(\tau(u) < \infty)L_{\mathbb{Q}}(u)]$, where now $L \equiv L_{\mathbb{Q}}(u)$ is a function of u. The main problem is that the $L_{\mathbb{Q}}(u)$ becomes highly variable on the set of sample paths where $\tau(u)$ is 'large' but finite. So why not isolate the part where $L_{\mathbb{Q}}(u)$ is highly variable, and just estimate the part where it is not? Do we lose much by leaving out the part on which $L_{\mathbb{Q}}(u)$ is highly variable, i.e., how much bias do we incur? In particular, let us say that $L_{\mathbb{Q}}(u)$ is 'well-behaved' when $\tau(u) \le k_0$ for some $k_0 \equiv k_0(u)$ that may be a function of u, and not well-behaved otherwise. Then if $\mathbb{E}[I(\tau(u) < \infty)] \approx \mathbb{E}[I(\tau(u) \le k_0(u)] = \mathbb{E}_{\mathbb{Q}}[I(\tau(u) \le k_0(u))L_{\mathbb{Q}}(u)]$ and $L_{\mathbb{Q}}(u)$ has a low variance over the set $\{\tau(u) \le k_0\}$ (in particular, $\mathbb{V}\mathrm{ar}[I(\tau(u) \le k_0(u))L_{\mathbb{Q}}(u)]$ satisfies the conditions for $\mathbb{E}[I(\tau(u) \le k_0(u)]$ to be estimated a.o.), then we are not losing much.

One question that arises here is: what kind of bias is one willing to tolerate? A reasonable compromise is stated in the first criterion of the large set a.o. criteria. If we set $\alpha(u) = \mathbb{E}[I(\tau(u) < \infty)]$ and $\gamma(u) = \mathbb{E}[I(\tau(u) \le k_0(u)]$, then the criterion states that the asymptotic relative bias be less than some given value δ, or equivalently, $\mathbb{E}[I(\tau(u) \le k_0(u)]/\mathbb{E}[I(\tau(u) < \infty)]$ be greater than $1-\delta$. In unbiased rare event simulations, the best one hopes for is bounded relative error (relative error is defined as confidence interval half-width divided by $\alpha(u)$) as $u \to \infty$. So it makes sense that in biased simulations one should be able to tolerate a bounded asymptotic relative bias.

How does one find such a $k_0(u)$? Fortunately, in certain cases, there exist large deviations results for the distribution of $\tau(u)$ under some scaling (e.g., Asmussen and Klüppelberg (1996), and Asmussen and Hojgaard (1996)) and one can use them to determine such a $k_0(u)$. Instead of proving large set a.o. on a case by case basis, we make the existence of such large deviations results as one of a set of sufficient conditions for our algorithm to produce large set a.o. estimators. We then give particular cases for which these conditions are satisfied. This is done in the next subsection.

4.1 Large Deviations Conditions

Condition 4.1 (Large deviations condition on $\psi(u)$.)
(i)The probability $\psi(u) = \mathbb{P}(\tau(u) < \infty)$ satisfies

$$\lim_{u\to\infty} \frac{-\log(\psi(u))}{\Lambda_I(u)} = 1.$$

Also, for any given $\delta > 0$, there exists a function $k_0(u)$ such that:
(ii)

$$\lim_{u\to\infty} \frac{\mathbb{P}(\tau(u) \le k_0(u))}{\mathbb{P}(\tau(u) < \infty)} \ge 1-\delta$$

for each arbitrary small $\delta > 0$,
(iii) $\log(k_0(u)) = o(\Lambda_I(u))$,
(iv) and there exists a constant $b > 1$ satisfying

$$\lim_{u\to\infty} \frac{k_0(u)}{\Lambda(u)^{b-1}} = 0$$

and

$$\lim_{u\to\infty} \frac{k_0(u)\left(F^{\leftarrow}\left(1-\Lambda(u)^{-b}\right)\right)}{u} = 0\ .$$

In Condition 4.1 (ii), $k_0(u)$ is some function of u, such that the asymptotic relative bias remains bounded. However we do not want such a $k_0(u)$ to grow very fast, as then both the effort and the variance in the estimation of $\mathbb{P}(\tau(u) < k_0(u))$ may grow very fast, thus making the estimation inefficient. Condition 4.1 (iii) and (iv) present restrictions on the growth of $k_0(u)$. In particular, (iii) is a restriction imposed to keep the effort per replication in control, and (iv) is a restriction imposed to keep the variance of the importance sampling estimator (under weighted delayed hazard rate twisting) of $\mathbb{P}(\tau(u) < k_0(u))$ under control.

Finally we need an assumption on the distribution F:

Assumption 2 $\quad \lim_{u\to\infty} \dfrac{\Lambda(u)}{\Lambda_I(u)} = 1$

This assumption is satisfied by distributions like the Weibull and lognormal, but not the Pareto.

We will now show that a broad class of insurance risk models satisfy Condition 4.1. If F is subexponential, then in both the renewal model and the Markov-modulated compound Poisson model,

$$\psi(u) \sim \frac{\rho}{1-\rho}\overline{F}_I(u) \quad (u \to \infty), \tag{5}$$

i.e., Condition 4.1 (i) is satisfied. For the proof, we refer to Embrechts and Veraverbeke (1982) for the renewal model (see also Pakes (1975)) and to Asmussen, Henriksen and Klüppelberg (1994), Asmussen (2000), and Jelenković and Lazar (1998) for the Markov-modulated model. In Asmussen, Schmidli and Schmidt (1999) some conditions are given under which (5) holds, as well as several examples. For Condition 4.1 (ii) consider models where F satisfies the following assumption:

Assumption 3 *$F_I \in \mathcal{S}$ and F is in the maximum domain of attraction of the Gumbel distribution (denoted by $F \in$MDA(Gumbel)).*

Two well-known distributions in this class are the Weibull and the lognormal. Define $\mathbb{P}_i^{(u)}(\cdot)$ by $\mathbb{P}(\cdot \mid \tau(u) < \infty;\ J(0) = i)$. If Assumption 3 holds, then in both the renewal model (see Asmussen and Klüppelberg (1996)) and the Markov-modulated compound Poisson model (see Asmussen and Hojgaard (1996)),

$$\frac{\tau(u)}{a(u)} \to \frac{\mathrm{Exp}(1)}{1-\rho} \tag{6}$$

in $\mathbb{P}_i^{(u)}$ distribution. Here Exp(1) denotes an exponential random variable with mean 1 and $a(u)$ is defined to be any function such that $a(u) \sim \int_u^\infty \overline{F}(x)\mathrm{d}x/F(u)$ (for details about $a(u)$, which is called the auxiliary function in extreme value theory, see, e.g., Asmussen and Hojgaard (1996), Asmussen and Klüppelberg (1996), Boots and Shahabuddin (2000b), Embrechts, Klüppelberg and Mikosch (1997), and Goldie and Resnick (1988)). With (6) and with $k_0(u) := -a(u)(\log\delta)/\mu$ we find for $i = 1, \ldots, m$,

$$\begin{aligned} & \frac{\mathbb{P}_i(k_0(u) < \tau(u) < \infty)}{\mathbb{P}_i(\tau(u) < \infty)} \\ = \; & \mathbb{P}_i(\tau(u) > k_0(u) \mid \tau(u) < \infty) \\ = \; & \mathbb{P}_i^{(u)}\left(\frac{\tau(u)}{a(u)} > \frac{k_0(u)}{a(u)}\right) \\ = \; & \mathbb{P}_i^{(u)}\left(\frac{\tau(u)}{a(u)} > \frac{-\log\delta}{\mu}\right) \to \delta \;(u \to \infty). \end{aligned}$$

This implies

$$\lim_{u\to\infty} \frac{\mathbb{P}(\tau(u) \le k_0(u))}{\mathbb{P}(\tau(u) < \infty)} = 1 - \delta. \tag{7}$$

Thus in these cases Condition 4.1 (ii) is satisfied. Also note that with $k_0(u)$ of this form, Condition 4.1 (iii) and (iv) become conditions on the $a(u)$ (or equivalently, on F). An example of a distribution that satisfies these three conditions is the Weibull with shape parameter less than 1. An example of a distribution that satisfies Condition 4.1 (iii) but not (iv) is the lognormal.

Another example of a distribution that does not satisfy Condition 4.1 is the Pareto. Even in this case, a $k_0(u)$ may be determined that satisfies 4.1 (ii) but this $k_0(u)$ does not satisfy Conditions 4.1 (iii) and (iv) simultaneously (see Boots and Shahabuddin (2000b)).

4.2 The Simulation Algorithm and its Efficiency

As mentioned before, we use weighted delayed hazard rate twisting on each of the claim-size distributions. For the parameters, we use

$$\theta_u = 1 - \frac{1}{\Lambda(u)}, \;\; w_u = \frac{c_1 \log\delta}{k_0(u)}$$
$$\text{and} \quad x_u^\star = F^{\leftarrow}\left(1 - \Lambda(u)^{-b}\right),$$

where $c_1 > 0$ is some constant and b is the constant in Condition 4.1 (iv). Note that $\Lambda(x_u^\star) = b\log\Lambda(u)$ and that $x_u^\star$ goes to infinity as u goes to infinity, because $\Lambda(u) \to \infty$. The reasons for these selections become clearer when one goes over the proofs of the main results.

Algorithm 1 (WDHRT for the estimation of $\psi(u)$.) *Instead of using $F_i = 1 - e^{-\Lambda_i(x)}$ to generate a particular claim size conditioned on its type being i, we use a distribution that is obtained by applying WDHRT on F_i, with θ_u, w_u and $x_u^\star$ chosen as above. Let Z be the resulting likelihood-ratio. We simulate a replication of*

$$S = \sum_{i=1}^{\min\{\tau(u),k_0(u)\}} (X_i - c\xi_i).$$

Set $V = Z$ if $S > u$ and set $V = 0$ otherwise. An average of k i.i.d. replications of V is taken as an unbiased estimator of $\mathbb{P}(\tau(u) \le k_0(u))$. Use this estimator of $\mathbb{P}(\tau(u) \le k_0(u))$ as an estimate of $\psi(u)$.

Theorem 4.2 *Under Condition 4.1, Assumption 1 and Assumption 2, Algorithm 1 yields a work-normalized large set a.o. estimator for $\psi(u)$.*

5 FINITE HORIZON RUIN PROBABILITIES

5.1 Large Deviations Conditions

Finding a change of measure for efficiently estimating $\psi(u)$ is more difficult than finding one for estimating $\psi(u, T)$, $T < \infty$, since in the latter case the distribution of the number of claims arriving in a sample path of the insurance risk process is independent of u. The $N(T)$ also has a light tail. Let a_0 be the infimum over all a, such that $\mathbb{P}(N(T) \ge k) \le a^k$ for all sufficiently large k.

In that case the following large deviations condition and Assumption 1, is sufficient to ensure a.o.

Condition 5.1 (Large deviations condition on $\psi(u, T)$, $T < \infty$.) *The $\psi(u, T)$ satisfies*

$$\lim_{u\to\infty} \frac{-\log(\psi(u, T))}{\Lambda(u)} = 1.$$

The following theorems show that a broad class of insurance risk processes with subexponential claims satisfy Condition 5.1.

Theorem 5.2 *Under Assumption 1, in both the renewal model and the Markov-modulated compound Poisson model, $\psi(u, T)$ satisfies Condition 5.1.*

Theorem 5.2 also holds in insurance risk models other than the ones considered in this paper, e.g., some special cases of the Markov-modulated renewal model in Jelenković and Lazar (1998).

5.2 Simulation Algorithm and its Efficiency

We use the same θ_u and x_u^* as before. But for the w_u, we use any w such that $w < a_0^{-1} - 1$. If a_0 is not known, then just use $w = 0$.

Algorithm 2 (WDHRT for the estimation of $\psi(u, T)$.) *Consider the insurance risk model. Instead of using distribution $F_i = 1 - e^{-\Lambda_i(x)}$ to draw from the* *distribution of X_n conditioned on its type being i, use the distribution that is obtained by applying hazard rate twisting on F_i with θ_u, x_u^* and w_u chosen as above. Let Z be the resulting likelihood-ratio. Generate a replication of*

$$S = \sum_{i=1}^{\tau(u,T)} X_i - c(\sigma(u) \wedge T).$$

Set $V = Z$ if $S > u$ and set $V = 0$ otherwise. An average of k i.i.d. replications of V is taken as an unbiased estimate of $\psi(u, T)$

Theorem 5.3 *Given Assumption 1, Algorithm 2 with the given θ_u, $x_u^\star$ and w_u produces an estimator that is a.o.*

5.3 Extensions to the Pollaczek-Khintchine representation of the compound Poisson model

As mentioned before, in the renewal model with Poisson arrivals, the $\psi(u)$ can be expressed as $\mathbb{P}(Y_1 + \cdots + Y_M > u)$ with the Y_i's i.i.d. with distribution F_I, and M independent of the Y_i's and having the (geometric) distribution $\mathbb{P}(M = k) = (1-\rho)\rho^k$, for $k \geq 0$. Analogous versions of Theorem 5.2, Algorithm 2, and Theorem 5.3 hold for this 'finite horizon representation' of $\psi(u)$ (replace F by F_I, Λ by Λ_I, $X_i - c\xi_i$ by Y_i and $\tau(u, T)$ by M, and take $m = 1$). In this way we improve over the algorithms and results in Asmussen and Binswanger (1997), and Juneja and Shahabuddin (1999) by having a technique that is a.o. with less assumptions on the claim-size distribution.

6 EXPERIMENTAL RESULTS

In this section we use the algorithms given in the previous sections (we refer to them generically as WDHRT) to estimate the finite and infinite horizon ruin probabilities in various models. We then experimentally evaluate the efficiency of the simulation algorithms.

The standard effort of a simulation algorithm using independent replications is defined to be the expected CPU time per replication times the variance per replication. We compute the *efficiency ratio* of WDHRT (the efficiency ratio of a simulation algorithm is defined as the standard effort of the new algorithm divided by the standard effort of naive simulation; see, e.g., Glynn and Whitt (1992)) to compare the performance of these algorithms with that of naive simulation. The variance per replication of naive simulation is estimated by $\widehat{\psi}(u, T)(1 - \widehat{\psi}(u, T))$, where $\widehat{\psi}(u, T), 0 < T \leq \infty$ denotes the estimator of $\psi(u, T), 0 < T \leq \infty$ obtained by WDHRT. We consider the examples described earlier, i.e., the renewal arrival model, the Markov-modulated compound Poisson model. In the more complex models (i.e., the Markov-modulated models) we also give the estimate of $\psi(u, T), 0 < T \leq \infty$ obtained by naive simulation, if the estimate is non-zero. We compare the WDHRT estimate of $\psi(u, T), 0 \leq T \leq \infty$, with asymptotic approximations obtained by other authors.

In the tables with experimental results we use some conventions. The relative error is estimated by the ratio of the estimated confidence interval half-width (HW) and the simulation estimate of the ruin probability. The numbers immediately after the estimates $\widehat{\psi}(u, T)$ of $\psi(u, T)$ obtained by naive simulation and WDHRT denote the estimates of the relative error. The number between brackets after the WDHRT estimate of the relative error denotes the efficiency ratio of WDHRT. The number in the brackets after the asymptotic approximation (AA) denotes the relative bias of AA, i.e. $|AA - \psi(u, T)|/\psi(u, T)$, where we approximate $\psi(u, T), 0 < T \leq \infty$ by the reasonably accurate WDHRT estimate.

In the estimation of $\psi(u)$, we use $k_0(u) = \max\{-a(u) \log \delta^\beta / \mu, 50\}$ where the "50" is added for practical considerations. For the w_u when F is Weibull, we use the optimal c_1 obtained via the heuristic for the infinite horizon ruin probability described in Section 5.3 in Boots and Shahabuddin (2000b). We call this choice $\tilde{w}_u$. For lognormal F we use $w = 0$.

6.1 Finite Horizon Ruin Probabilities in the Compound Poisson Model

In Table 1 we present estimates of $\psi(u, T)$ in the compound Poisson (non-Markov-modulated) model with Weibull claim sizes. We use $\beta = 2$, $\rho = 0.5$, $b = 2.1$ and $n = 300,000$ (n denotes the number of simulation replications).

For the Weibull distributed claim sizes, we use both $w = 0$ and $w \equiv \tilde{w}_u$. The results indicate that the quality of the WDHRT estimator heavily depends on the choice of w and that $\tilde{w}_u$ is preferable to $w = 0$.

Results for $\psi(u)$ for such models have been presented Asmussen and Binswanger (1997), Juneja and Shahabuddin (1999), and Boots and Shahabuddin (2000b). The first two papers use methods for simulating the geometric sum of subexponentially distributed random variables. The algorithm in Boots and Shahabuddin (2000b) is the same as WDHRT in this paper applied to the special case of the renewal model.

6.2 Markov-Modulated Compound Poisson Model

Tables 2 and 3 present estimates of $\psi(u, T)$ in a two state Markov-modulated compound Poisson model. We use $n = 1,000,000$, $b = 2.1$ and $w = \tilde{w}_u$ in Table 2, and $w = 0$ in Table 3. In both Tables 2 and 3, the F_2 is the Weibull(1,.75) distribution for which $\beta_2 = 1.1906$. Also, $\pi_1 = .5$, $\pi_2 = .5$, $\omega_1 = .2$, $\omega_2 = .3$ and $\rho = .38$. In Table 2, the F_1 is the Weibull(1,.5) distribution that has

Table 1: Estimates of $\psi(u, 100)$ in the Compound Poisson Model with Weibull(1, .5) Claim Sizes

u		
200	WDHRT, $w = 0$	$1.30E-5 \pm 10.9\%(29.8)$
	WDHRT, $w \equiv \tilde{w}_u$	$1.49E-5 \pm 3.9\%(2.6E2)$
	AA	$9.06E-6(39.3\%)$
800	WDHRT, $w = 0$	$1.04E-11 \pm 17.5\%(1.4E7)$
	WDHRT, $w \equiv \tilde{w}_u$	$1.01E-11 \pm 4.7\%(2.1E8)$
	AA	$8.94E-12(11.8\%)$

$\beta_1 = 2$. In Table 3, the F_1 is Pareto(1,.5) and $\beta_1 = 2$. The initial state of $J(t)$ is one with probability π_1 and two with probability π_2.

Table 2: Estimates of $\psi(u, \infty)$ in the Markov-Modulated Compound Poisson Model with Weibull(1,.5) and Weibull(1,.75) Claim Sizes

u		
100	WDHRT	$2.55E-4 \pm 3.4\%(41.1)$
	Naive Simulation	$2.83E-4 \pm 27.9\%$
	AA	$1.52E-4(40.3\%)$
400	WDHRT	$1.69E-8 \pm 3.8\%(4.5E5)$
	Naive Simulation	–
	AA	$1.32E-8(21.9\%)$

Table 3: Table 3: Estimates of $\psi(u, 50)$ in the Markov-Modulated Compound Poisson Model with Pareto(1,.5) and Weibull(1,.75) claim sizes

u		
200	WDHRT	$3.32E-5 \pm 6.8\%(15.2)$
	Naive Simulation	$3.39E-5 \pm 14.0\%$
	AA	$5.43E-5(63.6\%)$
400	WDHRT	$8.53E-6 \pm 6.5\%(62.7)$
	Naive Simulation	$1.03E-5 \pm 25.4$
	AA	$1.45E-5(70.0\%)$

In the Markov-modulated compound Poisson model, WDHRT gives in general worse results than in similar renewal models. WDHRT still improves considerably over naive simulation in the experiments we conducted.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation (U.S.A.) Career Award Grant DMI 96-25297 and an IBM University Partnership Program Award. It was partially performed while the first author was visiting Columbia University and while the second author was a guest of the Tinbergen Institute at Vrije University.

REFERENCES

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

Asmussen, S. 2000. *Ruin Probabilities*. World Scientific, Singapore, New Jersey, London, Hong Kong.

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

Asmussen, S., K. Binswanger and B. Hojgaard. 2000. Rare event simulation for heavy-tailed distributions. *Bernoulli* 6 (2): 303-322.

Asmussen, S., L.F. Henriksen, and C. Klüppelberg. 1994. Large claims approximations for risk processes in a Markovian environment. *Stochastic Processes and their Applications* 54: 29-43.

Asmussen, S., and B. Hojgaard. 1996. Finite horizon ruin probabilities for Markov-modulated risk processes with heavy tails. *Th. Random Processes*, 2: 96-107.

Asmussen, S., and C. Klüppelberg. 1996. Large deviations results for subexponential tails, with applications to insurance risk. *Stochastic Processes and their Applications* 64: 103-125.

Asmussen, S., H. Schmidli, and V. Schmidt. 1999. Tail probabilities for non-standard risk and queueing processes with subexponential jumps. *Advances in Applied Probability* 31: 422-447.

Boots, N.K., and P. Shahabuddin. 2000a. Simulating GI/GI/1 queues and insurance risk processes with subexponential distributions. *Proceedings of the 2000 Winter Simulation Conference* 656-665, IEEE Press, Piscataway, New Jersey.

Boots, N.K., and P. Shahabuddin. 2000b. Simulating tail probabilities in GI/GI/1 queues and insurance risk processes with subexponential distributions. Research Report, Dept. of Industrial Engineering and Operations Research, Columbia University, NY 10027.

Boots, N.K., and P. Shahabuddin. 2001. A framework for simulating small ruin probabilities in insurance risk processes with subexponential distributions. Research Report, Dept. of Industrial Engineering and Operations Research, Columbia University, NY 10027.

Bucklew, J.A. 1990. *Large Deviations Techniques in Decision, Simulation, and Estimation*. John Wiley & Sons, Inc.

Cottrell, M., J.C. Fort and G. Malgouyres. 1983. Large deviations and rare events in the study of stochastic algorithms. *IEEE Transactions on Automatic Control* AC28: 907-920.

Chistyakov, V.P. 1964. A theorem on sums of independent positive random variables and its applications to branching processes. *Theory of Probability and its Applications* 9: 640-648.

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

Embrechts, P., and C. Klüppelberg. 1993. Some aspects of insurance mathematics. *Theory of Probability and its Applications* 38 (2): 262-295.

Embrechts, P., and N. Veraverbeke. 1982. Estimates for the probability of ruin with the special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics* 1: 55-72.

Glynn, P.W., and Whitt, W. 1992. The asymptotic efficiency of simulation estimators. *Operations Research* 40: 505-520.

Goldie, C.M., and S. Resnick. 1988. Distributions that are both subexponential and in the domain of attraction of an extreme-value distribution. *Advances in Applied Probability* 20: 706-718.

Jelenković, P., and A.A. Lazar. 1998. Subexponential asymptotics for a Markov-modulated random walk with a queueing application. *Journal of Applied Probability* 35: 325-347.

Juneja, S., and P. Shahabuddin. 1999. Simulating heavy-tailed processes using delayed hazard rate twisting. Research Report, Dept. of Industrial Engineering and Operations Research, Columbia University, NY 10027.

Heidelberger, P. 1995. Fast simulation of rare events in queueing and reliability models. *ACM Transactions on Modeling and Computer Simulation* 6: 43-85.

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

Pakes, A.G. 1975. On the tails of waiting time distributions. *Journal of Applied Probability* 12: 555-564.

Sadowsky, J.S. 1991. Large deviations and efficient estimation of excessive backlogs in GI/G/m queue. *IEEE Transactions on Automatic Control* 36 (1991): 1383-1394.

Shahabuddin, P. 1994. Importance sampling for the simulation of highly reliable Markovian systems. *Management Science* 40: 333-352.

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

AUTHOR BIOGRAPHIES

NAM KYOO BOOTS is a Ph.D. student at the Department of Economics and Business Administration at Vrije University in the Netherlands since 1997. He received a M.S. in Econometrics (1997) and a M.S. in Mathematics (1998), both from Vrije University. His research interests include rare event simulation in stochastic models with heavy-tailed random variables. His e-mail address is `<nboots@feweb.vu.nl>`.

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 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). He is a recipient of a 1996 National Science Foundation (U.S.A) Career Award, a 1998 IBM University Partnership Award and the 1996 Outstanding Simulation Publication Award given by INFORMS College on Simulation. He serves on the editorial boards of *Management Science, Stochastic Models, IEEE Transactions on Reliability, IIE Transactions on Operations Engineering,* and *ACM Transactions on Modeling and Computer Simulation* (Guest Editor). His e-mail address is `<perwez@ieor.columbia.edu>`.