Proceedings of the 2011 Winter Simulation Conference
S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, eds.

IMPORTANCE SAMPLING FOR ACTUARIAL COST ANALYSIS UNDER A HEAVY TRAFFIC MODEL

Jose Blanchet

Columbia University
500 W 120th Street
New York, NY 10027, USA

Henry Lam

Boston University
111 Cummington Street
Boston, MA 02215, USA

ABSTRACT

We explore a bottom-up approach to revisit the problem of cash flow modeling in insurance business, and propose a methodology to efficiently simulate the related tail quantities, namely the fixed-time and the finite-horizon ruin probabilities. Our model builds upon the micro-level contract structure issued by the insurer, and aims to capture the bankruptcy risk exhibited by the aggregation of policyholders. This distinguishes from traditional risk theory that uses random-walk-type model, and also enhances risk evaluation in actuarial pricing practice by incorporating the dynamic arrivals of policyholders in emerging cost analysis. The simulation methodology relies on our model's connection to infinite-server queues with non-homogeneous cost under heavy traffic. We will construct a sequential importance sampler with provable efficiency, along with large deviations asymptotics.

1 INTRODUCTION

This paper explores a bottom-up approach to revisit the problem of cash flow modeling in insurance business, and proposes a methodology to simulate the related tail quantities efficiently. We set up our model from both the perspectives of risk theory and actuarial pricing. In brief, given the types and structures of insurance contracts issued, we formulate and compute ruin problems that evaluate the risk carried by the insurer's business.

Classical risk theory uses a random walk model to describe the probability of bankruptcy for an insurance company. The most basic Cramer-Lundberg model assumes a constant rate of premium earned by the insurance company i.e. the aggregate premium grows linearly with time. The claims, on the other hand, follow a compound Poisson process, the arrival being the occurrence of claims and the summands being the claim sizes. The aggregate net asset value of the insurance company is then the difference of claims and premiums. Restricting to uniformly discretized observations, this is equivalent to a negatively drifted random walk. The ruin probability of the insurance company is modeled as the probability that the negatively drifted compound Poisson process, or random walk in discrete-time version, hits a high level that represents the company's initial surplus. This model has been extended and modified in many different directions, such as Markov economic environment, generalization to Levy processes, combination with an independent

investment process, incorporation of operational costs, among others (see, for example, Asmussen 2000). Due to the celebrated Cramer-Lundberg asymptotic and large deviations theory, analytical approximations of the ruin probability is available in many interesting examples, as well as the corresponding rare-event simulation algorithms.

There is a related but nonetheless different way of looking at cash flow's tail probability in an insurance company, in the context of what is called emerging cost analysis. This is used by product teams in life insurance companies to estimate the risk of particular insurance products. In this context, the payoff of the product is well defined, and stochastic simulation is run to obtain the value-at-risk of the product. Usually the important case is when the product involves financial risk, and when the product has sophisticated payoff structure. The standard practice is to assume a single account entering the contract of the product initially, then run a scenario to obtain a path sample. This is repeated thousands of times to calculate the quantile. Stochastic simulation has become standard practice for risk management and capital calibration in big insurance companies (see, for example, Hardy 2003).

Despite its simplicity, this practice may sometimes miscalculate the actual risk involved, because it ignores the dynamic arrival of policyholders as the company signs new insurance contracts over time. The goal of this paper is to explore this aspect of risk in the practice of emerging cost analysis. In particular, we look at the dynamic aggregation of cash flow built up from these individual policyholder accounts over time. This aggregate cash flow can also be viewed as the cash flow process from a risk theoretic viewpoint, but instead of being a negatively drifted random walk, it is now a functional of all signed policies and statuses of policyholders. Although it may sound infeasibly complicated, we will show that this new process is readily analyzable. Indeed, we will derive a large deviations asymptotic and construct efficient algorithm to calculate its tail probability. This includes fixed-time and finite-horizon level-crossing probabilities.

From the risk theoretic perspective, besides the bottom-up approach using micro-level policy accounts, our work also distinguishes from traditional ruin problems by focusing on the risk effect of aggregation. Our model aims to capture the characteristics of the large number of policyholders that large insurance companies usually have in shaping their finite-horizon risk levels. This is in contrast to classical risk theory which focuses on the eventual ruin probability i.e. risk involved over long time horizon, that is primarily triggered by a deviation from the drift. The rationale behind our approach is that in product-level risk analysis, the horizon in consideration by risk management personnel is in the same scale as the duration of products, and their main concern is the risk exhibited by the payoff of the particular products in mass sales. Nevertheless, it can be shown that a moderate twist in our model formulation by elongating the time horizon and fixing the scaling would uncover a connection with classical ruin theory (Blanchet and Lam 2011a).

A salient feature of our new model is its connection to infinite-server queue. The rich literature in the study of such objects allows us to obtain analytical and computational results with little effort. This connection can be briefly seen as follows. The entrance of policyholders into insurance contracts can be represented by customer arrivals in a queue. Once a policyholder signs into a contract, he has to pay the premium accordingly, and in return he receives benefit claims during accident or death. After the accident happens, the insurance contract typically ends automatically. This is analogous to the end of service in a queue. Since there is no "server", the "insurance system" is an infinite-server queue with a time-changing cost function imposed on the service. Moreover, a typical insurance company has hundreds of thousands of policyholders, so it is plausible to invoke approximation in heavy-traffic theory.

Another driver of our bottom-up approach is the ease in incorporating various extensions and practical considerations, as we will discuss in Section 6. We hope that the model could ultimately be useful in a real-life framework. Despite such ambition, we acknowledge that our model is still far from perfection, since it certainly oversimplifies many important operational and investment issues. We will briefly discuss these future directions also in Section 6.

The organization of the paper is as follows. In Section 2 we describe our model and state the assumptions. In Section 3 we derive a large deviations result for a fixed-time tail probability of our model. Then, in

Section 4, we will construct an efficient algorithm for calculating the fixed-time tail probability, while in Section 5 we will generalize to finite-horizon framework and also show a finite-horizon asymptotic result. Section 6 is devoted to several extensions of the algorithm. Finally, we close the paper with some numerical results in Section 7.

2 MODEL ASSUMPTIONS

2.1 Insurance Contract

In this section we introduce our model and notations. First let us describe the insurance contract structure. For ease of computation and also due to practical data collection considerations, we use a discrete time formulation i.e. we focus on time $t = 0, 1, \ldots$. The time increment represented is typically a week or month. When a policyholder enters a contract with the insurance company, he pays a premium $p(t)$ at each time $t = 0, 1, \ldots$. This lasts until the casualty happens, when instead he is paid a benefit of $b(T)$, where T is the time of casualty.

This framework can be used for, say, life and disability accidents. These insurances share the feature that claim is a one-time event i.e. after a claim is made, the contract ceases automatically. In health and property insurance, the contract usually lasts until the stated period ends, and many claims can occur during the covered period. While our framework can potentially be applied in such contexts, we shall not discuss this here and will leave it to future research. For convenience, from now on we will focus on life insurance (with notations easily translatable to the case of disability and similar insurance types).

We assume a constant compound interest rate δ, and let $d = e^{-\delta}$ be the discount factor. We assume policyholders all follow the same mortality distribution at the time of arrival, and their mortalities are independent. Let T be the random time between arrival and death, and $f(t)$, $F(t)$ and $\bar{F}(t)$ be its probability mass, distribution and survival functions respectively. We assume that T has support on $1, \ldots, M$ i.e. $f(t) > 0$ if and only if $t = 1, \ldots, M$. For simplicity we assume no operational cost. All these assumptions will be relaxed in Section 6 when we discuss the extensions.

We now introduce a few actuarial notations. We let $a(t) = \sum_{i=0}^{t-1} p(i)d^i$ be the annuity immediate. This quantity captures the accumulated premium paid from time 0 to t, discounted at time 0. Also, we let $A(t) = b(t)d^t$ be the discounted benefit. For example, in the case of whole life insurance with constant premium rate, we have $p(t) = p$ and $b(t) = b$ for constants p and b. Then $a(t) = p(1 - d^t)/(1 - d)$ and $b(t) = bd^t$.

2.2 Arrival Process

We now discuss the arrival process of policyholders, a feature that distinguishes our work from the prevailing actuarial literature. This also connects our work to infinite-server queueing system and allows us to draw upon established results in the area.

To begin, we assume the number of new arrivals at each time t to be i.i.d. following the distribution generated by the random variable N^s. The number s is a scale parameter that we assume to be large (to denote the large number of arrivals). We assume N^s has exponential moment i.e. $\psi_N^s(\theta) := \log Ee^{\theta N^s} < \infty$ for all $\theta \in \mathbb{R}$. Moreover, we assume the scaling s is such that

$$\lim_{s\to\infty} \frac{1}{s}\psi_N^s(\theta) = \psi_N(\theta) \tag{1}$$

We also assume that $\psi_N(\cdot)$ is steep (i.e. its derivative $\psi_N'(\cdot)$ has range $(0, \infty)$), differentiable and strictly convex. These conditions hold, for example, for Poisson random variable with rate λs. In this case, $\psi_N(\theta) = \lambda(e^\theta - 1)$.

Each of these new arrivals then follow i.i.d. mortality distribution given by the random variable T described in the previous subsection. To facilitate our discussion, we let N_k be the number of policyholders who arrive at time k (i.e. N_k is each distributed as i.i.d. N^s), and $N_k(i)$ be the number of policyholders

who arrive at time k and die i time units after k. Note that we have suppressed the superscripts s in N_k and $N_k(i)$ for notational convenience. It is then easy to see that given N_k, $N_k(i), i = 1, \ldots, M$ follows a multinomial distribution with probability $f(i)$.

2.3 Regularity Conditions

For convenience we will assume there are no policyholders before time 0, and the first batch of policyholders is N_0. On the other hand, we assume the insurance company holds an initial surplus of xs for some constant $x > 0$. This surplus grows at an accumulation rate δ (so the time-0 discounted value of surplus at any future time is always xs). The scaling assumption means that the insurance company holds a surplus in proportion to the expected basis of its business.

Regarding the insurance contract, we impose the following regularity conditions:

Assumption 1 (feasibility) There exists i in $\{1, \ldots, M\}$ such that $A(i) - a(i) > 0$.

This implies that there is a positive chance the contract will induce loss to the insurance company. For many popular insurance contracts, $A(t) - a(t)$ is non-increasing in t and $A(1) - a(1) = b(1)d - p > 0$, which implies Assumption 1. Examples of monotonic-payoff contracts include whole life insurance, term life insurance, endowment, etc. The principle is that longevity is desirable to the insurance company. Nevertheless, if Assumption 1 is violated, then no rational individual has an incentive to enter the contract since this would cause loss to him with probability one. We will see in the next section that Assumption 1 guarantees a steepness condition for our Gartner-Ellis limit.

Assumption 2 (profitability) It holds that $Ea(T) > EA(T)$.

This assumption states that the actuarial present value of premium is larger than that of the benefit paid for an individual contract. This ensures in the long run the insurance business is profitable. On the contrary, if this does not hold, the company is doomed to go bankrupt eventually since an average contract leads to loss. Note that the standard Equivalence Principle in actuarial literature (Bowers et al. 1997) states that $Ea(T) = EA(T)$ i.e. the net profit to the insurance company is zero. While this holds in a perfectly competitive market, insurance companies often charge a premium loading, among other extra income to make lucrative business in practice. These loadings are built into the premium $p(t)$ in our formulation.

In the case of whole life insurance with constant premium rate, for example, Assumption 2 states that $p(1 - Ed^T)/(1 - d) > bEd^T$.

3 LARGE DEVIATIONS FOR FIXED-TIME PROBABILITY

We first look at the net asset of the insurance company at a fixed time t. For a policyholder who arrives at time $k < t$ and has death time T, his contribution to the net asset is described as follows. For convenience we focus on the accumulated cash outflow, discounted at time 0, from the insurance company up to time t. This is the negative of net asset. For succinctness we merely say "cash outflow" to refer to this quantity. Also, we let $t \geq \min\{i : A(i) - a(i) > 0\}$ where the minimum exists by Assumption 1. Any t smaller than this minimum will not be interesting since the probability of positive $C(t)$ is 0.

1. If $k + T > t$, then the policyholder is still alive at time t. In this case, benefit paid is 0 and accumulated premium up to time t is $a(t-k)d^k$. So the cash outflow is $-a(t-k)d^k$.
2. If $k+T \leq t$, then the policyholder deceases before or at t. The discounted benefit paid is $A(T)d^k$, whereas the accumulated premium up to $k+T$ is $a(T)d^k$. So the cash outflow is $(A(T)-a(T))d^k$.

Now denote $C(t)$ as the cash outflow at time t. Recall that we assume no policyholders initially. For notational convenience, denote

$$h_u(i) = \begin{cases} A(i) - a(i) & \text{for } i \leq u \\ -a(u) & \text{for } i > u \end{cases}$$

Then

$$C(t) = \sum_{k=0}^{t} \left[\sum_{i=1}^{M \wedge (t-k)} N_k(i)(A(i)-a(i))d^k - \sum_{i=t-k+1}^{M} N_k(i)a(t-k)d^k I(M > t-k) \right]$$
$$= \sum_{k=0}^{t} \sum_{i=1}^{M} N_k(i) h_{t-k}(i) d^k \tag{2}$$

We want to study the probability $P(C(t) > xs)$ i.e. the probability that the cash outflow exceeds the (cumulated) initial surplus. To do so, we find the large deviations rate using Gartner-Ellis theorem (Dembo and Zeitouni 1998). First, the logarithmic moment generating function of $C(t)$ is

$$\psi^s_{C,t}(\theta) := \log E e^{\theta C(t)} = \log E \left[E \left[\exp \left\{ \theta \sum_{k=0}^{t} \sum_{i=1}^{M} N_k(i) h_{t-k}(i) d^k \right\} \middle| N_k, k = 0, \ldots, t \right] \right]$$
$$= \sum_{k=0}^{t} \psi^s_N \left(\log \left(\sum_{i=0}^{M} f(i) e^{\theta h_{t-k}(i) d^k} \right) \right)$$

Hence

$$\psi_{C,t}(\theta) := \lim_{s \to \infty} \frac{1}{s} \psi^s_{C,t}(\theta) = \sum_{k=0}^{t} \psi_N \left(\log \left(\sum_{i=0}^{M} f(i) e^{\theta h_{t-k}(i) d^k} \right) \right) = \sum_{k=0}^{t} \psi_N (\log g(\theta; t, k)) \tag{3}$$

by (1), where for convenience we define

$$g(\theta; t, k) = \sum_{i=0}^{M} f(i) e^{\theta h_{t-k}(i) d^k} \tag{4}$$

Consider the equation

$$\psi'_{C,t}(\theta) = x \tag{5}$$

Note that

$$\psi'_{C,t}(\theta) = \sum_{k=0}^{t} \psi'_N (\log g(\theta; t, k)) \times \frac{\sum_{i=1}^{M} f(i) h_{t-k}(i) d^k e^{\theta h_{t-k}(i) d^k}}{g(\theta; t, k)} = \sum_{k=0}^{t} \psi'_N (\log g(\theta; t, k)) \times$$
$$\frac{\sum_{i=1}^{M \wedge (t-k)} f(i)(A(i)-a(i)) d^k e^{\theta (A(i)-a(i)) d^k} - \sum_{i=t-k+1}^{M} f(i) a(i) d^k e^{-\theta a(i) d^k}}{g(\theta; t, k)} \tag{6}$$

Since $f(i) > 0$ for all $i = 1, \ldots, M$ and $A(i) - a(i) > 0$ for some $i = 1, \ldots, M$ by Assumption 1, together with the steepness assumption on $\psi_N(\cdot)$, we see that $\psi_{C,t}(\cdot)$ also has domain $\mathbb{R}$ and is steep. Moreover, by Assumption 2 and that $-a(w) \leq A(w) - a(w)$ for any $w \geq 0$, we have $\psi_{C,t}(0) < 0$. The strict convexity of $\psi_{C,t}(\cdot)$ inherits from $\psi_N(\cdot)$. Therefore there is a unique positive solution to (5). Call it θ_t.

By Gartner-Ellis theorem (Dembo and Zeitouni 1998), we have

$$\lim_{s \to \infty} \frac{1}{s} \log P(C(t) > xs) = -I_t \tag{7}$$

where $I_t = \theta_t x - \psi_{C,t}(\theta_t)$.

Equation (7) is the starting point of our analysis. This in particular allows us to follow the scheme in Szechtman and Glynn (2002), Blanchet et al. (2009) and Blanchet and Lam (2011) to tackle the finite-horizon ruin problem.

4 IMPORTANCE SAMPLING FOR FIXED-TIME PROBABILITY

We aim to construct an efficient importance sampling algorithm for $P(C(t) > xs)$, using the criterion of so-called asymptotic optimality in the rare-event literature. Suppose we want to simulate efficiently a rare event probability $P(A_s)$ where $P(A_s) \searrow 0$ as $s \nearrow \infty$ using an importance sampler with likelihood ratio L. We say that the importance sampler is asymptotically optimal, or logarithmically efficient, if

$$\lim_{s\to\infty} \frac{\log \tilde{E} I(A_s) L^2}{\log P(A_s)} \geq 2 \tag{8}$$

This criterion is standard and ensures that the relative error of the unbiased estimator (ratio of standard deviations to probability of interest) does not grow exponentially as the event parametrized by s gets rarer. We will use criterion (8) throughout the paper for evaluating algorithmic efficiency. For more discussion on rare-event simulation via importance sampling, see Bucklew (2004) and Asmussen and Glynn (2007).

We first note that, as in the literature in rare-event simulation, a natural change of measure is the optimal exponential tilting formed by the parameter in (5). This algorithm in theory would induce an exponential reduction in variance, but is unfortunately not implementable because the distribution of $C(t)$ is not directly analyzable.

Instead, we follow the methodology in Szechtman and Glynn (2002), Blanchet et al. (2009) and Blanchet and Lam (2011). The idea is to replace direct exponential tilting with sequential tilting such that the tilting at each step becomes computable. The sequential tilting procedure is hinted by comparing (6) to the same expression with $\theta = 0$, which suggests that asymptotically the mean of the θ-tilted number of arrivals at time k has mean $s\psi_N'(\log g(\theta; t, k))$ and the mean of the corresponding death distribution has mean $\sum_{i=1}^M f(i) h_{t-k}(i) d^k e^{\theta h_{t-k}(i) d^k} / g(\theta; t, k)$. The following tilting scheme possesses these properties. Define the new measure $\tilde{P}_t$ such that for each $k = 0, \ldots, t$,

$$\tilde{P}_t(N_k = n) = e^{-s\psi_N(\log g(\theta_t; t, k))} g(\theta_t; t, k)^n P(N_k = n) \tag{9}$$

and for each arrival at time k,

$$\tilde{P}_t(T = i) = \frac{f(i) e^{\theta_t h_{t-k}(i) d^k}}{g(\theta_t; t, k)} \tag{10}$$

Note that for $k = 0, \ldots, t$, we have

$$\frac{P(N_k = n)}{\tilde{P}_t(N_k = n)} = e^{s\psi_N(\log g(\theta_t;t,k)) - n \log g(\theta_t;t,k)} \quad \text{and} \quad \frac{P(T = i)}{\tilde{P}_t(T = i)} = e^{\log g(\theta_t;t,k) - \theta_t h_{t-k}(i) d^k}$$

Hence under the new measure, with $N_k = n$ and $N_k(i) = n_i$, the sequential likelihood ratio at time k is

$$\begin{aligned} L_{t,k}(n_i, i = 1, \ldots, M) &= \frac{P(N_k = n)}{\tilde{P}_t(N_k = n)} \cdot \frac{\binom{n}{n_1,\ldots,n_M} \prod_{i=1}^M P(T = i)^{n_i}}{\binom{n}{n_1,\ldots,n_M} \prod_{i=1}^M \tilde{P}_t(T = i)^{n_i}} \\ &= \frac{P(N_k = n)}{\tilde{P}_t(N_k = n)} \cdot \prod_{i=1}^M \left(\frac{P(T = i)}{\tilde{P}_t(T = i)} \right)^{n_i} = e^{s\psi_N(\log g(\theta_t;t,k)) - \sum_{i=1}^M n_i \theta_t h_{t-k}(i) d^k} \end{aligned} \tag{11}$$

In particular, at time t, the likelihood ratio is

$$L_{t,t}(n) := e^{s\psi_N(\theta_t p(0) d^t) - \theta_t n} \tag{12}$$

From (11), the overall likelihood ratio is

$$\begin{aligned} L_t = \prod_{k=1}^t L_{t,k}(N_k(i), i = 1, \ldots, M) &= \exp\left\{ s \sum_{k=1}^t \psi_N(\log g(\theta_t; t, k)) - \theta_t \sum_{k=1}^t \sum_{i=1}^M N_k(i) h_{t-k}(i) d^k \right\} \\ &= e^{s\psi_{C,t}(\theta_t) - \theta_t C(t)} \end{aligned} \tag{13}$$

by (2) and (3).

We state our algorithm precisely. The following procedure generates one unbiased sample:

Algorithm 1

Step 1 Initialize $L = 1$, $C = 0$.

Step 2 For $k = 1, \ldots, t-1$, do:

1. Generate N_k according to (9).
2. Generate $N_k(i), i = 1, \ldots, M$ according to a multinomial distribution with parameter being the sampled N_k and the probability vector given by (10).
3. Update $L \leftarrow L \cdot L_{t,k}(N_k(i), i = 1, \ldots, M)$, where $L_{t,k}(N_k(i), i = 1, \ldots, M)$ is given by (11).
4. Update $C \leftarrow C + \sum_{i=1}^{M} N_k(i) h_{t-k}(i) d^k$.

Step 3 For $k = t$, do:

1. Generate N_t according to (9).
2. Update $L \leftarrow L \cdot L_{t,t}(N_t)$, where $L_{t,t}(N_t)$ is defined in (12).
3. Update $C \leftarrow C + N_t p(0) d^t$.

Step 4 Output $I(C > xs)L$.

This procedure leads to an asymptotically optimal estimator for $P(C(t) > xs)$, by the following simple argument. From (13), we have the second moment of our estimator

$$\tilde{E}[I(C(t) > xs)L_t^2] = e^{-2sI_t} \tilde{E}[e^{2\theta_t(xs - C(t))}; C(t) > xs] \leq e^{-2sI_t}$$

This shows that (8) holds and Algorithm 1 is asymptotically optimal.

5 IMPORTANCE SAMPLING AND LARGE DEVIATIONS FOR FINITE-HORIZON PROBLEM

Our goal in this section is to construct a rare-event simulation algorithm for the finite-horizon ruin probability, namely $P(\max_{t=1,\ldots,T} C(t) > xs)$. We follow the time-randomization method in Blanchet et al. (2009) and Blanchet and Lam (2011), which also gives the finite-horizon large deviations result simultaneously. The idea is to sample a random time, say $R \in \{1, \ldots, T\}$, independent of the system, followed by Algorithm 1 putting t equal to the sampled R. The distribution of R is flexible in a finite-horizon problem. For example, one can merely choose uniform distribution over $[0, T]$. The sequential tilting can stop whenever $C(t)$ hits xs, possibly a time before the sampled R, since anything that happens after that does not affect the occurrence of the rare event. Nevertheless, this stopping is not essential. We call this first passage time τ.

An important feature of this algorithm is that one should view the sample $\{N_t(\cdot), t = 1, \ldots, \tau \wedge T\}$ in a pathwise and vector-valued sense i.e. we define the probability space on the sample path of $N_t(\cdot)$ and the element in the likelihood ratio is a path of vectors. Let $\tilde{P}$ be the probability measure with the random time R, under which the sample path of $N_t(\cdot)$ is generated. The likelihood ratio is then a mixture given by

$$L^{-1} := \frac{d\tilde{P}}{dP} = \sum_{t=1}^{T} P(R = t) L_t'^{-1} \tag{14}$$

where $L_t' = \prod_{k=1}^{t \wedge \tau} L_{t,k}$ and $L_{t,k}$ is defined in (11). Note that the elements in the individual likelihood ratios L_t' depend on the configuration of $N_t(\cdot)$ at different times t, in contrast to L_t defined in (13) that only acts on a single value of $C(t)$.

Note that we can write out L'_t as follows:

$$L'_t = \begin{cases} e^{s\psi_{C,t}(\theta_t)-\theta_t C(t)} & \text{for } t \le \tau \\ \exp\{s\sum_{k=0}^{\tau}\psi_N(\log g(\theta_t;t,k)) - \theta_t C[\tau,t]\} & \text{for } t > \tau \end{cases} \tag{15}$$

where $C[u,t]$ is defined as the discounted cash outflow at time t contributed by the arrivals before or at u, with $u \le t$ i.e.

$$C[u,t] = \sum_{k=0}^{u}\sum_{i=1}^{M} N_k(i)h_{t-k}(i)d^k$$

To implement the scheme, we first compute $N_k(i)$ for every $k = 1,\ldots,T$ and $i = 1,\ldots,M$ under the new measure $\tilde{P}_r$ after sampling a realization $R = r$. Then we can compute $C(t)$ for $t = 1,\ldots,T$ to locate $\tau \wedge T$ and also to compute the likelihood ratio. Let us present our algorithm as follows:

Algorithm 2

Step 1 Sample R from uniform distribution on $\{1,\ldots,T\}$. Call the realization $R = r$.

Step 2 Sample $N_k(i)$ for $k = 1,\ldots,r$, $i = 1,\ldots,M$ under $\tilde{P}_r$. If $r < T$, then sample $N_k(i)$ for $k = r+1,\ldots,T$, $i = 1,\ldots,M$ under the original measure P.

Step 3 Compute $C(t)$ for $t = 1,\ldots,T$ using (2).

Step 4 Find τ in $\{1,\ldots,T\}$ from the computed $\{C(t)\}$. If $C(t) \le xs$ for all $t \in \{1,\ldots,T\}$, then set $\tau = \infty$.

Step 5 If $\tau \le T$, output the likelihood ratio L from (14) and (15); otherwise output 0.

Note that this algorithm needs order TM space to implement. In Blanchet and Lam (2011a) we will provide an alternate algorithm that can reduce the memory space to order $2M + T$, using a Markov representation to capture the dynamic evolution of $C(t)$.

Our main result is the following:

Theorem 1 Algorithm 2 is asymptotically optimal.

Proof. Define $I^* = \min\{I_t : t = 1,\ldots,T\}$. We show that the second moment of the estimator from Algorithm 2 and 3 has an exponential decay rate faster than $2I^*$. Note that

$$L = \frac{1}{\sum_{u=1}^{T} P(R=u)L_u^{-1}} \le P(R=\tau)^{-1}L_\tau = T\exp\left\{\sum_{k=1}^{\tau} s\psi_N(\log g(\theta_\tau;\tau,k)) - \theta_\tau C(\tau)\right\}$$

Then

$$\begin{aligned} \tilde{E}[L^2;\tau\le T] &\le T\tilde{E}\left[\exp\left\{\sum_{k=1}^{\tau} s\psi_N(\log g(\theta_\tau;\tau,k)) - \theta_\tau xs + \theta_\tau(xs - C(\tau))\right\};\tau\le T\right] \\ &\le Te^{-2sI^*}\tilde{E}[e^{\theta_\tau(xs-C(\tau))};\tau\le T] \le Te^{-2sI^*} \end{aligned} \tag{16}$$

which shows our claim.

On the other hand, we can see that I^* is also the slowest decay rate that the probability $P(\tau \le T) = P(\max_{t=1,\ldots,T} C(t) > xs)$ can achieve, by noting that $P(\max_{t=1,\ldots,T} C(t) > xs) \ge P(C(t) > xs)$ for any $t = 1,\ldots,T$. From (7) we then get $\liminf_{s\to\infty}(1/s)\log P(\max_{t=1,\ldots,T} C(t) > xs) \ge -I_t$ for any $t = 1,\ldots,T$. In particular, we take $t = \operatorname{argmin} I_t$ and get $\liminf_{s\to\infty}(1/s)\log P(\max_{t=1,\ldots,T} C(t) > xs) \ge -I^*$. Coupled with (16) and Jensen's inequality, we get

$$-2I^* \le \liminf_{s\to\infty}\frac{1}{s}\log P\left(\max_{t=1,\ldots,T} C(t) > xs\right)^2 \le \limsup_{s\to\infty}\frac{1}{s}\log\tilde{E}[L^2;\tau\le T] \le -2I^*$$

This shows that Algorithm 2 is asymptotically optimal, and the probability $P(\max_{t=1,\ldots,T} C(t) > xs)$ satisfies a large deviations asymptotic $\lim_{s\to\infty}(1/s)\log P(\max_{t=1,\ldots,T} C(t) > xs) = -I^*$. □

6 EXTENSIONS

The model that we introduce so far is very basic and ignores many practical complications. In this section we discuss how we can improve the model in several directions.

6.1 Multiclass Policyholders

Suppose there are more than one type of policyholders. These types can be classified by age group, sex, social-economic background, health condition, etc. Suppose that the arrivals of these types follow independent random variables $N_{t,j}$, where t denotes the time and j denotes the type. Also suppose that the types elicit different mortality random variables T_j. We let j be in the range $1,\ldots,J$. These assumptions hold naturally for a Poisson arrival model with thinning probabilities.

Hence we can treat the cash outflow contributed by each type, called $C_j(t)$, as independent cash outflow process. We are then interested in the probability $P\left(\max_{t=1,\ldots,T}\sum_{j=1}^{J} C_j(t) > xs\right)$. Note that the logarithmic moment generating function of $\sum_{j=1}^{J} C_j(t)$ is then $\psi_{C,t}(\theta) := \sum_{j=1}^{J}\psi_{C,t,j}(\theta)$, where $\psi_{C,t,j}(\cdot)$ is the logarithmic moment generating function of $C_j(t)$. We can solve for θ_t using this new $\psi_{C,t}(\cdot)$ to get the rate function.

The algorithm for fixed-time and finite-horizon problem remains largely the same, but instead of tilting one type of policyholders, we have to tilt the measure for all types of policyholders using the parameter θ_t simultaneously over time. The time randomization method still holds for the finite-horizon problem.

We note that the same idea can be applied to a multi-product situation.

6.2 Time-Varying Model

Suppose that the interest rate, arrivals and death distribution all vary with time in a known fashion. We introduce the notations d_t, $\psi_{N,t}(\cdot)$ and $f_t(\cdot)$ to denote the interest rate, the logarithmic moment generating function of arrivals and the probability mass function of death time for an arrival at time t. Also define $D_{l,k} = \prod_{j=l}^{k} d_j$.

The logarithmic moment generating function of $C(t)$ is then

$$\psi_{C,t}(\theta) = \sum_{k=0}^{t}\psi_{N,k}\left(\log\left(\sum_{i=1}^{M} f_k(i)h_{t,k}(i)D_{1,k}\right)\right)$$

where

$$h_{t,k}(i) = \begin{cases} A_{t,k}(i) - a_{t,k}(i) \text{ for } i \le t-k \\ -a_{t,k}(t-k) \text{ for } i > t-k \end{cases}$$

and $A_{t,k}(i) = b(i)D_{k+1,i}$ and $a_{t,k}(i) = \sum_{j=0}^{i-1} p(j)D_{k+1,j}$.

The rest is the same as before. In particular, the large deviations asymptotic, fixed-time and finite-horizon importance samplers remain. We can even use different policies at different times i.e. premium $p_k(i)$ and $b_k(i)$ depends on the arrival time k.

6.3 Markov Economic Environment

Suppose now that there is an underlying finite state space Markov chain $Y_t \in \mathcal{S}$ that controls the parameters in the system i.e. interest rate, arrival and death time distribution at time t. This case can be reduced easily to the time-varying scenario. More precisely, suppose we want to find the finite-horizon ruin probability.

In running the simulation, we first sample the Markov states from time 0 to T. This realizes the values of the parameters in the system. Now this time-varying system can be simulated using the same procedure as above. It can be shown by similar argument as Theorem 1 that asymptotic optimality and large deviations asymptotic still remains, now with the exponential decay rate $I^* := \min_{t=1,\dots,T, y\in\mathcal{S}^T} I_{t,y}$, where $I_{t,y}$ is the rate function of the fixed-time probability given $Y := \{Y_t\}_{t=1,\dots,T} = y \in \mathcal{S}^T$. We further note that the Markov assumption is not crucial, as other finite dimensional processes would work as well.

6.4 Operational Cost

There are a few ways to incorporate operational cost. If cost is incurred every time a contract is signed, then we merely modify the first premium to be $p(0) - c$ where c is the cost. If cost is incurred when an accident occurs, then we modify benefit to $b(i) + c$.

Suppose that cost is incurred over time (e.g. for infrastructure, etc.). We can then modify the initial surplus to change in time i.e. we have a decreasing $x(t)$ as our first passage level. In this case, we can solve $\psi_{C,t}(\theta) = x(t)$ to get θ_t. The rest remains similar as before.

6.5 Policy Withdrawal

When policyholders are allowed to elapse, the contract terminal distribution will be adjusted from only death to include withdrawal, and the payoff structure at the decrement time for the two cases would be different. We can calculate the new decrement distribution, say $\tilde{f}(i)$, and given that decrement occurs at i, we calculate the probabilities, say $p_d(i)$ and $p_e(i)$, that the policyholder deceases or elapses. The cash outflow $C(t)$ then becomes

$$\begin{aligned} C(t) &= \sum_{k=0}^{t} \Bigg[\sum_{i=1}^{M\wedge(t-k)} \sum_{j=1}^{N_k(i)} [(A(i)-a(i))d^k I_d(j) + (A_e(i)-a(i))d^k I_e(j)] \\ &\quad - \sum_{i=t-k+1}^{M} N_k(i) a(t-k) d^k I(M > t-k) \Bigg] \end{aligned}$$

where $I_d(j)$ and $I_e(j) = 1 - I_d(j)$ are indicator variables to denote whether person j decrements due to death or elapse, and $A_e(i)$ denotes the withdrawal benefit at period i. The logarithmic moment generating function of this new $C(t)$ can be obtained similarly as before and the same analysis follows.

6.6 Actuarial Reserve

Actuarial reserve is the amount of back-up capital, required by statutory law and for risk control, to be set aside by the insurance company to account for future benefit payment. This quantity is often calculated on a contract basis, and the amount set aside for a particular contract is the expected future cost incurred by the contract. Note that if a contract ends i.e. a policyholder deceases, nothing has to be backed up and so the actuarial reserve is zero. However, if the policyholder still survives at time t since his arrival, then the actuarial reserve is

$$V_t = \sum_{i=t+1}^{M} \frac{f(i)}{\bar{F}(t)} \left[b(i)d^i - \sum_{j=t+1}^{i-1} p(j) d^j I(i > t+1) \right]$$

Note that V_t can be positive or negative because of Assumption (2). Now, with the reserve requirement, the net cash outflow of the insurance company becomes

$$C(t) = \sum_{k=0}^{t} \sum_{i=1}^{M} N_k(i) [h_{t-k}(i) d^k + V_{t,k} d^k]$$

where

$$V_{t,k} = \begin{cases} V_{t-k} & \text{if } k+i > t \\ 0 & \text{if } k+i \leq t \end{cases}$$

We can then analyze the asymptotic and algorithmic efficiency as before.

6.7 Other Discussions

There are other important issues that are not addressed in our current model. One of them is the financial risk involved when capital is invested. The interesting case is when the interest rate δ fluctuates in a scale close to or larger than the fluctuation due to arrivals and deaths. Note that our Markov-modulated formulation does not address this, since the effect of the Markov chain does not scale with s. Along the same line, investment-linked insurance products, such as guaranteed minimum death benefit and accumulation benefit schemes, are not tackled in our model.

Other issues include the applicability to insurance with multiple claims, a situation that arises in property and casualty insurance as mentioned previously. There are other types of life products, such as pension and retirement schemes, that are also subject to further investigation. Finally, an interesting future direction is the case of correlated policyholders, for example in common shock decrements.

7 NUMERICAL EXAMPLE

We close this paper by a numerical example. Assume Poisson arrivals with rate λs and $\lambda = 1$. Also assume a uniform mortality distribution over $[1, M]$ with $M = 10$. Set $\delta = 0.01$ and so $d = e^{-\delta} = 0.99$. We use whole life insurance with $b(i) = 1$ for all $i = 1, \ldots, M$ and a uniform premium rate $p(i) = p$ with a multiplicative loading of 10% i.e. $0.9Ea(T) = EA(T)$. Note that in this case $A(i) = d^i$ and $a(i) = p(1-d^i)/(1-d)$. Hence $EA(T) = bd(1-d^M)/(M(1-d))$ and $Ea(T) = p(1/(1-d) - d(1-d^M)/(M(1-d)^2))$. The solution is $p = 0.20$.

Note that Poisson arrival leads to $\psi_N(\theta) = \lambda(e^\theta - 1)$ and $\psi_{C,t}(\theta) = \sum_{k=0}^{t} \lambda(g(\theta; t, k) - 1)$ with $g(\theta; t, k)$ defined in (4). For importance sampling, the sequential exponential tilting on the arrivals leads to a change in Poisson rate to $\lambda s g(\theta_t; t, k)$ at time $k \leq t$. The tilting on mortality distribution is given in (10).

To test our algorithm, we compute the finite-horizon ruin probability $P(C(t) > xs$ for some $t = 1, \ldots, T)$. We set the time horizon $T = 100$ and the initial surplus parameter $x = 0.05$ (for comparison, the expected gain to the insurance company per contract is $0.1Ea(T) = 0.11$). We run Algorithm 2 and compare it to crude Monte Carlo. Specifically, for each $s = 1, 5, 10, 15, 20, 30, 50$, we run crude Monte Carlo and Algorithm 2 each for one minute, then tabulate the estimate, relative error (ratio of empirical standard deviation to mean), and 95% confidence interval. We can see that our importance sampler consistently outperforms crude Monte Carlo with a smaller relative error. When s is 20 or above, crude Monte Carlo fails to give an estimate while importance sampler still gives reasonable output.

Table 1: Results of Crude Monte Carlo Experiment

s	Estimate	R.E.	C.I.
1	0.53	0.94	$(0.45, 0.56)$
5	0.074	3.53	$(0.058, 0.090)$
10	0.011	9.31	$(0.0050, 0.018)$
15	2.9×10^{-3}	18.6	$(-3.8 \times 10^{-4}, 6.1 \times 10^{-3})$
20	0	N/A	N/A
30	0	N/A	N/A
50	0	N/A	N/A

Table 2: Results of Monte Carlo Experiment with Importance Sampling

s	Estimate	R.E.	C.I.
1	0.49	0.62	$(0.47, 0.51)$
5	0.082	1.95	$(0.071, 0.093)$
10	0.0086	5.86	$(0.0049, 0.012)$
15	2.2×10^{-3}	6.38	$(1.2 \times 10^{-3}, 3.2 \times 10^{-3})$
20	4.2×10^{-4}	9.35	$(1.2 \times 10^{-4}, 7.2 \times 10^{-4})$
30	8.2×10^{-6}	11.5	$(1.3 \times 10^{-6}, 1.5 \times 10^{-5})$
50	8.2×10^{-9}	17.8	$(-2.5 \times 10^{-9}, 1.9 \times 10^{-8})$

REFERENCES

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

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

Asmussen, S, and P. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. Springer, New York.

Blanchet, J., P. Glynn, and H. Lam. 2009. Rare-event simulation for a slotted-time $M/G/s$ model. *Queueing Systems: Theory and Applications*, 63:33-57.

Blanchet, J., and H. Lam. 2011. Rare-event simulation for many-server queues. *Preprint*.

Blanchet, J., and H. Lam. 2011a. Computing ruin probability under a bottom-up model. *Working paper*.

Bowers, N., H. Gerber, J. Hickman, D. Jones, and C. Nesbitt. 1997. *Actuarial Mathematics*. Society of Actuaries.

Bucklew, J. 2004. *Introduction to rare-event simulation*. Springer-Verlag, New York.

Dembo, A., and O. Zeitouni. 1998. *Large Deviations Techniques and Applications*, 2nd Edition. Springer-Verlag, New York.

Hardy, M. 2003. *Investment Guarantees: Modeling and Risk Management for Equity-Linked Life Insurance*, John Wiley & Sons, New Jersey.

Szechtman, R., and P. Glynn. 2002. Rare event simulation for infinite server queues. *Proceedings of the 2002 Winter Simulation Conference*, 416-423.

AUTHOR BIOGRAPHIES

JOSE H. BLANCHET is an Assistant Professor in the Department of Industrial Engineering and Operations Research at Columbia University. Jose holds a M.Sc. in Engineering- Economic Systems and Operations Research and a Ph.D. in Management Science and Engineering, both from Stanford University. He also holds two B.Sc. degrees: one in Actuarial Science and another one in Applied Mathematics from ITAM (Mexico). Jose worked for two years as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability, computational finance, performance engineering, queueing theory, risk management, rare-event analysis, statistical inference, stochastic modelling, and simulation.

HENRY LAM is an Assistant Professor in the Department of Mathematics and Statistics at Boston University. He recently graduated from Harvard University with a Ph.D. degree in statistics. His research interests lie in applied probability and Monte Carlo methods with applications in queueing, operations management and insurance modeling.