

IMPORTANCE SAMPLING FOR INDICATOR MARKOV CHAINS

Kay Giesecke

Alexander D. Shkolnik

Department of Management Science and Engineering
Stanford University
Stanford, CA, USA

Institute for Computational and Mathematical Engineering
Stanford University
Stanford, CA, USA

ABSTRACT

We consider a continuous-time, inhomogeneous Markov chain M taking values in $\{0, 1\}^n$. Processes of this type arise in finance as models of correlated default timing in a portfolio of firms, in reliability as models of failure timing in a system of interdependent components, and in many other areas. We develop a logarithmically efficient importance sampling scheme for estimating the tail of the distribution of the total transition count of M at a fixed time horizon.

1 INTRODUCTION

An indicator Markov chain is a continuous-time Markov chain that takes values in $\mathbb{S} = \{0, 1\}^n$, starts at $0_n = (0, \dots, 0)$ and for which each of the n components has two states, 0 and 1, with 1 an absorbing state. A transition indicates the arrival of an event. While there are no simultaneous component transitions, the transitions of different components may be correlated. Thus, an indicator Markov chain arises naturally as a stochastic model of the timing of different events. In finance, for example, an indicator Markov chain can serve as a model of default timing in a portfolio of firms. Here, the value of the chain describes the state of the firms in the pool. In reliability, it can serve as a model of failure timing in a component system. Here, the value of the chain represents the state of the components.

An indicator Markov chain can also be defined as an indicator point process with an intensity that is a deterministic function of time and the value of the point process at that time. When defined in this manner, an indicator Markov chain appears to have a restrictive structure. However, its scope is broader. Consider an indicator point process $N = (N^1, \dots, N^n) \in \mathbb{S}$ with intensity $\lambda = (\lambda^1, \dots, \lambda^n)$. The process λ may be a function of auxiliary stochastic processes that represent randomly varying risk factors. In particular, N may not be Markov. By Proposition 3.1 in (Giesecke et al. 2009), there exists an indicator Markov chain $M = (M^1, \dots, M^n)$ such that $\mathbb{P}(N_T = B) = \mathbb{P}(M_T = B)$ for fixed T and all $B \in \mathbb{S}$. A component M^i has the transition rate function

$$p_n^i(t, B) = \mathbb{E}(\lambda_t^i \mathbf{1}(N_t^i = 0) | N_t = B), \quad B \in \mathbb{S}. \quad (1)$$

The existence of M reduces the problem of computing $\mathbb{E}(f(N_T))$ for an indicator point process N with general stochastic intensity to the simpler problem of computing the Markov chain expectation $\mathbb{E}(f(M_T))$. Thus, we are led to consider indicator Markov chains also in settings that are much more complex a priori.

Let 1_n be an n -vector of ones. Given an indicator Markov chain M , consider the process $C = 1_n \cdot M \in \{0, 1, \dots, n\}$ counting the transitions of M . We are often interested in the probability distribution of C_T for some fixed horizon T . In portfolio credit risk, where M serves as a (mimicking) default timing model, C_T represents the number of defaults in the portfolio at T , and the distribution of C_T measures the default risk in the portfolio. Of particular interest is the tail of that distribution, which describes the risk of extreme default scenarios. In reliability, when M models failures in a component system, the distribution of C_T describes the dependability of the system, and we are again interested in the tail of the distribution, i.e., the risk of system failure.

In practice, the dimension n is often large. In portfolio credit risk, n represents the number of names in the pool, and can reach several thousand. We consider estimating the distribution of C_T by Monte Carlo simulation of M rather than through alternative numerical methods that would be plagued by the high dimensionality of the state space \mathbb{S} . The main difficulty in computing the tail of the distribution of C_T using plain Monte Carlo is that the number of trials

required to estimate the probability of interest to a given relative precision scales in rough proportion to one over the probability. As a consequence, plain Monte Carlo is highly inefficient for estimating tail probabilities. This paper develops an importance sampling scheme for efficiently estimating the tail of the distribution of C_T . More precisely, our goal is to efficiently estimate the probabilities of the events

$$\xi_n = \{C_T \geq \mu n\}, \quad \mu \in (0, 1), \tag{2}$$

when $\mathbb{P}(\xi_n) \rightarrow 0$ as $n \rightarrow \infty$, i.e., when (ξ_n) is a rare-event sequence.

Our technical assumptions are as follows. We fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with right-continuous and complete information filtration (\mathcal{F}_t) . We let $M = (M^1, \dots, M^n) \in \mathbb{S}$ be a continuous-time Markov chain with the following properties: $M_0^i = 0$ almost surely and 1 is an absorbing state for M^i , for each $i = 1, \dots, n$. Further, there are no simultaneous transitions of any of the M^i . Hence, M is completely specified by a vector $(p_n^1, \dots, p_n^n) : \mathbb{R}_+ \times \mathbb{S} \rightarrow \mathbb{R}_+^n$ of transition functions satisfying $p_n^i(\cdot, B) = 0$ if $B^i = 1$. For simplicity, we also assume that $p_n^i(\cdot, B) > 0$ if $B^i = 0$.

2 RARE EVENT REGIME

A sequence of events (ξ_n) is called a rare-event sequence indexed by n if $\mathbb{P}(\xi_n) = \mathbb{P}(C_T \geq \mu n) \rightarrow 0$ as $n \rightarrow \infty$. As our goal is to take the sequence (ξ_n) out of the rare event regime, we wish to better understand the asymptotic behavior of the variable C_T with respect to n . To accomplish this task we relate C to two auxiliary counting processes which will be used to stochastically bound the random variable C_T and facilitate the analysis of its asymptotic behavior. As described in Section §3, we adopt a measure change which transforms C into a Poisson process stopped at the n th event time. For this reason we consider the Poisson process as our main instrument in studying the asymptotics of C_T . For extensions of this approach see the discussion following Theorem 1 of section §4.

More precisely, we say that a random variable X (stochastically) dominates C at time T whenever $\mathbb{P}(C_T \geq x) \leq \mathbb{P}(X \geq x)$ for all x . If the inequality is reversed we say X yields to C at time T . It follows that if X dominates C at time T and $\mathbb{P}(X \geq \mu n) \rightarrow 0$ as $n \rightarrow \infty$ we deduce that $\mathbb{P}(\xi_n) \rightarrow 0$ which implies that (ξ_n) is a rare-event sequence. Conversely, if X yields to C at time T and $\mathbb{P}(X \geq \mu n) \rightarrow 0$ as $n \rightarrow \infty$ we deduce that (ξ_n) is not in the rare event regime. In §2.1 we utilize a Poisson random variable to state our rare event regime condition. In §2.2 we analyse its asymptotics which will be used in proving our main result in Section §4.

2.1 Rare Event Conditions

Consider a Poisson process K with rate θ_n and event times (U_m) . Define the random variable $K_n = K_{U_n \wedge T}$. Note that both C_T and K_n take values in the set $\{0, 1, \dots, n\}$. Then, since $\{K_T \geq \mu n\} = \{U_{\lceil \mu n \rceil} \leq T\}$ we have

$$\mathbb{P}(K_T \geq \mu n) = \mathbb{P}\left(\frac{U_{\lceil \mu n \rceil} - \theta_n^{-1} \lceil \mu n \rceil}{\theta_n^{-1} \lceil \mu n \rceil^{1/2}} \leq \frac{T - \theta_n^{-1} \lceil \mu n \rceil}{\theta_n^{-1} \lceil \mu n \rceil^{1/2}}\right) = \mathbb{P}\left(\frac{U_{\lceil \mu n \rceil} - \theta_n^{-1} \lceil \mu n \rceil}{\theta_n^{-1} \lceil \mu n \rceil^{1/2}} \leq -\psi_n(v)\right) \tag{3}$$

where we define the sequence $v_n = \frac{T\theta_n}{\lceil \mu n \rceil}$ and the function ψ as

$$\psi_n(v) = (1 - v_n) \lceil \mu n \rceil^{1/2}. \tag{4}$$

Since the random variables $(U_m - U_{m-1})$ are i.i.d. exponential of mean θ_n^{-1} and variance θ_n^{-2} we apply the CLT to

$$g_n(z) = \mathbb{P}\left(\frac{U_{\lceil \mu n \rceil} - \theta_n^{-1} \lceil \mu n \rceil}{\theta_n^{-1} \lceil \mu n \rceil^{1/2}} \leq -z\right) \tag{5}$$

to conclude that $\lim_{n \rightarrow \infty} g_n(z) = \Phi(-z)$ where $\Phi(\cdot)$ is the standard Normal distribution function. Observe that if $\psi_n(v) \rightarrow \infty$ then there exists n sufficiently large so that $0 \leq g_n(\psi_n(v)) \leq g_n(z) \rightarrow \Phi(-z)$ for any finite z . Since $\Phi(-\infty) = 0$ we deduce the following rare event condition.

Condition 1. Let $\mu \in (0, 1)$ and K be a Poisson process with rate $T^{-1} \lceil \mu n \rceil v_n$ such that $K_n = K_{U_n \wedge T}$ dominates C at time T . Then $\xi_n = \{C_T \geq \lceil \mu n \rceil\}$ is a rare-event sequence and $\mathbb{P}(\xi_n) \rightarrow 0$ as $n \rightarrow \infty$ if

$$\liminf_{n \rightarrow \infty} \psi_n(v) = \infty \tag{6}$$

Conversely, if K_n yields to C at time T and

$$\liminf_{n \rightarrow \infty} \psi_n(v) < \infty \tag{7}$$

then $\mathbb{P}(C_T \geq \lceil \mu n \rceil) \rightarrow 0$ and (ξ_n) is not a rare-event sequence.

A simpler condition for the rare event regime and one that we will eventually use is as follows.

Condition 2. Let $\mu \in (0, 1)$ and K be a Poisson process with rate $T^{-1} \lceil \mu n \rceil v_n$ such that $K_n = K_{U_n \wedge T}$ dominates C at time T . Then $\xi_n = \{C_T \geq \lceil \mu n \rceil\}$ is a rare-event sequence and $\mathbb{P}(\xi_n) \rightarrow 0$ as $n \rightarrow \infty$ if

$$\limsup_{n \rightarrow \infty} v_n < 1 \tag{8}$$

Notice that Condition 2 is more restrictive than Condition 1 but is simpler and, as reflected in the next section, yields a simpler analysis of the asymptotics of the dominating Poisson process which we have defined.

2.2 Poisson Asymptotics

Here we consider K to be a Poisson process with rate $T^{-1} k v_k$ and as before define the random variable $K_n = K_{U_n \wedge T}$. The distribution of K_n is given by

$$\mathbb{P}(K_n = \ell) = \exp(-k v_k) \frac{(k v_n)^\ell}{\ell!} \quad \ell < n \tag{9}$$

$$\mathbb{P}(K_n = \ell) = \exp(-k v_k) \sum_{m=n}^{\infty} \frac{(k v_k)^m}{m!} \quad \ell = n \tag{10}$$

$$\mathbb{P}(K_n = \ell) = 0 \quad \ell > n. \tag{11}$$

It follows that for $k \leq n$

$$\mathbb{P}(K_n \geq k) = \exp(-k v_k) \frac{(k v_n)^k}{k!} r_k \tag{12}$$

where $r_k \geq 1$ is defined as

$$r_k = \lim_{\ell \rightarrow \infty} r_{k\ell} = \lim_{\ell \rightarrow \infty} \sum_{m=0}^{\ell} \frac{(k v_k)^m k!}{(k+m)!} = \lim_{\ell \rightarrow \infty} \sum_{m=0}^{\ell} v_k^m \frac{k^m k!}{(k+m)!}. \tag{13}$$

The partial sums satisfy $r_{k\ell} - r_{k(\ell-1)} \leq v_k^\ell$, thus if $\limsup_{k \rightarrow \infty} v_k = b < 1$, the sequence $r_{k\ell}$ is uniformly Cauchy convergent and thus uniformly convergent. This implies we can exchange the limit and the sum, yielding

$$\limsup_{k \rightarrow \infty} r_k = \lim_{\ell \rightarrow \infty} \left(\limsup_{k \rightarrow \infty} r_{k\ell} \right) \leq \frac{1}{1-b}. \tag{14}$$

Similarly, $\liminf_{k \rightarrow \infty} v_k = a \leq b < 1$ implies $\liminf_{k \rightarrow \infty} r_k \geq 1/(1-a)$. We now use Sterling's approximation $\log k! = k \log k - k + \log \sqrt{2\pi k} + O(k^{-1})$ to deduce that for $k \leq n$

$$\frac{1}{k} \log \mathbb{P}(K_n \geq k) = -v_k + \log(k v_k) - k^{-1} \log k! + k^{-1} \log r_k \tag{15}$$

$$= 1 - v_k + \log v_k + k^{-1} \log r_k - O(k^{-1} \log k) \tag{16}$$

Letting $\varphi_k(v) = -1 + v_k - \log v_k$ we have for $0 \leq a = \liminf_{k \rightarrow \infty} v_k \leq \limsup_{k \rightarrow \infty} v_k = b < 1$ and $k \leq n$

$$\frac{1}{k} \log \mathbb{P}(K_n \geq k) = -\varphi_k(v) - O(k^{-1} \log k). \tag{17}$$

Known as the rate function, φ dictates the rate of exponential decay of $\mathbb{P}(K_n \geq k)$ for $n \geq k \rightarrow \infty$. Note that φ is strictly decreasing on $[0, 1]$ with $\varphi_n(0) = \infty$ and $\varphi_n(1) = 0$ and is strictly positive on $[0, 1)$.

Since our asymptotic analysis holds only for $\limsup_n v_n < 1$ we use Condition 2 as the relevant rare event regime setting. To this end, letting K be a Poisson process with rate $T^{-1} \lceil \mu n \rceil v_n$ with $\limsup_{n \rightarrow \infty} v_n < 1$, we have that

$n \mapsto \{K_{U_n \wedge T} \geq \mu n\}$ is a rare-event sequence. Furthermore, if K dominates C at time T we have

$$\frac{1}{\lceil \mu n \rceil} \log \mathbb{P}(C_T \geq \mu n) \leq \frac{1}{\lceil \mu n \rceil} \log \mathbb{P}(K_{U_n \wedge T} \geq \mu n) = -\varphi_n(v) - O(n^{-1} \log n) \tag{18}$$

where

$$\varphi_n(v) = -1 + v_n - \log v_n. \tag{19}$$

Similarly, if K yields to C at time T then the inequality in (18) is reversed.

3 IMPORTANCE SAMPLING

We wish to efficiently estimate the probabilities $y_n = \mathbb{P}(\xi_n) = \mathbb{P}(C_T \geq \mu n)$ when (ξ_n) is rare. We could obtain samples of the estimator $\mathbf{1}(\xi_n)$ of y_n by simulating M under the reference probability measure \mathbb{P} . Averaging the samples gives an estimate \hat{y}_n which by the LLN converges to y_n almost surely as the number of samples tends to infinity. However, the variance σ_n^2 of the plain Monte Carlo (MC) estimator $\mathbf{1}(\xi_n)$ is $y_n(1 - y_n)$ and consequently, the relative error is given by $\zeta_n = \sigma_n/y_n = 1/\sqrt{y_n}$. This relative error tends to infinity as $n \rightarrow \infty$ in the rare event regime. We use importance sampling (IS) to circumvent this difficulty.

Importance sampling entails an absolutely continuous change of probability measure from \mathbb{P} to \mathbb{Q} and the subsequent simulation of M under \mathbb{Q} . The key identity is

$$\mathbb{E}_{\mathbb{P}}[f(M_T)] = \mathbb{E}_{\mathbb{Q}}[Z_T f(M_T)] \tag{20}$$

where f is a \mathbb{P} -integrable function on \mathbb{S} and Z_T is the Radon-Nikodym derivative of \mathbb{P} with respect to \mathbb{Q} defined on \mathcal{F}_T . We consider the indicator function $f(B) = \mathbf{1}(1_n \cdot B \geq \mu n)$ and the IS estimator of y_n is

$$Y_n = Z_T \mathbf{1}(\xi_n). \tag{21}$$

This estimator satisfies $\mathbb{E}_{\mathbb{Q}}[Y_n] = \mathbb{P}(\xi_n)$. We chose an importance measure \mathbb{Q} so that the relative error of Y_n under \mathbb{Q} is much smaller than that of the plain MC estimator $\mathbf{1}(\xi_n)$ under \mathbb{P} .

3.1 Measure Change

We propose a family of importance measures $(\mathbb{Q})_{\theta_n}$ parametrized by $\theta_n \geq 1$ under which M is still an indicator Markov chain and $C = 1_n \cdot M$ is a Poisson process with rate θ_n stopped at the n^{th} event time. In this setting, the IS estimator is easily generated. Note that the \mathbb{P} -dynamics of C are more complicated: the intensity of C is $p_n(\cdot, M)$ where

$$p_n(t, B) = \sum_{i=1}^n p_n^i(t, B). \tag{22}$$

Under the conditions stated in Section §1, a vector of rate functions (q_n^1, \dots, q_n^n) specifies M under the new measure \mathbb{Q} . We define these new rate functions as follows, using the convention $0/0 = 0$:

$$q_n^i(t, B) = \frac{p_n^i(t, B)}{p_n(t, B)} \theta_n. \tag{23}$$

Since $q_n^i(t, B) = 0 \Rightarrow p_n^i(t, B) = 0$, the measure \mathbb{P} is absolutely continuous with respect to \mathbb{Q} . Furthermore, under \mathbb{Q} the counting process C with event times (S_n) has intensity on $[0, T]$ given by

$$q_n(t, M_t) = \sum_{i=1}^n q_n^i(t, M_t) = \theta_n \mathbf{1}(C_t < n) = \theta_n \mathbf{1}(S_n > t). \tag{24}$$

It follows that under \mathbb{Q} , C is a Poisson process with rate θ_n stopped at S_n . There are alternative definitions of the $q_n^i(t, B)$ that accomplish such a transformation. But, the attractive nature of the measure change corresponding to (23) is that the \mathbb{Q} -distribution of M mimics the “shape” of its \mathbb{P} -distribution which is desirable for a good IS scheme (Glynn and Asmussen 2007).

The next step is to pick a parameter θ_n . From Theorem VI.3 in (Bremaud 1981), the density Z_T takes the form

$$Z_T = \prod_{i=1}^n \exp \left(\int_0^T \log \left(\frac{p_n^i(s-, M_{s-})}{q_n^i(s-, M_{s-})} \right) dM_s^i - \int_0^T (p_n^i(s, M_s) - q_n^i(s, M_s))(1 - M_s^i) ds \right). \tag{25}$$

With the parametrization (23), we get

$$Z_T = Z_T(\theta_n) = \exp((S_n \wedge T) \theta_n - C_T \log(T \theta_n) + \mathcal{D}_T) \tag{26}$$

where \mathcal{D}_T , the \mathbb{P} -log-likelihood of the event times $\{S_m : m \leq C_T\}$ is

$$\mathcal{D}_T = \int_0^T \log(T p_n(s-, M_{s-})) dC_s - \int_0^T p_n(s, M_s) ds. \tag{27}$$

To minimize the relative error produced by the IS estimator $Y_n = Z_T(\theta_n) \mathbf{1}(\xi_n)$, we consider its second moment:

$$\mathbb{E}_{\mathbb{Q}} [Y_n^2] = \mathbb{E}_{\mathbb{P}} [Z_T(\theta_n) ; \xi_n] \leq \exp(\theta_n T - \lceil \mu n \rceil \log(T \theta_n)) \mathbb{E}_{\mathbb{P}} [\exp(\mathcal{D}_T) ; \xi_n] \tag{28}$$

for all $\theta_n \geq 1$. Minimizing the second moment is difficult, but minimizing the bound (28) over $\theta_n \geq 1$ is easy. Since the \mathbb{P} -expectation on the right side of (28) does not depend on θ_n , for $\lceil \mu n \rceil \geq T$ the minimizer θ_n^* is given by

$$\theta_n^* = T^{-1} \lceil \mu n \rceil. \tag{29}$$

To better understand the measure change corresponding to (29) note that $C - \int_0^\cdot \theta_n \mathbf{1}(S_n > s) ds$ is a \mathbb{Q} -martingale on $[0, T]$ which implies that

$$\mathbb{E}_{\mathbb{Q}} [C_T] = \mathbb{E}_{\mathbb{Q}} \left[\int_0^T \theta_n^* \mathbf{1}(S_n > s) ds \right] = \frac{\lceil \mu n \rceil}{T} \mathbb{E}_{\mathbb{Q}} [T \wedge S_n] \approx \lceil \mu n \rceil \tag{30}$$

for large n . The \mathbb{Q} -rates of M are shifted such that ξ_n is no longer a rare event with respect to n . Indeed, since under \mathbb{Q} , the event time $S_{\lceil \mu n \rceil}$ of C is the sum of $\lceil \mu n \rceil$ independent exponential random variables with parameter $T^{-1} \lceil \mu n \rceil$, similarly to (3), we obtain

$$\mathbb{Q}(\xi_n) = \mathbb{Q}(C_T \geq \mu n) = \mathbb{Q}(S_{\lceil \mu n \rceil} \leq T) \rightarrow \frac{1}{2} \tag{31}$$

as $n \rightarrow \infty$. For large n , roughly half of the samples of C_T generated under \mathbb{Q} will be larger than $\lceil \mu n \rceil$.

3.2 IS Algorithm

To obtain estimates of $\mathbb{P}(\xi_n)$ using IS we simulate M under the importance measure $\mathbb{Q}_{\theta_n^*}$ and compute samples of the estimator $Y_n = Z_T(\theta_n^*) \mathbf{1}(\xi_n)$ where $\theta_n^* = T^{-1} \lceil \mu n \rceil$. The implementation is particularly simple since the waiting times $(S_m - S_{m-1})$ between transitions of M are i.i.d. exponential with parameter $T^{-1} \lceil \mu n \rceil$ and

$$\mathbb{Q}(I_m = i | \mathcal{F}_{S_m^-}) = \frac{q_n^i(S_m, M_{S_{m-1}})}{q_n(S_m, M_{S_{m-1}})} = \mathbb{P}(I_m = i | \mathcal{F}_{S_m^-}), \quad i = 1, 2, \dots, n \tag{32}$$

where $I_m \in \{1, \dots, n\}$ is the component of M in which a transition occurs at time S_m . We generate a sufficient number of IS trials, performing the following steps on each:

- Generate event times (S_m) up to time T of a Poisson process with rate $T^{-1} \lceil \mu n \rceil$.
- For each S_m draw the component I_m of M at which the transition took place from the distribution (32).
- Using the values (S_m, I_m) compute a sample $Z_T(\lceil \mu n \rceil / T) \mathbf{1}(\xi_n)$ of Y_n using $Z_T(\cdot)$ defined in (26).

The samples are then averaged to obtain the IS estimate, which then gives an estimate of $\mathbb{P}(\xi_n)$ by (20).

4 LOGARITHMIC EFFICIENCY

A standard measure of the performance of an IS estimator is logarithmic efficiency. This property states that the relative error of the estimator Y_n grows slower than the rate at which $\mathbb{P}(\xi_n) \rightarrow 0$. Letting

$$\rho_n = \frac{\log \mathbb{E}_{\mathbb{Q}} [Y_n^2]}{\log \mathbb{E}_{\mathbb{Q}} [Y_n]}, \tag{33}$$

an estimator Y_n is said to be *logarithmically efficient* if and only if

$$\liminf_{n \rightarrow \infty} \rho_n = 2, \tag{34}$$

see (Glynn and Asmussen 2007). We next prove that our IS estimator is logarithmically efficient. We do not require specific information about the rates of the indicator chain M at hand, and prove (34) under the rare event condition and a condition which relates C to a dominating and a yielding Poisson processes. In particular, we require that the rate functions (19) of the dominating and yielding Poisson processes have the same asymptotics. The interpretation of such a condition is that the oscillations in $p_n(\cdot, \cdot)$ must not grow linearly in n as $n \rightarrow \infty$.

Theorem 1. Define the sequences (α_n) and (β_n) by

$$\alpha_n = \inf_{\mathcal{B}} \frac{Tp_n(s, B)}{\lceil \mu n \rceil} \quad \text{and} \quad \beta_n = \sup_{\mathcal{B}} \frac{Tp_n(s, B)}{\lceil \mu n \rceil} \tag{35}$$

where $\mathcal{B} = \{(s, B) : s \in [0, T] \text{ and } B \in \{0, 1\}^n \setminus \{1_n\}\}$. If

$$\limsup_{n \rightarrow \infty} \beta_n < 1 \tag{36}$$

$$\liminf_{n \rightarrow \infty} \frac{\varphi_n(\beta)}{\varphi_n(\alpha)} = 1, \tag{37}$$

then the estimator $Y_n = Z_T(\lceil \mu n \rceil / T) \mathbf{1}(\xi_n)$, where $Z_T(\cdot)$ is given by (26), is logarithmically efficient.

Proof. It is easy to show that $\rho_n \leq 2$ for all n by simple calculation. Thus, it is sufficient to show that $\liminf_{n \rightarrow \infty} \rho_n \geq 2$. First, a thinning argument as in Proposition 2.1 of (Giesecke and Kim 2008) shows that the random variable $K_{U_n \wedge T}$, where K is a Poisson process with rate $T^{-1} \lceil \mu n \rceil \beta_n$, (stochastically) dominates C_T . Following Condition 2, (36) implies we are in the rare event regime. So, from (18) we deduce that for large enough n and $0 < \delta_n = O(n^{-1} \log n) \downarrow 0$

$$0 > \frac{\log \mathbb{E}_{\mathbb{Q}} [Y_n]}{\varphi_n(\alpha) \lceil \mu n \rceil} \geq \frac{\log \mathbb{P}(K_{U_n \wedge T} \geq \mu n)}{\varphi_n(\alpha) \lceil \mu n \rceil} = -1 + \delta_n \tag{38}$$

where φ_n is the Poisson rate function defined in (19).

Then, starting from $Y_n = Z_T(\lceil \mu n \rceil / T) \mathbf{1}(\xi_n)$ and using (26) and (35) yields the following bound:

$$\begin{aligned} Y_n &= \mathbf{1}(\xi_n) \exp((S_n \wedge T) T^{-1} \lceil \mu n \rceil - C_T \log \lceil \mu n \rceil + \mathcal{D}_T) \\ &\leq \mathbf{1}(\xi_n) \exp((S_n \wedge T) T^{-1} \lceil \mu n \rceil - C_T \log \lceil \mu n \rceil + C_T \log(\lceil \mu n \rceil \beta_n) - (S_n \wedge T) T^{-1} \lceil \mu n \rceil \alpha_n) \\ &\leq \mathbf{1}(\xi_n) \exp((S_n \wedge T) T^{-1} \lceil \mu n \rceil (1 - \alpha_n) + C_T \log \beta_n). \end{aligned}$$

Again, by (36) we have that for n large enough this satisfies

$$\begin{aligned} Y_n &\leq \mathbf{1}(C_T \geq \lceil \mu n \rceil) \exp(-\lceil \mu n \rceil (-1 + \alpha_n - \log \beta_n)) \\ &= \mathbf{1}(\xi_n) \exp(-\lceil \mu n \rceil \varphi_n(\beta)) \exp((\beta_n - \alpha_n) \lceil \mu n \rceil). \end{aligned} \tag{39}$$

Note that since $\varphi_n(\cdot)$ is strictly decreasing on $[0, 1]$, we have the following two relationships:

$$0 \geq \varphi_n(\beta) - \varphi_n(\alpha) = \beta_n - \alpha_n - \log \frac{\beta_n}{\alpha_n} \tag{40}$$

$$0 \leq \varphi_n(\alpha) - \varphi_n(\beta) = \alpha_n - \beta_n + \log \frac{\beta_n}{\alpha_n} \leq \log \frac{\beta_n}{\alpha_n}. \tag{41}$$

Then $\beta_n - \alpha_n \leq \log(\beta_n/\alpha_n)$ and using (39) we obtain

$$\begin{aligned} \frac{\log \mathbb{E}_{\mathbb{Q}} [Y_n^2]}{\varphi_n(\alpha) \lceil \mu n \rceil} &\leq -2 \frac{\varphi_n(\beta)}{\varphi_n(\alpha)} + \frac{2(\beta_n - \alpha_n)}{\varphi_n(\alpha)} + \frac{\log \mathbb{Q}(\xi_n)}{\varphi_n(\alpha) \lceil \mu n \rceil} \\ &\leq -2 \frac{\varphi_n(\beta)}{\varphi_n(\alpha)} + \frac{2}{\varphi_n(\alpha)} \log \frac{\beta_n}{\alpha_n} + \frac{\log \mathbb{Q}(\xi_n)}{2\varphi_n(\alpha) \lceil \mu n \rceil}. \end{aligned} \tag{42}$$

Taking the ratio of (42) and (38) yields

$$\frac{\log \mathbb{E}_{\mathbb{Q}} [Y_n^2]}{\log \mathbb{E}_{\mathbb{Q}} [Y_n]} \geq \frac{2 \varphi_n(\beta)}{\varphi_n(\alpha) (1 - \delta_n)} - \frac{2}{\varphi_n(\alpha) (1 - \delta_n)} \log \frac{\beta_n}{\alpha_n} - \frac{\log \mathbb{Q}(\xi_n)}{2\varphi_n(\alpha) (1 - \delta_n) \lceil \mu n \rceil} \tag{43}$$

where we observe by (41) and condition (37) that

$$0 \geq \frac{1}{\varphi_n(\alpha) (1 - \delta_n)} \log \frac{\beta_n}{\alpha_n} \geq \frac{\varphi_n(\alpha) - \varphi_n(\beta)}{\varphi_n(\alpha) - \delta_n} \geq 1 - \frac{\varphi_n(\beta)}{\varphi_n(\alpha)} \downarrow 0.$$

Finally, taking the infimum and again applying (37) along with the fact that $\mathbb{Q}(\xi_n) \rightarrow 1/2$ yields

$$\liminf_{n \rightarrow \infty} \frac{\log \mathbb{E}_{\mathbb{Q}} [Y_n^2]}{\log \mathbb{E}_{\mathbb{Q}} [Y_n]} \geq 2 \tag{44}$$

as required. Therefore, Y_n is logarithmically efficient. □

Theorem 1 could be extended to incorporate milder conditions on the oscillations of the transition rate function p_n of C by replacing the dominating and yielding Poisson processes with pure birth processes stopped at the n^{th} event time. This would allow us to consider oscillations not over the entire state space \mathbb{S} but over the states B for which $1_n \cdot B$ is constant. Here, we would define α_n^m and β_n^m in (35) to be the infimum and supremum taken over the set $\mathcal{B}^m = \{(s, B) : s \in [0, T] \text{ and } B \in \{0, 1\}^n, 1_n \cdot B = m\}$. This extension would then take into account that p_n also depends on the number of components of M which are in state 0.

5 NUMERICAL RESULTS

We demonstrate the effectiveness of the IS scheme on a model of correlated default timing in a portfolio of n firms proposed by (Giesecke et al. 2009, Section 6). Here, the indicator chain M of interest mimics an indicator point process $N = (N^1, \dots, N^n)$ with a stochastic intensity $\lambda = (\lambda^1, \dots, \lambda^n)$ that follows a jump-diffusion process:

$$\lambda_t^i = X_t^i + \sum_{j \neq i} \beta^{ij} N_t^j \tag{45}$$

where $\beta^{ij} \geq 0$ and the risk factors X^i follow mutually independent Feller diffusions

$$dX_t^i = \kappa_i (c_i - X_t^i) dt + \sigma_i \sqrt{X_t^i} dW_t^i, \quad X_0^i > 0. \tag{46}$$

Here, κ_i is a parameter controlling the speed of mean-reversion of X^i , c_i is the level of mean reversion, and σ_i controls the diffusive volatility of X^i . The process (W^1, \dots, W^n) is a standard Brownian motion. The parameter β^{ij} determines the impact on firm i of firm j 's default. The corresponding jump terms generate correlation between the firm intensities. The matrix (β^{ij}) governs the default dependence structure. The transition rates $p_n^i(t, B)$ of the indicator chain M mimicking N in the sense that $\mathbb{P}(M_T = B) = \mathbb{P}(N_T = B)$ for all $B \in \mathbb{S}$ are computed according to formula (1). Giesecke et al. (2009) show that, for $\gamma_i = (\kappa_i^2 + 2\sigma_i^2)^{1/2}$,

$$p_n^i(t, M_t) = \frac{4X_0^i \gamma_i^2 \exp(\gamma_i t)}{(\gamma_i - \kappa_i + (\gamma_i + \kappa_i) \exp(\gamma_i t))^2} + \frac{2\kappa_i c_i (\exp(\gamma_i t) - 1)}{\gamma_i - \kappa_i + (\gamma_i + \kappa_i) \exp(\gamma_i t)} + \sum_{j \neq i} \beta^{ij} M_t^j \tag{47}$$

We consider a portfolio of $n = 100$ firms. The model parameters are drawn randomly, but such that the portfolio represents a pool of high-quality firms. With the exception of the matrix (β_{ij}) , the parameters are sampled as in (Giesecke et al. 2009). The matrix (β_{ij}) is scaled down by a factor of 10 to ensure that the indicator chain M is in the rare event regime. Consequently, following Condition 2 of Section §2, we ensure that $p_n(\cdot, B) < T^{-1} \lceil \mu n \rceil$, assuming that $n = 100$ is large enough for the asymptotic regime to be in effect.

$\lceil \mu n \rceil$	IS Estimate \pm 95% CI	IS Scheme Variance	Plain MC Variance	Variance Ratio
1	9.480484E-01 \pm 5.89732E-02	9.053097E-00	6.494702E-02	7.174011E-03
2	7.522197E-01 \pm 1.46434E-02	5.023614E-01	1.916260E-01	3.814504E-01
3	4.906572E-01 \pm 8.91635E-03	1.862536E-01	2.499024E-01	1.341732E-00
4	2.660202E-01 \pm 5.64038E-03	8.281410E-02	1.962571E-01	2.369851E-00
5	1.196635E-01 \pm 3.07897E-03	2.467735E-02	1.080868E-01	4.379999E-00
6	4.823202E-02 \pm 1.39821E-03	5.089037E-03	4.594010E-02	9.027268E-00
7	1.679726E-02 \pm 5.32909E-04	7.392553E-04	1.622778E-02	2.195152E+01
8	4.851613E-03 \pm 1.70985E-04	7.610350E-05	4.923527E-03	6.469515E+01
9	1.321480E-03 \pm 4.91585E-05	6.290501E-06	1.300307E-03	2.067096E+02
10	3.043514E-04 \pm 1.23209E-05	3.951607E-07	3.099045E-04	7.842493E+02
11	6.616864E-05 \pm 2.79334E-06	2.031115E-08	8.199344E-05	4.036868E+03
12	1.267209E-05 \pm 5.61857E-07	8.217485E-10	1.799971E-05	2.190416E+04
13	2.199999E-06 \pm 1.02546E-07	2.737301E-11	3.999992E-06	1.461291E+05
14	3.484863E-07 \pm 1.70216E-08	7.542041E-13	(*)	
15	5.439357E-08 \pm 2.71767E-09	1.922568E-14	(*)	
16	7.538963E-09 \pm 3.89113E-10	3.941290E-16	(*)	
17	1.013618E-09 \pm 5.27912E-11	7.254569E-18	(*)	
18	1.165917E-10 \pm 6.47477E-12	1.091282E-19	(*)	
19	1.293590E-11 \pm 7.43057E-13	1.437251E-21	(*)	
20	1.499824E-12 \pm 8.51218E-14	1.886121E-23	(*)	

Table 1: IS estimates of probabilities $\mathbb{P}(C_T \geq \mu n)$ for various values of μn . For comparison the variance of the IS and plain MC estimators are supplied along with the variance ratios showing the variance reduction obtained by the IS scheme. (*) indicates that no events were observed for the given μ value.

We compare the plain Monte Carlo (MC) estimator of $\mathbb{P}(C_T \geq k)$ with the corresponding IS estimator. The MC estimator $\mathbf{1}(\xi_n)$ is obtained by simulating M using the rates (p_n^1, \dots, p_n^n) up to time $T \wedge S_{\lceil \mu n \rceil}$. Event times (S_m) of C are generated using the acceptance/rejection Algorithm 4.1 of Giesecke et al. (2009). In the plain MC setting we only need to simulate up to time $T \wedge S_{\lceil \mu n \rceil}$, and the samples lead to estimates of $\mathbb{P}(C_T \geq \mu n)$ for any value of μ simultaneously. In the IS scheme we need to generate the full path up to time T and we cannot use the same samples for different μ values. For each value of $\mu \in \{0.01, 0.02, \dots, 0.2\}$, we use 10K trials to compute the IS estimator. The plain Monte Carlo simulation is based on 500K trials. The generation of the 500K trials takes roughly the same time as the generation of the $20 \times 10K$ IS trials. All source code was implemented with R package and numerical experiments conducted on an 64bit Intel® Core™ 2 Quad CPU at 2.40GHz. The plain MC simulation and all IS simulations took roughly 14 hours in total to complete.

Table 1 shows the estimates of the tail probabilities $\mathbb{P}(C_T \geq \mu n)$. The 95% confidence intervals on the IS estimates are included. The table also shows the variance reduction achieved by the IS scheme relative to plain Monte Carlo. This is measured by the ratio of the sample variance of the IS estimator and the sample variance of the plain MC estimator. The results show that the IS scheme starts exhibiting variance reduction as early as the center of the distribution of C_T near $\lceil \mu n \rceil = 3$. The variance ratio grows rapidly as we estimate values into the tail.

Figure 1 shows the estimates of the tail probabilities $\mathbb{P}(C_T \geq \mu n)$ obtained by plain MC and the IS scheme on a log-plot. We plot the 95% confidence intervals which, for demonstration purposes, have been magnified by a factor of 5. We do not display lower error bars for the estimates at which the lower confidence interval is below zero. Each estimate is indicated with the specified symbol and the upper and lower error bars are interpolated to demonstrate the way in which the estimate confidence changes for each scheme as μ increases. Plain MC loses accuracy early on as the 95% confidence intervals diverge away from the estimated probability. The IS estimates are accurate well into the tail with each estimate having relatively the same degree of precision.

REFERENCES

- Bremaud, P. 1981. *Point processes and queues: Martingale dynamics*. New York: Springer-Verlag.
- Giesecke, K., H. Kakavand, M. Mousavi, and H. Takada. 2009. Exact and efficient simulation of correlated defaults. *SIAM Journal on Financial Mathematics*, forthcoming.
- Giesecke, K., and B. Kim. 2008. Risk analysis of collateralized debt obligations. *Operations Research*, forthcoming.
- Glynn, P., and S. Asmussen. 2007. *Stochastic simulation: Algorithms and analysis*. New York: Springer.

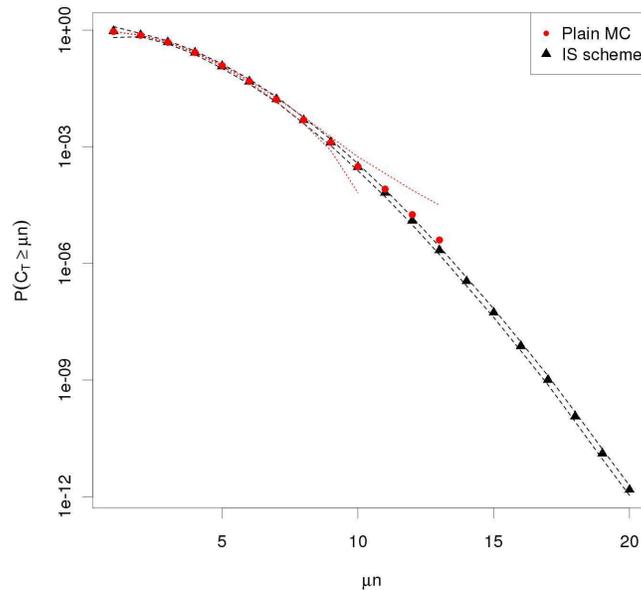


Figure 1: IS and plain MC estimates of the probabilities $\mathbb{P}(C_T \geq \mu n)$. The estimated values for each μn are shown by the corresponding symbol and 95% confidence intervals are indicated with dashed lines above and below the estimated values. The confidence intervals have been magnified by a factor of 5.

ACKNOWLEDGMENTS

We kindly thank David Sigmund for some pointers to literature as well as useful suggestions on section §5 of the manuscript.

AUTHOR BIOGRAPHIES

KAY GIESECKE is an assistant professor in the Department of Management Science and Engineering at Stanford University. He has previously held a position in the School of Operations Research and Information Engineering at Cornell University. His research is in financial engineering, particularly credit risk modeling. His email is giesecke@stanford.edu and his web page is www.stanford.edu/giesecke.

ALEXANDER D. SHKOLNIK is a Ph.D. candidate in the Institute for Computational and Mathematical Engineering at Stanford. He graduated with B.S. degrees in Applied Mathematics, Physics and Computer Science from Carnegie Mellon University. His research interests include applied probability, rare event simulation algorithms and credit risk modelling.